Revisiting Pierre Gy's formula (TOS) – A return to size-density classes for applications to contaminated soils, coated particular aggregates and mixed material systems

Dubé, Jean Sébastien; Esbensen, Kim H.

Published in:
Analytica Chimica Acta

DOI (link to publication from Publisher):
10.1016/j.aca.2021.339227

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Analytica Chimica Acta 1193 (2022) 339227

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

Revisiting Pierre Gy's formula (TOS) – A return to size-density classes for applications to contaminated soils, coated particular aggregates and mixed material systems

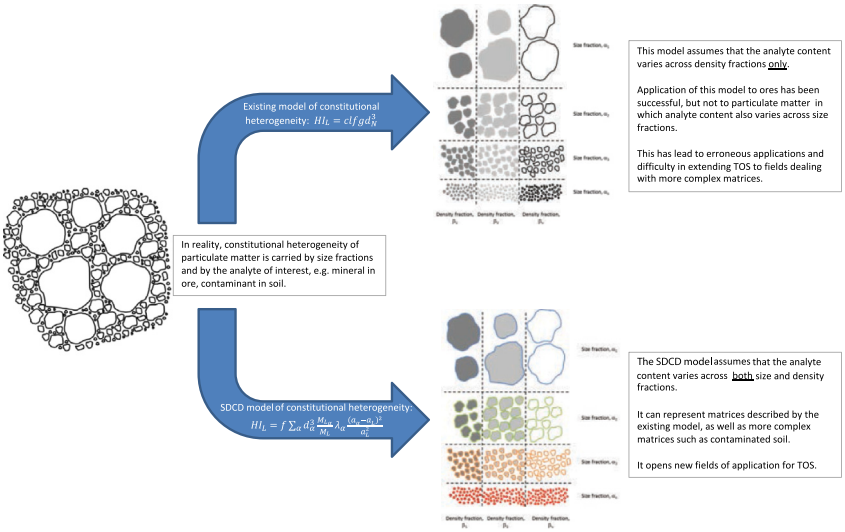
Jean-Sébastien Dubé [a, *], Kim H. Esbensen [b, c, d, e]

[a] Laboratory for Geotechnical and Geoenvironmental Engineering, École de Technologie Supérieure (ETS). 1100 Notre-Dame Ouest, Montreal, H2L 3M4, Canada
[b] KHE Consulting[1], Copenhagen, Denmark
[c] Visiting professor (2017), ETS, Montreal, Canada
[d] Aalborg University, Denmark
[e] Geological Survey of Denmark and Greenland (GEUS), Denmark

HIGHLIGHTS

- A revised heterogeneity invariant model is presented which assumes that constitutional heterogeneity is correlated to size fraction and phase density.
- The model successfully reproduced sampling variance due to constitutional heterogeneity from previous studies on analyte-enriched particulate matter.
- The model successfully corrected sampling variance due to constitutional heterogeneity from previous studies on analyte-coated particulate matter.
- It is a practical model of constitutional heterogeneity extended to mixed particulate matter composed of analyte-enriched and analyte-coated particles.
- The model can be practically implemented in sampling protocol design.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 4 June 2021
Received in revised form
5 October 2021
Accepted 27 October 2021
Available online 2 November 2021

ABSTRACT

For some real-world material systems, estimations of the incompressible sampling variance based on Gy's classical $s^2(FSE)$ formula from the Theory of Sampling (TOS) show a significant discrepancy with empirical estimates of sampling variance. In instances concerning contaminated soils, coated particular aggregates and mixed material systems, theoretical estimates of sampling variance are larger than empirical estimates, a situation which does not have physical meaning in TOS. This has led us to revisit the development of estimates of $s^2(FSE)$ from this famous constitutional heterogeneity equation and explore the use of size-density classes for *mixed material systems* (mixtures of both analyte-enriched and

* Corresponding author.
E-mail address: jean-sebastien.dube@etsmtl.ca (J.-S. Dubé).
[1] www.khe.consult

https://doi.org/10.1016/j.aca.2021.339227

Keywords:
Sampling particulate matter
Constitutional heterogeneity
Heterogeneity invariant
Gy's formula
Fundamental sampling error
Sampling variance

coated particles), an approach which has been mostly unused since Gy's original derivation. This approach makes it possible to avoid taking into account the granulometric and liberation factors from Gy's classical treatment, and present grounds for criticising the use of 'standard' input values of critical parameters such as f: = 0.5, and g: = 0.25. But, as always, the "liberation factor" (l) issue still plays an important role, which is paid due attention. The constitutional heterogeneity formula based on size-density classes is presented in a form that allows for easy implementation in practice, within specified limitations. We present extensive experimental results from real-world systems. Using the "SDCD model" with published data *reproduced* the relative sampling variances calculated for the standard "mineral-like matrices", but more importantly *corrected* the relative sampling variance calculated for real contaminants by several orders of magnitudes. In all cases, the recalculated relative sampling variances were *decreased* to below their corresponding experimental measurements, now fully as expected from TOS, substantiating our development.

1. Introduction

There exists a long tradition for enthusiastic use of the famous "Gy's formula", aka "Gy's equation", in several important sectors of technology and industry, dominantly within the exploration, mining, minerals processing and waste management sectors. This popularity is understandable as this equation purports to relate the Fundamental Sampling Error, $s^2(FSE)$, to the effective *heterogeneity* of the material to be sampled, on the basis of which Gy managed a monumental theoretical breakthrough in deriving a first approximative relationship between the FSE sampling error as a function of a set of *in-principle* measurable physical and chemical material parameters. This was quickly welcomed by all manner of practically minded samplers because of its alluring promise to be able to determine an operable minimum sample mass that in some way, under a set of specific assumptions, would appear to honour TOS′ objective demands for representativity. In the industry sectors that first embraced Gy's formula this formalism met with a lot of practical success.

However, through its use over many decades in several other sectors, the critical underlying assumptions were not always at the forefront of the samplers' minds, but were in fact often ignored (as shall be traced in some detail below) in the view of the obviously desirable holy grail: Estimate four or five material parameters, plug them into Gy's equation, and obtain a minimum sample mass which, with reference to the published formula, allows one to *claim* sampling representativity.

Alas, this is a road filled with many dangers, especially when Gy's *mandate* that all Incorrect Sampling Errors (ISE) first must have been eliminated before even beginning to put estimated numbers in the formula ([1] p.52-53[2], [2]); one must also be wary of the crucial random distributional heterogeneity assumptions regarding Gy's formula (treated more in full below).

Nevertheless, agreement between calculated $s^2(FSE)$ estimates and measured experimental values of the relative sampling variance has been broadly satisfactory for many types of ore, some *similar* rock types and aggregate mixtures, as well as for certain *artificial samples* with an aim to *simulate* natural material types, e.g. Refs. [3,4,5] [however these latter may show more or less correspondence with most of the real-world sampling targets in science, technology and industry].

Specifically, it has recently surfaced that agreement is distinctly poor for the specific case of *contaminated soils* [6,7]. Typically, predicted theoretical relative sampling variances can overestimate experimental relative sampling variances by several orders of magnitude when sampling typical contaminated soils encountered in environmental science and monitoring [6,7], *indicating* that the assumptions behind the use of Gy's equation for $s^2(FSE)$ (see eq. [5]) are perhaps not applicable for the constitution of this type of compound material. This point is further reinforced below by reporting other such usages, and misusages, of the classical equation.

The present thrust takes its point of departure in the realization that in contaminated soils, the heterogeneity carried by the contaminants both comes from *intrinsic* contaminants, i.e. contaminant particles either liberated or embedded in neutral matrix particles (i.e. soil and waste particles) and from *extrinsic* contaminants, i.e. contaminants concentrated at the surface of neutral matrix particles, for example caused by reactions with solutions percolating through the soil. This latter situation constitutes a distinct departure from the traditional assumptions behind the use of Gy's equation.

This forces us to outline the different nature of contaminated soils in relation to the more 'traditional' particulate material classes and types for which this famous equation has met with notable success.[3] It is also possible that the material class presently in focus, i.e. contaminated soils and related mixed materials, can be extended to other *similar* materials, to be delineated further below.

The objectives of the present paper are therefore to:

1) Investigate the source of the discrepancies between theoretical and experimental relative variances for contaminated soil and *similar* materials;
2) Examine the relevance of the current practical equation and its parameters for the constitutional heterogeneity of contaminated soils and *similar* materials;
3) Derive and validate an equation of constitutional heterogeneity which is better applicable to contaminated soils and *similar materials*. It is imperative that this equation shall be easy to implement in practice.

[2] After a long career, Gy in private expressed his displeasure with too many applications in which this mandate was not complied with. He was pleased that his name would be associated with many compliant applications - but distinctly unhappy with "all too many" incorrect usages (pers. com. 2005 (KHE)).

[3] The present authors acknowledge the extensive discussion that has taken place within the TOS community on the conditions and limitations of applicability of Gy's equation, in particular the role played by the liberation parameter, which has undergone a tortuous journey of evolving interpretations since Gy's original definition; see Minnitt [8] and François-Bongarçon [9,10] for the most recent overviews. In the latter François-Bongarçon present a historical survey of recent treatments of both the shape factor and the granulometric factor in addition to the liberation factor; these factors are also commented upon in similar context by Minkkinen and Esbensen [11].

2. The "mineral model" – successes and shortcomings

In order to understand the possible shortcomings of the current model of constitution heterogeneity and its relative sampling variance when it is used in the sampling of contaminated soil, and mixed materials, it is necessary to briefly review a few salient principles of Gy's Theory of Sampling (TOS).

Herein, the relative variance of the Fundamental Sampling Error, FSE, is defined as

$$s^2(FSE) = (1/m_s - 1/m_L)HI_L \quad [1]$$

where.

$s^2(FSE)$ = relative variance due to the FSE (dimensionless)
HI_L = Heterogeneity Invariant of the lot material to be sampled (g)
m_S = sample mass (g)
m_L = mass of the lot (g)

Basically, the general form of the lot Heterogeneity Invariant, HI_L, is defined by Gy [12] for a population of lot fragments, F_i, as

$$HI_L = \sum_i \frac{(a_i - a_L)^2}{a_L^2}\frac{m_i^2}{m_L} \quad [2]$$

i.e. the mass-weighted sum of concentration heterogeneities carried by each fragment, F_i, of the lot and where:

a_i = critical content of the analyte of interest in each fragment F_i (dimensionless);
a_L = critical content in the lot of the analyte of interest (dimensionless);

m_i = mass of each fragment F_i [M];

m_L = mass of the lot [M].

From Equation [2], it can be seen that the constitution heterogeneity carried by fragment F_i increases with its compositional deviations $(a_i - a_L)$ and with relative fragment mass. Conversely, a lot can only be constitutionally homogeneous if all fragments, which all have positive mass, carry the exact same analyte grade *equal to* a_L, i.e. all $(a_i - a_L) = 0$. As is well acknowledged, there does not exist homogenous materials in this definition in the world of technology and industry, see e g. DS 3077 [13].

Pitard [14,15] developed Equation [2] further, by considering that the critical content of a fragment in general is more *correlated* with its density than with its size. Therefore, following a set of reasonable approximations, Pitard [14,15] defines the lot Heterogeneity Invariant as

$$HI_L = \sum_\alpha v_\alpha \sum_\beta \rho_\beta \frac{(a_{\alpha\beta} - a_L)^2}{a_L^2}\frac{m_{L_{\alpha\beta}}}{m_L} = \left(\sum_\alpha v_\alpha \frac{m_{L_\alpha}}{m_L}\right) \cdot \left(\sum_\beta \rho_\beta \frac{(a_\beta - a_L)^2}{a_L^2}\frac{m_{L_\beta}}{m_L}\right) = X \cdot Y \quad [3]$$

where.

v_α = volume of the average fragment $F_{\alpha\beta}$ of size fraction α [L^{-3}];
ρ_β = density of fragment $F_{\alpha\beta}$ [M L^{-3}];
$a_{\alpha\beta}$ = critical content of density fraction β, of size fraction, α (dimensionless);
a_β = critical content of density fraction β (dimensionless);
$m_{L_{\alpha\beta}}$ = mass of density fraction β, of size fraction, α [M];
m_{L_α} = mass of size fraction α [M];
m_{L_β} = mass of density fraction β [M];
X = heterogeneity carried by the *ensemble* of *differentiated* size fractions (Fig. 1a–b);
Y = heterogeneity carried by the *ensemble* of *differentiated* density fractions (Fig. 1c).

After further approximations intending to honor reality w.r.t. *most* aggregate material classes (see Refs. [1,12,14–19]), Equation [3] leads to the well-known form of HI_L, which is practically implementable in principle - *provided* its parameters are all relevant and amendable to estimation, or can be measured directly without undue difficulty, namely:

$$HI_L = clfgd_N^3 \quad [4]$$

where.

d_N = nominal particle size at 95% passing [L];
f = particle shape factor (dimensionless; $0 < f < 1$);
g = granulometric factor (dimensionless; $0 < g < 1$);
c = mineralogical factor = $\rho_M \frac{(1-a_L)^2}{a_L} + \rho_g(1-a_L)$ [M L^{-3}]
ρ_M = density of the pure mineral (the 'analyte', subscript "M") [M L^{-3}]
ρ_g = density of the gangue (subscript "g") [M L^{-3}]
l = liberation factor (dimensionless; $0 < l < 1$)

Equation [4] will be referred in the present treatment as the "mineral model". Note that Equation [4] describes a *binary* density-class based model, since it describes the heterogeneity *carried by the analyte* as being correlated with two density phases, the analyte-carrying mineral and the gangue.

It is important that the material parameters *f, g, c* and *l* in Equation [4] can be assumed to be *meaningful averages* in their role of describing real-world particulate matter. In Gy's work, and that of others (e.g. Refs. [12,14,15], scientifically-based expressions are provided for these parameters; also treated in Minkkinen and Esbensen [11]. However, practical use of Gy's equation has often been performed by directly plugging-in 'typical values' rather than by characterizing them using these basic expressions. This has often led to inappropriate use of Gy's equation for certain important material classes as described later. Therefore, using standard *proxy values* for *f* or *g* will often be debatable, for example: how well will a grain size distribution be adequately represented by a sorting or *size range factor*, *g*, or: are all fragment shapes indeed describable by a singular *particle shape factor*, *f*. Nevertheless, this is very nearly always *assumed* to be realistic with standard $g = 0.25$ and $f = 0.5$ values, which is one of the dissatisfactions of the present authors and others [20].

The *mineralogical composition factor*, *c*, has units of (but not the meaning of) specific gravity (g cm^{-3}), and takes the average grade

and densities of the two component phases into account. Factor c is to be understood as the density of the 'mineral of interest' (i.e. the phase carrying the analyte) divided by the grade - so c becomes larger as the average grade of the material decreases – the sampling variance increases the lower the analyte grade.

The *liberation factor, l*, was originally defined as the grain size at which embedded analyte particles were liberated upon comminution. It was later defined as $l = (d_l/d_N)^{0.5}$, where d_l is the analyte liberation diameter, and which varies between 0 (for totally non-liberated fragments) and 1 (for completely liberated fragments); see historical overviews in François-Bongarçon & Gy [20]; François-Bongarçon [9,10].

As shall be clear below, the differing interpretation as to the meaning and validity of the liberation factor, l, is of particular importance for the realism one can ascribe to a variance estimate calculated by Gy's formula. It is of considerable interest that Pierre Gy was of the opinion that far too much credence has been put on this formula by many direct calculations based on *standard values* for f, g, c and l. Gy himself stated vigorously that this formula was but a very first attempt to establish a quantitative relationship between lot materials properties and the related sampling variance, *were* all other sampling errors *completely* eliminated (strong emphasis on the conditional). Indeed, Gy was adamant that the formula often was subjected to "simplistic" quantifications with a debilitating effect vis-à-vis realism (see footnote 2). It is worth mentioning that Gy emphasized that even under the strongest adherence to its assumptions, calculation of $s^2(FSE)$ should clearly be understood as but an *order-of-magnitude* estimate only [11,21].

Against the background of these qualifications, Equation [4] constitutes the classical "mineral model" of constitution heterogeneity, because it explicitly considers the critical element, the analyte, either as fully liberated (e.g. gold-quartz mineralizations) or as embedded in one, or more, analyte-enriched particle types. This *implies* that the heterogeneity carried by different empirical size fractions will be *dominated* by the *largest* fragments and that the heterogeneity carried by the density fractions is maximal when $l = 1$, meaning that the heterogeneity carried by the separate density fractions is maximal, or $c = Y_{max}$. This is so because the mineralogical factor, c, is defined as the heterogeneity of two completely liberated density fractions, i.e. the pure mineral of interest *versus* the gangue. Then the liberation factor, l, defines the degree to which these two phases are actually liberated *from each other*. Only when the liberation factor is exactly equal to 1 is the analyte *available* to be sampled *independently*.

In practice, the liberation factor is known to be difficult and very laborious to measure or estimate with realism and validity, which explains the rare application areas in which it has been found 'profitable' to do so [6,15,18,22–24].

Combining Equations [1] and [4] leads to the historically most popular version of Gy's formula:

$$s^2(FSE) = \left(\frac{1}{m_s} - \frac{1}{m_L}\right) clfgd_N^3 \approx \frac{clfgd_N^3}{m_s} = \frac{Kd_N^3}{m_s} \qquad [5]$$

which can be found in many standard TOS references, for application to the sampling of ores in particular but also for other, *similar* materials. The parameter K is known as the "sampling constant" allowing characterisation of the ease of sampling of specific materials with different compositional makeups. The sampling constant is specific for a given state of comminution, i.e. for a given top nominal grain size of the mineral of interest [12,15,22–24]. Estimates of $s^2(FSE)$ are consequently only valid for a specific comminution stage or state – and for the specific test material investigated, Minkkinen and Esbensen [11].

The historical record shows that Gy's formula has been found particularly useful in gold exploration, mining and minerals processing and within a few other application areas (other ores of broadly similar granular makeup). The excessive efforts needed when trying to apply it to more complex matrices has very often been felt prohibitive, i.e. the more complex the heterogeneity of the material the less applicability.

Equation [5] was e.g. used by Gerlach and Nocerino [4] to develop a rationale for sampling *environmental matrices* (including contaminated soil), where it was successful in estimating the fundamental sampling variance for simple two – or three-component artificial granular matrices (such as those in Refs. [3,4], i.e. quartz sand with NaCl and magnetite particles. Application to such matrices (which by design are composed of fully liberated density phases and which are constitutionally similar to the mineral-like material described by the mineralogical factor, c, in Equation [4]) are understandably successful. Recently an extension to this kind of laboratory experiments was presented by Svensmark [5] (who treats extensions to Gy's equation in a completely different way than here, summing classes with globally similar properties).

However, Equation [5] failed to provide meaningful estimates of $s^2(FSE)$ for real-world soils contaminated with trace metals, as reported in Boudreault et al. [6] and Dubé et al. [7]. In these studies, Equation [5] provided estimates of $s^2(FSE)$ which were much larger than the experimental estimates of sampling variance, an impossible situation *per* TOS, as $s^2(FSE)$ is defined as the minimum irreducible variance caused by the constitutional heterogeneity of a lot in which the analyte is *ideally* randomly distributed. In reality however, even if all incorrect sampling errors are diligently eliminated, there always exist some residual pattern or form of segregation within the lot which will cause the sampling variance to be greater than $s^2(FSE)$. This aspect is discussed further below.

The latter authors have examined the influence of the liberation factor, l, on these disagreements. They suggested that the rule-of-thumb $l = 1$ when there is no satisfactory model available for l, as proposed by Gy [16], is actually *detrimental* to the applicability of Equation [5] to sampling of contaminated soil and similar materials. Unfortunately, this rule is currently more or less *formalized* in USEPA guidelines [4]; p. 44) under the sweeping justification that "for most pollutants, [$l = 1$] *should* be acceptable, as the methods are for total analyte".[4]

However this assumption is *misleading*, because l does not connote availability of the analyte to chemical *extraction* but specifically to the degree to which individual analyte particles are physically susceptible of being selected i.e. *sampled* independently from non-analyte particles. Based on experimental evidence, Boudreault et al. [6] and Dubé et al. [7] suggested that the model of complete liberation of a mineral from its gangue may in fact not be *relevant* for the heterogeneity of trace metal contents in contaminated soil. Minkkinen and Esbensen [11] also suggest that the liberation factor is not only a measure of liberation but also of segregation within the lot.

Boudreault et al. [6] also investigated the use of the model for l presented by Pitard [15]:

$$l = (d_l/d_N)^b \qquad [6]$$

where,

d_l = analyte liberation diameter [L];

[4] Emphasis on the conditional by present authors.

b = empirical exponent.

Equation [6] implies that after material comminution to the "liberation limit" at $d_N = d_l$, the mineral of interest (i.e. the analyte) is now supposed to be completely liberated from the gangue. Boudreault et al. [6] and Dubé et al. [7] have stressed that the problem with the use of Equation [6] when sampling contaminated soils lies in the determination of d_l and b as there is a lack of available data on these parameters for such a diverse class of particulate matter. It may even be that these concepts are not relevant for this, and *similar* cases - as is further explored below.

2.1. The need for a different formula for specific material systems

The liberation factor can alternatively be formulated directly from Equation [3] following another set of assumptions, see Pitard [15]; to obtain:

$$l = \frac{a_{max} - a_L}{1 - a_L} \quad [7]$$

where a_{max} = critical content of the largest fragments (dimensionless).

Therefore, a significantly large difference between the critical content of the largest fragments and the average content of the lot *implies* that the analyte is more located in the set of large fragments. Again, the hypothesis of a mineral-like analyte occurrence may well be inconsistent with the state of the analyte in matrices like soil and similar, where the analyte may be, fully or partially, distributed both "mineralogically" but also in a *different manner*, e.g. as *coatings* on otherwise inert gangue particles.

In point of fact, in contaminated soils, the observation is prevalent that the largest grainsize fractions can have a *smaller* analyte content than the smaller grainsize fractions due to their smaller particular *surface area*. This would lead to a situation where $a_{max} < a_L$ and thus $l < 0$, which violates all assumptions leading to Equation [7].

Therefore, due to such difficulties in estimating l, instead of using $Y = cl$, one alternative would be to use $Y = \left(\sum_{\beta} \rho_{\beta} \frac{(a_{\beta} - a_L)^2}{a_L^2} \frac{m_{L_{\beta}}}{m_L} \right)$ directly in Equation [3]. This would require identification of each contaminant-bearing phase, its corresponding average content and density. Using the technique of QEMSCAN, Lyman and Schouwstra [26] have successfully performed exactly this on rocks with minerals with elemental grades between 3 and 40% (i.e. 30 000 and 400 000 mg kg^{-1}).

However, pollutant-bearing minerals and chemical complexes are very much more diverse and perhaps also differently distributed in contaminated soil [27,28] and critically, usually occur at much smaller grades, often close to or below the detection limit relevant for the standard mineralogical techniques. Therefore, calculating Y from Equation [3] may not be practical for solving problems in sampling of contaminated soils and similar mixed materials matrices, because of technical limitations in the identification and quantification of the relevant geochemical phases. Again the work involved would fast become prohibitive for a system as complex as soil.

Based on such apparent shortcomings, we venture below to revisit the development of a variant of Gy's equation to be used on material systems with contaminated soils used as a starting point.

3. Modelling constitutional heterogeneity using size-density classes

To circumvent the above shortcomings and difficulties, we consider that the analyte, i.e. the contaminant of interest, may not only be present as a set of, or embedded in, more-or-less enriched particles, but may also be present on the *surface* of soil particles through common *retention mechanisms* such as ion exchange, adsorption, precipitation or complexation [27,28]; see also the thought-experiment presented in Svensmark [5] in which PAHs are considered sorbed onto organic matter (clearly very fine grained), which in turn was very unevenly distributed in the pertinent soil. Importantly, the retention capacity of soil particles is *proportional* to their surface area and *inversely proportional* to their size. Therefore, the standard approximation that the critical content of a fragment is more strongly correlated to its density than to its size will manifestly have to be *revised* for matrices such as contaminated soils and aggregates of coated mixtures, i.e. the critical content of a soil particle shall have to be considered to be correlated with both particle size and density. A diagrammatic framework for the conceptual formula derivation described below, is given in Fig. 1a–c. The authors owe a debt of gratitude to Lyman [29] for an early inspired physical illustration on the conceptual decomposition into size and density classes (which may make it easier to contemplate the physical meanings of the many single and double-summations presented below).

In order to obtain a mathematical expression capable of describing all cases of constitutional heterogeneity illustrated by Fig. 1b and c, Equation [3] shall first be rewritten as

$$HI_L = \sum_{\alpha} v_{\alpha} \sum_{\beta} \rho_{\alpha\beta} \frac{(a_{\alpha\beta} - a_L)^2}{a_L^2} \frac{m_{L_{\alpha\beta}}}{m_L} = \sum_{\alpha} v_{\alpha} \frac{m_{L_{\alpha}}}{m_L} \sum_{\beta} \rho_{\alpha\beta} \frac{(a_{\alpha\beta} - a_L)^2}{a_L^2} \frac{m_{L_{\alpha\beta}}}{m_{L_{\alpha}}} \quad [8]$$

where $\rho_{\alpha\beta}$ = density of fragment $F_{\alpha\beta}$ of density fraction β from size fraction α. Also, indices L_{α} and $L_{\alpha\beta}$ represent size fraction α and density fraction β of size fraction α respectively.

Equation [8] constitutes a model of constitutional heterogeneity based on size-density classes; it is equivalent to the model developed in Lyman and Schouwstra [26]. Equation [8] and its variants, Equations [9] to [13], will be referred to herein collectively as the size-density-class derived model or the SDCD model for conciseness.

The application of Equation [8] can be rationalized by considering three general cases, as illustrated in Fig. 1b and c.

The first two are partially idealized, and not often likely to be fully realistic in real-world matrices, but they serve well for verification of the proposed "SDCD model" and the development of methods for its practical implementation. The third case represents a reasonable hypothesis about the state of many analytes occurring in real-world contaminated soils or aggregates of coated particulate mixtures, e.g. precipitated or adsorbed metals, as well as adsorbed organic compounds, amongst others. The heterogeneity invariant for each case can be derived from a specific variant of Equation [8]. It needs emphasising that Equation [8] and its variants developed in the following sections, describe the heterogeneity carried by solid-phase analytes. While in contaminated soil, as well as possibly in other matrices, there can also exist liquid-phase, or gaseous-phase analytes, Equation [8] and its variants do not take these into account. The SDCD model specifically extends the expression of the heterogeneity carried by solid-phase analytes to analyte-coated particulate

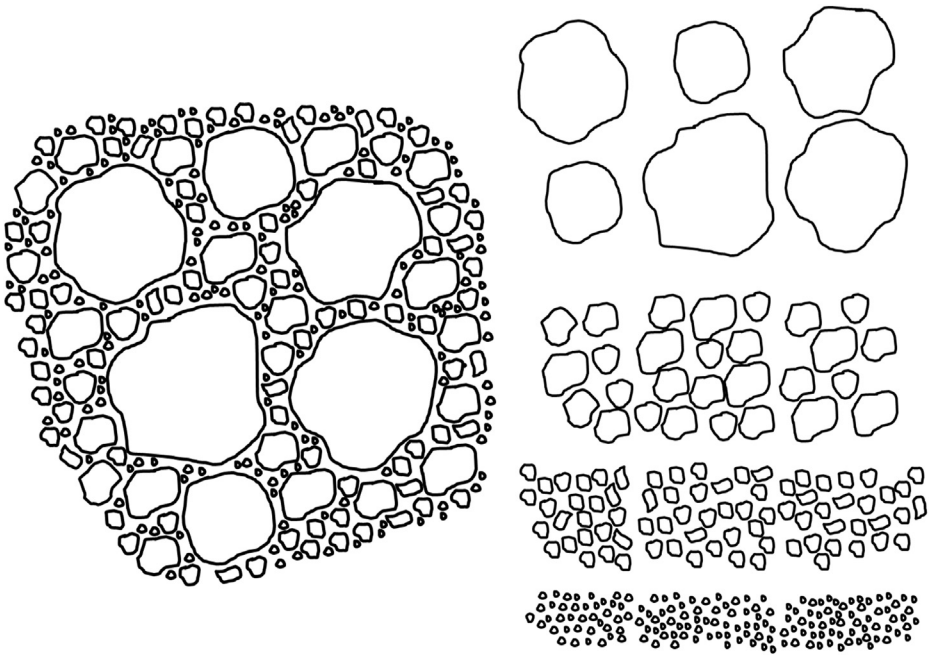


Fig. 1a. A simplified soil model as a mixture of distinct granular classes of fragments of various size and mineral composition (**left**) and sorted into size fractions (**right**). The density of naturally occurring soil particles is between 2.6 and 2.8 g cm^{-3} as usually assumed, which is indeed the range of density of the Earth's crust from which the soils are formed (e.g. Refs. [30,31]. Therefore, this basic soil model does not have density fractions, only size fractions.

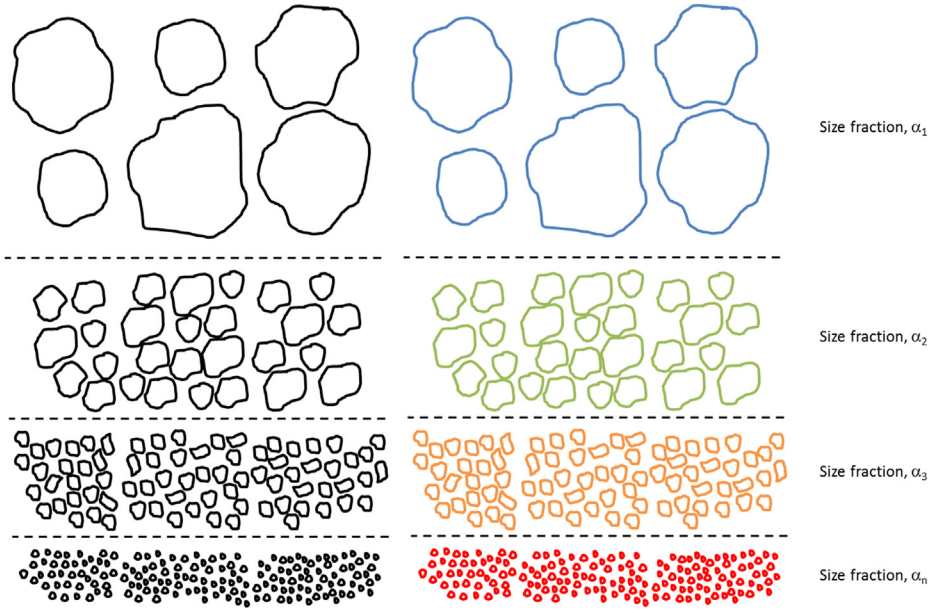


Fig. 1b. Same granular soil model as in Fig. 1a. **Left:** Heterogeneity carried by the fragments as a function of their size. If a soil is sampled for the determination of its particle size distribution or the quantification of a specific size fraction, this schematic representation of constitutional heterogeneity is sufficient. It is mathematically described by Equations [9] and [10]; **Right:** Heterogeneity carried by the size fraction as well as an analyte present as a coating (color) on the soil particles. The 'warmer' the colors (from blue to red), the larger is the analyte content in a given size fraction. In this representation, analyte content is correlated with fragment size. If this soil is sampled for the determination of its analyte content, this schematic representation of constitutional heterogeneity is appropriate. It is mathematically described by Equation [12], where a reasonable approximation is that the density of soil particles is not changed significantly by analyte coating.

materials or mixed analyte-enriched and analyte-coated particulate materials only. It will need to be further developed to be extended to liquid-phase and gaseous-phase analytes, if/where/when relevant. Also, to be absolutely clear, the terms "coating" and "analyte-coated" refer herein to any form of concentration of the analyte at the surface of soil particles.

3.1. Case 1: analyte totally contained in mineral or organic coatings on soil particles

This case is schematically represented by Fig. 1b. Examples of analyte-coated particles would be the cases of fine gold particles coating fragments of crushed material presented in Pitard [15];

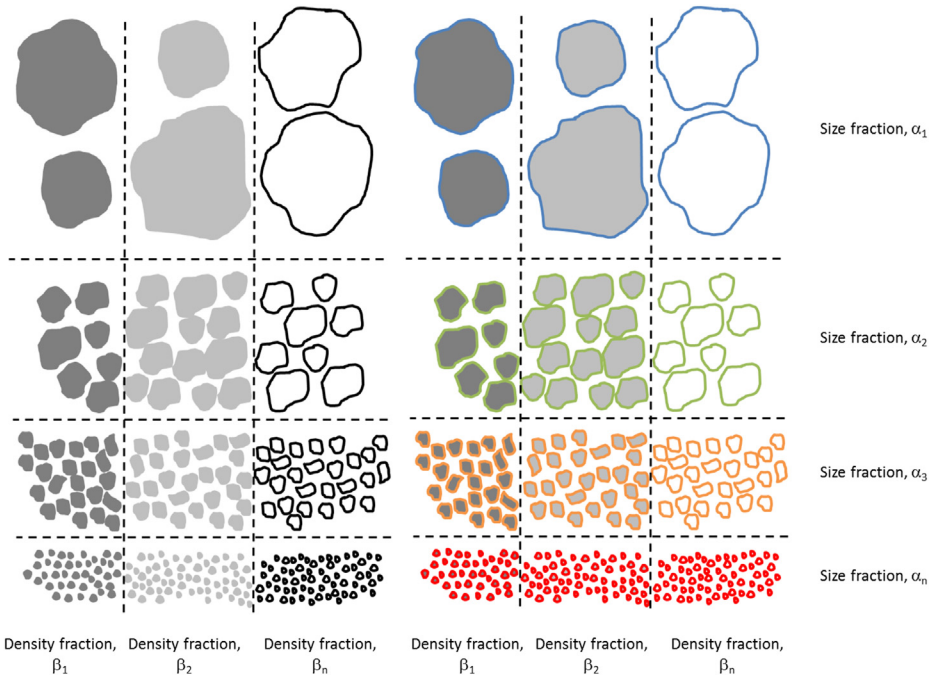


Fig. 1c. Same basic soil model as in Fig. 1a. **Left:** Particles differ in density. Increasing grey tone darkness reflects increasing density. In a soil, particle density will mostly change due to the presence of waste particles (e.g. metals scraps, concrete, brick, slag, clinker, ash). This schematic representation shows heterogeneity carried both by the size fractions and an analyte contained in the density fractions. When analyte content varies across the density fractions, but not across the size fractions, this is mathematically described by Equations [3] or [4]. **Right:** This schematic representation illustrates the same soil matrix as in the left figure, but with added analyte coatings (identical to the right-hand illustration of Fig. 1b. Therefore, it illustrates the heterogeneity carried by size fractions by an analyte, the content of which varies across both density and size fractions. This case represents the "size-density-class derived model" developed here; it is mathematically described by Equation [8] or [11].

aflatoxin-contaminated pistachio nuts in Lyn et al. [32]; and salt-coated sand particles in Gerlach et al. [4]. These are only a very few published examples of coated particulate matter that have been subjected to sampling and/or $s^2(FSE)$ estimation.

In such cases, $\rho_{\alpha\beta} = \rho_\alpha$ since it can be assumed that the coating does not modify significantly the density of the particles, i.e. for analyte contents approximately below 1%, as investigated by the thought-experiment in Appendix A. In contaminated soil, pollution criteria usually correspond to concentrations lower than 1%, or 10 000 mg kg^{-1} of a single contaminant. Therefore, instances of soil being considered contaminated at concentrations below 1% are very common.

Thus an estimate of HI_L could in this case be made based solely on the distribution of particle sizes as illustrated on the left of Fig. 1b. It is then critical to obtain a representative sample of all size fractions to make a relevant and reliable measurement of the analyte content of each size fraction [15]. Moreover, the critical content of size fraction L_α, namely a_α, becomes the content of fragments of that size fraction in itself, which is 1 by definition [15]. Therefore, Equation [8] can be reduced and rewritten as follows,

$$HI_L = \sum_\alpha v_\alpha \frac{m_{L_\alpha}}{m_L} \rho_\alpha \frac{(a_\alpha - a_L)^2}{a_L^2} = \sum_\alpha v_\alpha a_{L\alpha} \rho_\alpha \frac{(1 - a_{L\alpha})^2}{a_{L\alpha}^2} = \sum_\alpha v_\alpha \rho_\alpha \frac{(1 - a_{L\alpha})^2}{a_{L\alpha}} \quad [9]$$

where $a_{L\alpha}$ is the content of fragments of size fraction L_α in the lot and ρ_α is the density of the coated particles in size fraction L_α. Equation [9] can be further simplified, by considering HI_L as the dichotomous sum of i) the constitutional heterogeneity of a specific size fraction of interest and ii) the complement of the particle size distribution, namely

$$HI_L = \sum_\alpha v_\alpha \rho_\alpha \frac{(1 - a_{L\alpha})^2}{a_{L\alpha}} = f\rho \left[d_c^3 \frac{(1 - a_{Lc})^2}{a_{Lc}} + \sum_\alpha d_\alpha^3 a_{L\alpha} \right] \quad [10]$$

Moreover, Pitard [15] assumes that if $a_{Lc} < 0.05$, i.e. the size fraction of interest corresponds to the top nominal particle size, or larger, then $\sum_\alpha d_\alpha^3 a_{L\alpha} \approx 0$ in Equation [10].

However, if the material is well graded and can be effectively screened into several size fractions, the analyte content can be measured in each size fraction. The resulting distribution of the analyte content as a function of particle size can then be used to provide a more accurate estimate of HI_L using Equation [11] or [12], as illustrated on the right of Fig. 1b and presented in section 3.3 below, since it would account for the constitution heterogeneity carried by the analyte. This latter HI_L estimate would have to be smaller than that obtained from Equation [9] because $\frac{m_{L_\alpha}}{m_L} \frac{(a_\alpha - a_L)^2}{a_L^2} < \frac{(1 - a_{L\alpha})^2}{a_{L\alpha}}$.

3.2. Case 2: analyte exclusively contained in liberated enriched particles

In this classical case, the analyte solely occurs as identifiable and quantifiable liberated enriched particles, as illustrated in the left part of Fig. 1c. Examples of liberated and enriched particles found in the literature are the cases of the mixtures of sand, salt and magnetite in Gerlach et al. [3,4] or the mixture of steel microspheres and crushed stone in Sona and Dubé [33]. One can therefore use Equation [3] or [4], which are now particular cases of Equation [8], and in which each type of analyte-enriched particles can be considered as a distinct density fraction for each size fraction in which it is found.

3.3. Analyte contained in coated and enriched particles

In this more realistic case, the assumption is that the analyte is present as a mixture of different mineral and organic phases, some enriched in the analyte (classical case) and some as coating on soil particles, the latter coated to any degree from 0 to 100%. This is schematically represented by the illustration of the right of Fig. 1c. If each phase, coating or enriched, can be identified and quantified, Equation [8] can again be used as in the preceding case, at least in principle. However, technical limitations, or economic constraints, will assuredly very often hamper such a detailed characterization in practice. Moreover, in environmental sciences and engineering, samples are commonly analyzed for their total content of an element or an organic compound, e.g. Pb or benzo(a)pyrene. Equation [8] must then be simplified for its use in this context, wiz. that the total analyte content and density are now averaged over each size fraction. Thus,

$$HI_L = \sum_\alpha v_\alpha \frac{m_{L_\alpha}}{m_L} \sum_\beta \rho_{\alpha\beta} \frac{(a_{\alpha\beta} - a_L)^2}{a_L^2} \frac{m_{L_{\alpha\beta}}}{m_{L_\alpha}} = \sum_\alpha v_\alpha \frac{m_{L_\alpha}}{m_L} \rho_\alpha \frac{(a_\alpha - a_L)^2}{a_L^2}$$
$$= f \sum_\alpha d_\alpha^3 \frac{m_{L_\alpha}}{m_L} \rho_\alpha \frac{(a_\alpha - a_L)^2}{a_L^2} \quad [11]$$

where a_α is the total analyte content in size fraction L_α.

If density does not vary amongst size fractions, such that $\rho_\alpha = \rho$, Equation [11] reduces to

$$HI_L = f \rho \sum_\alpha d_\alpha^3 \frac{m_{L_\alpha}}{m_L} \frac{(a_\alpha - a_L)^2}{a_L^2} \quad [12]$$

The last part of Equation [11] and Equation [12] were originally obtained by Gy (e.g. [12]). However, Equation [12], which is especially useful with materials such as contaminated soil, was subsequently dismissed as inapplicable to "the metal, mining and processing industries" "due to an unusual density contrast between the components" in matrices encountered in these fields [12]; p. 79). It may have been this early dismissal which caused Equation [12] to be almost totally forgotten and left unused in other fields.

In this last HI_L expression for the specific classes of materials treated here, the general shape factor, f, remains in the form used in the Gy's classical formula. Therefore, the various issues related to the shape factor remain in Equation [12]. Gerlach and Nocerino [25] indicate that "the majority of particulate samples have shape factors from 0.3 to 0.5" and that hazardous waste particles have shape factors closer to 0.5 but there is little empirical evidence available; this kind of blanket argumentation does not sit well with the present authors.

It would be possible to extend Equation [11] to include an explicit definition of the shape factor for the average particle of each size-density class, namely $f_\alpha = m_{p\alpha}/(\rho_\alpha d_\alpha^3)$ [15], where $m_{p\alpha}$ is the mass of the average particle of size class L_α, so that

$$HI_L = \sum_\alpha f_\alpha d_\alpha^3 \frac{m_{L_\alpha}}{m_L} \rho_\alpha \frac{(a_\alpha - a_L)^2}{a_L^2} = \sum_\alpha \frac{m_{p\alpha}}{\rho_\alpha d_\alpha^3} d_\alpha^3 \frac{m_{L_\alpha}}{m_L} \rho_\alpha \frac{(a_\alpha - a_L)^2}{a_L^2}$$
$$= \sum_\alpha m_{p\alpha} \frac{m_{L_\alpha}}{m_L} \frac{(a_\alpha - a_L)^2}{a_L^2} \quad [13]$$

Characterizing $m_{p\alpha}$ for each size class would be time consuming, but it could conceivably be achievable in most material science laboratories. However, some published studies used below to validate the use Equation [8] and its variants for contaminated soils and similar materials do not provide $m_{p\alpha}$. Lyn et al. [32] did calculate a shape factor specific to their studied material, but the other studies simply used a 'typical value' for f. Therefore, Equation [12] was used with the data in the published studies presented below with the aforementioned typical values for f rather than Equation [13].

4. Materials and method

4.1. Validation of the "SDCD model" using data from Lyn et al. [32] and Gerlach et al. [3,4]

In Lyn et al. [32], the uncertainty in sampling pistachio nuts for the determination of aflatoxin concentration was estimated by both theoretical and empirical means. The theoretical sampling uncertainty was estimated by modelling the heterogeneity invariant of the nuts using Equation [4]. The parameters used were as follows [32]: m_s = 250 g, ρ = 0.9951 g cm^{-3}, a_L = 8.6 x 10^{-10}, c = 288.3 g cm^{-3}, l = 1, f = 0.27, g = 0.75, d_N = 2 cm. Empirical estimation of the sampling uncertainty was made using duplicate sampling, measurement of aflatoxin concentrations and robust analysis of variance (RANOVA). The authors estimated sampling, analysis and measurement uncertainties respectively.

In Gerlach et al. [3], artificial soil, made up by layering sand, salt and magnetite particles, was sampled using five different mass reduction techniques: riffle splitting, paper cone riffle splitting, fractional shoveling, coning & quartering, and grab sampling. These techniques were ranked based on their bias and precision. The latter was compared to relative standard deviation (RSD) values obtained from $s^2(FSE)$ values calculated with the "mineral model", i.e. Equation [5].

In Gerlach et al. [4], three studies were conducted. The first compared the bias and precision of incremental sampling and rotary sectorial splitting used to sample an artificial soil made of a mixture of sand and salt particles. The second study was designed to examine the effect of large particles on bias and precision of rotary sectorial splitting by using artificial samples made of sand, salt and coarse sandstone particles, while the third was designed to examine the constitution heterogeneity of particles coated with salt. In these studies, the experimental RSDs were compared to those obtained from $s^2(FSE)$ values calculated with the "mineral model", i.e. Equation [5], for the first two studies, and with Equations [1] and [2] for the third study.

In the present study, variants of Equation [8] are used with Equation [1] to recalculate $s^2(FSE)$ and corresponding RSD values obtained by Lyn et al. [32] (case 1 variant) and Gerlach et al. [3,4] (cases 2 and 3 variants).

4.2. Comparing the "mineral model" and the "SDCD model" for real-world contaminated soils

The agreement between theoretical sampling variances calculated using applicable variants of Equation [8] (with Equation [1]) and experimentally determined sampling variances was investigated using existing data from three published studies, namely Boudreault et al. [6] and Dubé et al. [7,34]. Details about these lots and their materials, sampling protocols, and analytical methods are provided in the aforementioned studies. A brief description of each will suffice below.

4.2.1. Boudreault et al. [6] and Dubé et al. [7] studies

These studies were conducted using samples from an urban brownfield consisting of a surface fill of waste fragments (bricks, concrete fragments, clinker, ash, metal scraps) mixed with remolded soil-type material to a depth of 2–3 m, overlying an uncontaminated till deposit [6,34]. The water table was at an average depth of 2–3 m, thus just underneath the fill. Previous site assessment studies confirmed that the fill was contaminated with several trace metals (Cd, Co, Cr, Cu, Mn, Ni, Pb, and Zn), as well as PAHs (e.g. chrysene, phenanthrene, benzo(a)pyrene). The aforementioned studies focused on sampling the fill with the purpose of determining the uncertainty associated with estimating trace metal concentrations in the soil.

Fig. 2a–b shows the brownfield site and the sampling stations where samples were taken for Boudreault et al. [6] and Dubé et al. [7] studies. Fig. 2a shows the locations of 32 sampling stations along a systematic sampling plan with a grid spacing of 6 m × 7 m.

Fig. 2b shows 6 supplemental sampling stations intercalated between stations of the systematic grid, and which were used for assessing short-range variability.

Each station was sampled at two depths. Therefore, a set of 64 samples, from the systematic grid, and another of 12 samples, from the short-range stations, were used for a study on the assessment and control of uncertainty in estimating concentrations at the field scale using conditional simulations [35]. Moreover, the set of 12 samples was used for the studies by Boudreault et al. [6] and Dubé et al. [7]. Finally, from the set of 64 samples, 9 samples were randomly selected for the present study in order to determine characteristic HI_L values for this brownfield using Equation [12] (see section 4.2.3).

4.2.1.1. Sampling protocols and data analysis in Boudreault et al. [6].

In Boudreault et al. [6], two protocols for contaminated soil sampling and analysis were compared based on measured contaminant concentrations and their variance, namely grab sampling and alternative sampling procedures (i.e. GSP and ASP). Both protocols involved subsampling steps from the field to the laboratory. Details of both protocols can be found in Boudreault et al. [6].

For the present purposes, only the results of the laboratory sampling stage in Boudreault et al. [6] are used, i.e. the specific field sampling issues and their associated uncertainties were excluded in order to present a clear experimental design, without loss of generalisation. Therefore, the comparison between theoretical and experimental relative sampling variances below takes its point of departure at the laboratory sampling stages.

For the ASP, 96 500-g tertiary field samples were produced from the 12 primary field samples and were brought to the laboratory for further subsampling. These were first dried at 60 °C and sieved at 1 cm. Each was subsequently ground to a nominal particle size of 1 mm, and divided into eight 50-g subsamples with a rotary sectorial splitter. Two subsamples were then randomly selected among these, further ground separately to a nominal particle size of 0.212 mm, and further divided again using the rotary sectorial splitter into 1-g analytical subsamples for chemical analysis. Three 1-g analytical subsamples were then randomly selected for trace metal content analysis. In total, trace metal content was determined in 576 1-g analytical subsamples produced from the 12 field sampling locations. This procedure is considered to comply well with the tenets of TOS.

For the GSP, 24 300-g field samples were brought to the laboratory, dried at a temperature of 60 °C, and sieved at 2 mm using a

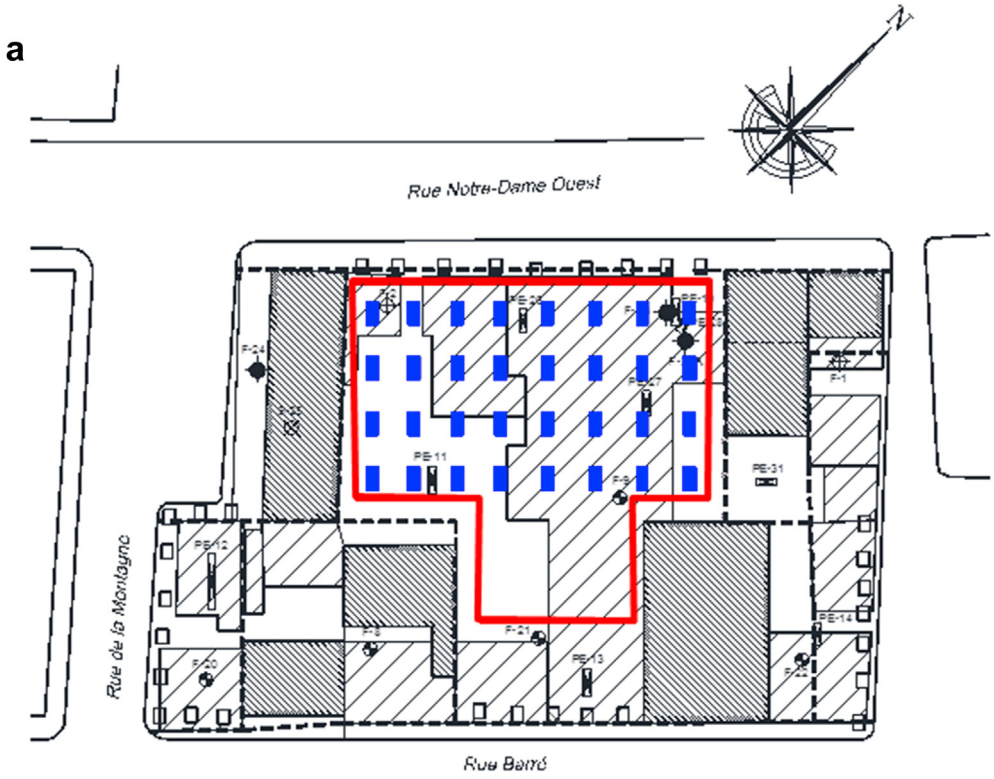


Fig. 2a. Site plan for Boudreault et al. [6] and Dubé et al. [7] showing systematic sampling locations. The red line delineates the open brownfield area located between buildings (dark grey polygons). Hatched areas show the locations of former buildings. Each blue rectangle represents a sampling location. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

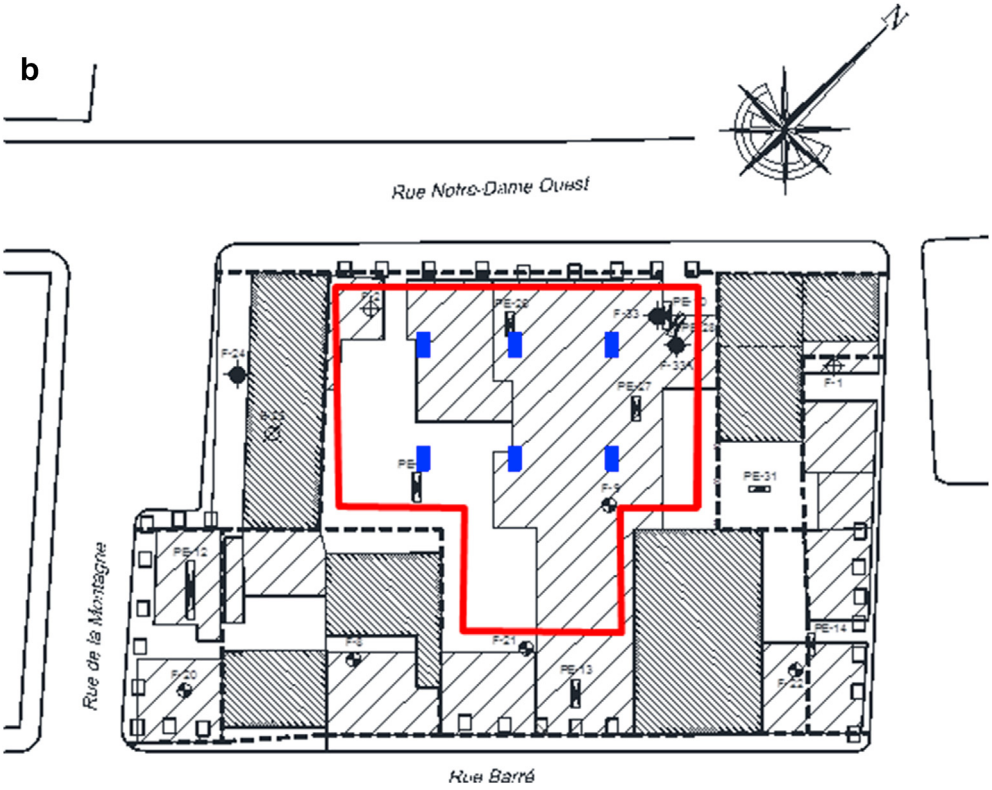


Fig. 2b. Site plan for Boudreault et al. [6] and Dubé et al. [7] showing supplemental sampling locations. The red line delineates the open brownfield area located between buildings (dark grey polygons). Hatched areas show the locations of former buildings. Each blue rectangle represents a supplemental sampling location. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

plastic sieve. Then, six 1-g analytical subsamples were obtained from the surface of each sieved field sample by grab sampling using a plastic spatula. A total of 144 1-g analytical subsamples were produced from the 24 field samples and were analyzed for trace metal content. This procedure is obviously not in accordance with TOS' stipulations for representative subsampling.

All analytical samples were analyzed for their trace metal content by ICP-OES (Varian Vista-MPX) following acid digestion according to CEAEQ [36]. Experimental relative sampling variances were estimated for each step of each protocol using a fully-nested ANOVA. Theoretical relative sampling variances were calculated following the mineral model (Equation [5]).

The purpose is not to compare these opposing sampling procedures against one another, but for each procedure, comparison shall be made between the measured experimental relative sampling variance, $s^2(EXP)$, and the corresponding calculated theoretical relative sampling variance, $s^2(THEO)$.

4.2.1.2. Sampling protocols and data analysis in Dubé et al. [7]. Five laboratory sampling protocols (termed A through E) were assessed using the 12 field samples obtained from the Boudreault et al. [7] study, which had initially been dried and sieved at a nominal particle size of 1 cm.

Protocol A consisted in sieving the soil at a nominal particle size of 2 mm followed by grabbing three replicate 1-g sub-samples with a small plastic spatula. Protocols B and C consisted in reducing the mass of the field samples by rotary sectorial splitting in three successive divisions, in order to produce the same size 1-g analytical samples. Protocols D and E used the same rotary splitting procedure as protocols B and C, but comprised an initial comminution to a nominal particle size of 1 mm before the first mass division, and a second comminution to a nominal particle size of 0.212 mm between the first and second mass divisions. Protocols C and D were similar to protocols B and E respectively, except for a preliminary screening of the soil at a nominal particle size of 2 mm.

Again, comparison shall be made between the measured experimental relative sampling variance of each procedure, $s^2(EXP)$, to the corresponding calculated theoretical relative sampling variance, $s^2(THEO)$.

Trace metal analysis were performed as in [6]. Experimental relative variances were determined using a fully nested ANOVA as in [6].

4.2.3. Calculation procedures with the "mineral model" vs. the "SDCD model"

For the "mineral model", Equation [5] was used to calculate the $s^2(FSE)$ of sampling techniques from each study.

For the "SDCD model", HI_L was first calculated using Equation [12] and then inputted in Equation [1] to calculate $s^2(FSE)$. These characteristic HI_L values were used to recalculate theoretical relative sampling variances from Boudreault et al. [6] and Dubé et al. [7]. These were compared to the theoretical relative sampling variances previously calculated using the mineral model using Equation [5] and to the experimental (measured) relative sampling variances reported in these studies.

Using Equation [12] requires dividing the soil into grain size fractions and determining the analyte content in each resolved size fraction. Nine primary field samples were randomly selected from the set of 64 primary field samples and were reduced to 530-g tertiary field samples using fractional shoveling. Each tertiary

field samples was then dried in an oven at 60 °C and subsequently divided in seven grain size fractions by mechanical sieving as presented in Table B1 in the appendices.

Total extractible trace metal content was determined in each grainsize fraction after acid digestion [36] and analysis by inductively coupled plasma atomic emission spectroscopy (ICP-AES) (Agilent VISTA-MPX). The following trace metals were analyzed as in Boudreault et al. [6] and Dubé et al. [7]: Cd, Co, Cr, Cu, Mn, Ni, Pb, and Zn.

Polypropylene, ceramic, or aluminum laboratory tools and equipment were used throughout in order to prevent contamination of the samples from the sampling equipment. Before each use, any piece of equipment in contact with the samples was washed with distilled water and soap, rinsed with water, soaked in a solution of dilute nitric acid (1 : 10 v/v), and rinsed in distilled water to avoid cross-contamination. All chemicals used were of Reagent A.C.S. grade. In addition, reference materials (SCP Science) were used for quality control. The detection limits (DL) and the quantification limits (QL) for the trace metals were as shown in Table 1. The DLs and QLs were determined according to CEAEQ [37].

Trace metal content in each grainsize fraction, a_α, was inputted in Equation [12] along with the average particle diameter and the relative mass of each grainsize fraction, d_α and $m_{L_\alpha}/\ m_L$ respectively. Moreover, particle density was set at $\rho = 2.65$ g cm^{-3}, i.e. the generally accepted average density of soil particles, and the shape factor was set identically at $f = 0.5$ for consistency (using the standard value for f is justified here because of the comparison setup). In this fashion, a comparative HI_L value was determined for each trace metal analyzed in the soil (i.e. Cd, Co, Cr, Cu, Mn, Ni, Pb, and Zn) for each primary field sample. A characteristic HI_L was then obtained for each trace metal by averaging HI_L values over these nine samples. Since some sampling protocols in Boudreault et al. [6] and Dubé et al. [7] involved grinding and sieving, the calculation procedure with Equation [12] had to account for the resulting modifications to grain size fractions. However, since grainsize distribution and trace metal content amongst grainsize fractions were not determined after each comminution or sieving step, $m_{L_\alpha}/\ m_L$ and a_α were adjusted based on the hypotheses and calculations presented in Appendix B.

5. Results

5.1. Validation of the "generalized model"

5.1.1. Case 1: sampling analyte-coated particles [4,32]

The case of Lyn et al. [32] is particularly interesting to review here since these authors used the "mineral model", i.e. Equation [5], to estimate the uncertainty in sampling pistachio nuts for the determination of their aflatoxin concentration.

Table 1
Detection and quantification limits.

Trace metal	Detection limit (mg kg^{-1})	Quantification limit (mg kg^{-1})
Cd	0.77	7.7
Co	0.49	4.9
Cr	0.35	3.5
Cu	0.31	3.1
Mn	13	130
Ni	1.1	11
Pb	3.0	30
Zn	3.3	33

Unfortunately, this is an obvious misuse of Equation [5] because in doing so, the authors implicitly assumed a liberation factor of 1 and, thus, that aflatoxin-bearing fungi and pistachio nuts could in fact be sampled independently, whereas the fungi coating was in reality sampled with the nuts. Thus, the authors obtained a theoretical uncertainty of 137%, which is coherent with the larger constitutional heterogeneity assumed by the "mineral model" for two liberated phases, but not with the actual heterogeneity carried by the analyte. Therefore, their theoretical estimate of sampling uncertainty was bound to much larger than their empirical estimates of 22.5% (1k, 68% confidence) and 45% (2k, 95% confidence). Lyn et al. [32] concluded that Gy's model of constitutional heterogeneity could not be used to adequately estimate sampling uncertainty and sampling variance in this material system. However, it is manifestly not Gy's mineral model which is at fault, but rather, its application to this particular material system. As shown earlier in this paper, it is possible to derive particular forms of Equation [2] for specific materials systems. Thus, Equation [5] was not appropriate for the material system studied by Lyn et al. [32] because the heterogeneity carried by the analyte, the aflatoxin, could not be adequately represented by two distinct phases, nut and aflatoxin, assumed to be completely liberated from one another.

However, when using Equations [1] and [10] with the same parameter values from Lyn et al. [32] the present authors obtain

$$HI_L = f\rho\left[d_c^3\frac{(1-a_{Lc})^2}{a_{Lc}} + \sum_\alpha d_\alpha^3 a_{L\alpha}\right]$$
$$= 0.27 \times 0.9951\,\frac{g}{cm^3} \times \left[(2\ cm)^3 \times \frac{(1-0.05)^2}{0.05} + 0\right] = 38.8\ g$$

and

$$s^2(FSE) = (1/m_s - 1/m_L)HI_L \approx HI_L/m_s = 38.8\ g/250\ g = 0.1552$$

$$RSD = \sqrt{0.1552} \times 100 = 39.4\%$$

This re-calculated sampling uncertainty using Equations [1] and [10] lies between the empirical uncertainty estimates of Lyn et al. [32] and thus provides a significantly improved theoretical estimate of the latter than the one obtained from applying Equation [5].

Equation [11] or [12] could not be used with the data from Lyn et al. [32]; as the distribution of aflatoxin content as a function of nut size was not available. Moreover, a batch of commercial nuts is a naturally sized, strongly sorted material with a minimal size range, so it would in any way have been difficult, if not impossible, to obtain a meaningful characterization of aflatoxin content as a function of nut size distribution. Hence, for such a case, Equation [10] would be a more appropriate model of constitutional heterogeneity.

Table 3 presents the RSD value obtained by Gerlach et al. [4] for salt-coated sand particles. Neglecting the salt concentration of size fractions, one can simply use Equations [1] and [9] and obtain

$$HI_L = \sum_\alpha v_\alpha \rho_\alpha \frac{(1-a_{L\alpha})^2}{a_{L\alpha}} = 2.65\ \frac{g}{cm^3} \times \frac{(1-0.33)^2}{0.33}$$
$$\times \left[(0.045)^3 + (0.136)^3 + (0.300)^3\right] cm^3 = 0.1067\ g$$

Table 2
HI_L and $s^2(FSE)$ recalculated with Equation [11] from data in Gerlach et al. [3].

Analyte	Soil composition		L_α	$L_{\alpha\beta}$	$m_{L\alpha}$ (g)	$m_{L\alpha\beta}$ (g)	m_L (g)	m_s (g)	$\rho_{\alpha\beta}$ (g cm^{-3})	d_N (cm)	d_{avg} (cm)	$a_{\alpha\beta}$	a_L	f_α	HI_L (g)	$s^2(FSE)$	RSD (%)	RSD‡ (%) [3]	Measured RSD¥ (%) [3]
Salt (NaCl)	Fine	Fine sand	1	1	70	70	80	5	2.650	0.015	0.010	0	0.0625	1	3.82×10^{-4}	7.16×10^{-5}	0.84	0.75	RS: 5.3
		Magnetite	2	1	10	5			5.175	0.025	0.023	0							PCRS: 7.0
		Salt		2		5			2.165	0.025	0.023	1							FS: 15
	Coarse	Coarse sand	1	1	80	70			2.650	0.025	0.023	0			4.08×10^{-4}	7.66×10^{-5}	0.88		C&Q: 18
		Magnetite		2		5			5.175	0.025	0.023	0							GS: 32
		Salt		3		5			2.165	0.025	0.023	1							
Magnetite	Fine	Fine sand	1	1	70	70			2.650	0.015	0.010	0			9.01×10^{-4}	1.69×10^{-4}	1.3	1.2	RS: 5.1
		Magnetite	2	1	10	5			5.175	0.025	0.023	1							PCRS: 5.7
		Salt		2		5			2.165	0.025	0.023	0							FS: 6.1
	Coarse	Coarse sand	1	1	80	70			2.650	0.025	0.023	0			9.28×10^{-4}	1.74×10^{-4}	1.3		C&Q: 8.2
		Magnetite		2		5			5.175	0.025	0.023	1							GS: 51
		Salt		3		5			2.165	0.025	0.023	0							

‡: Recalculated from Equation [8] with data from Gerlach et al. [3]. Note that the value of a_L in Gerlach et al. [3] was incorrect.
¥: RS = Riffle Splitter; PCRS: Paper Cone Riffle Splitter; FS = Fractional Shoveling; C & Q: Coning and Quartering; GS = Grab Sampling.

Table 3
HI_L and $s^2(FSE)$ recalculated with Equations [11] and [12] from data in Gerlach et al. [4].

Analyte	Study # from [4]	Soil composition	L_α	$L_{\alpha\beta}$	$m_{L\alpha}$ (g)	$m_{L\alpha\beta}$ (g)	m_L (g)	m_s (g)	$\rho_{\alpha\beta}$ (g cm^{-3})	d_N (cm)	d_{avg} (cm)	$a_{\alpha\beta}$	a_L	f_α	HI_L (g)	$HI_{L\ total}$ (g)	$s^2(FSE)$	RSD (%)	RDS‡ (%) [4]	Measured RSD (%) [4]
Salt (NaCl)	1	Sand	1	1	39.8	39.8	40	5	2.650	0.060	0.040	0	0.0050	0.50	8.44×10^{-5}	4.03×10^{-2}	7.05×10^{-3}	8.4	8.2	18
		Salt	2	1	0.2	0.2			2.165	0.050	0.050	1		0.75	4.02×10^{-2}					
	2	Sand	1	1	23	23	36	1	2.650	0.060	0.040	0	0.0278	0.50	5.42×10^{-5}	6.22×10^{-2}	6.04×10^{-2}	25	160	35
		Salt	2	1	1	1			2.165	0.050	0.050	1		0.75	6.91×10^{-3}					
		Coarse particles	3	1	12	12			2.650	0.500	0.500	0		0.50	5.52×10^{-2}					
	2 (no coarse particles)	Sand	1	1	23	23	24		2.650	0.060	0.040	0	0.0417	0.50	8.13×10^{-5}	4.56×10^{-3}	4.43×10^{-3}	6.7	6.7	17
		Salt	1	1	1	1			2.16	0.050	0.050	1		0.75	4.47×10^{-3}					
	3 (1:1:1 mixture)	Small particles	1	1	2	2	6		2.650	0.071	0.045	0.0038	0.0026	1	7.25×10^{-6}	6.62×10^{-3}	5.51×10^{-3}	7.4	7.4	12
		Medium particles	2	1	2	2			2.650	0.200	0.136	0.0029			9.64×10^{-6}					
		Large particles	3	1	2	2			2.650	0.400	0.300	0.0013			3.29×10^{-3}					

‡: Recalculated from Equation [8] for studies 1 and 2 and from Equation [5] for study 3 with data from Gerlach et al. [4]. Note that the HI_L values for the case without coarse particles in study 2 and the case of small particles in study 3 were incorrect in Gerlach et al. [4].

Table 4
Sample distribution of analyte content amongst size fractions.

		Trace metal															
		Cd		Co		Cr		Cu		Mn		Ni		Pb		Zn	
d_α (cm)	m_{L_α}/m_L	a_α	$(a_\alpha - a_L)^2$	a_α	$(a_\alpha - a_L)^2$	a_α	$(a_\alpha - a_L)^2$	a_α	$(a_\alpha - a_L)^2$	a_α	$(a_\alpha - a_L)^2$	a_α	$(a_\alpha - a_L)^2$	a_α	$(a_\alpha - a_L)^2$	a_α	$(a_\alpha - a_L)^2$
0.67	28%	5.10×10^{-6}	6.25×10^{-12}	5.30×10^{-6}	4.17×10^{-12}	1.30×10^{-5}	5.93×10^{-11}	4.16×10^{-5}	1.62×10^{-9}	3.65×10^{-4}	9.30×10^{-9}	1.33×10^{-5}	2.87×10^{-11}	2.76×10^{-4}	2.31×10^{-7}	3.92×10^{-4}	1.33×10^{-7}
0.27	15%	9.80×10^{-6}	4.84×10^{-12}	9.40×10^{-6}	4.23×10^{-12}	2.40×10^{-5}	1.09×10^{-11}	8.58×10^{-5}	1.54×10^{-11}	4.60×10^{-4}	2.75×10^{-12}	2.37×10^{-5}	2.54×10^{-11}	6.77×10^{-4}	6.28×10^{-9}	7.74×10^{-4}	2.68×10^{-10}
0.14	10%	8.60×10^{-6}	1.00×10^{-12}	8.40×10^{-6}	1.12×10^{-12}	2.24×10^{-5}	2.89×10^{-12}	9.60×10^{-5}	2.00×10^{-10}	5.17×10^{-4}	3.01×10^{-9}	2.12×10^{-5}	6.47×10^{-12}	6.72×10^{-4}	7.08×10^{-9}	7.83×10^{-4}	6.39×10^{-10}
0.06	14%	7.80×10^{-6}	4.00×10^{-14}	7.60×10^{-6}	6.61 x 10-14	2.38×10^{-5}	9.61×10^{-12}	8.40×10^{-5}	4.53×10^{-12}	5.13×10^{-4}	2.65×10^{-9}	1.98×10^{-5}	1.31×10^{-12}	8.94×10^{-4}	1.88×10^{-8}	8.43×10^{-4}	7.39×10^{-9}
0.032	5%	6.20×10^{-6}	1.96×10^{-12}	6.10×10^{-6}	1.54×10^{-12}	1.69×10^{-5}	1.44×10^{-11}	6.64×10^{-5}	2.39×10^{-10}	4.00×10^{-4}	3.79×10^{-9}	1.54×10^{-5}	1.06×10^{-11}	7.70×10^{-4}	1.97×10^{-10}	7.11×10^{-4}	2.10×10^{-9}
0.018	10%	6.30×10^{-6}	1.69×10^{-12}	6.00×10^{-6}	1.80×10^{-12}	1.75×10^{-5}	1.02×10^{-11}	5.74×10^{-5}	5.99×10^{-10}	3.87×10^{-4}	5.56×10^{-9}	1.41×10^{-5}	2.08×10^{-11}	6.65×10^{-4}	8.40×10^{-9}	5.94×10^{-4}	2.67×10^{-8}
0.008	18%	9.40×10^{-6}	3.24×10^{-12}	8.60×10^{-6}	1.58×10^{-12}	2.73×10^{-5}	4.36×10^{-11}	1.42×10^{-4}	3.60×10^{-9}	5.90×10^{-4}	1.64×10^{-8}	2.31×10^{-5}	1.97×10^{-11}	1.34×10^{-3}	3.42×10^{-7}	1.20×10^{-3}	2.00×10^{-7}
a_L		7.60×10^{-6}		7.34×10^{-6}		2.07×10^{-5}		8.19×10^{-5}		4.62×10^{-4}		1.87×10^{-5}		7.56×10^{-4}		7.57×10^{-4}	

and

$$s^2(FSE) = (1/m_s - 1/m_L)HI_L \approx HI_L/m_s = 0.1067\ g/1\ g = 0.1067$$

$$RSD = \sqrt{0.1067} \times 100 = 32.7\%$$

This theoretical RSD value is almost three times larger than the RSD of 12% estimated empirically by Gerlach et al. [4]. A better RSD estimate can be obtained by using the available data on the analyte content of each size fraction with Equation [12] as shown in section 5.1.3.

5.1.2. Case 2: sampling analyte-enriched particles [3,4] (studies 1 and 2))

Tables 2 and 3 present the RSD (i.e. $\sqrt{s^2(FSE)} \times 100$) values calculated with the mineral model (Equation [5]) and recalculated using Equations [1] and [11] from the data in Gerlach et al. [3] and Gerlach et al. [4] respectively. Densities and shape factors were taken as in Gerlach et al. [3,4]; while average mass fractions, i.e. a_L values, were recalculated, hence leading to a correction of the RSD values reported in Gerlach et al. [3,4].

RSD values calculated using Equations [1] and [11] were close to those reported in Gerlach et al. [3,4]; except for the case of coarse particles in study 2 [13]. As expected, all RSD values calculated using Equations [1] and [11] were smaller than the measured RSD (%) values, similarly to those from the Gerlach et al. studies [3,4].

However, except for the stated case of coarse particles, values calculated using Equations [1] and [11] were systematically larger, albeit only slightly, than values calculated using Equation [5]. A possible reason for this is that Equation [11] does not rely on a granulometric factor, which is an assumption regarding the characteristics of the grainsize distribution. Gerlach et al. [3,4] used $g = 0.55$, thus assuming a calibrated material [15]. However, when using $g = 0.75$, for a naturally calibrated material [15], the recalculated RSD = 0.88%. Therefore, the assumption made for g by Gerlach et al. may not have been appropriate for the mixture of sand, salt and magnetite – a clear warning against insertion of standard numbers without specific justification.

The case of the coarse particles in study 2 also illustrates well the limitations of Equation [4] for materials that do not conform to the underlying assumptions. Gerlach et al. [4] report an estimated RSD of 160% for study 2, compared to a measured RSD of 35%. Not only is this is a poor theoretical estimate of the measured RSD, but it is incoherent with the fact that it should be smaller since, by definition, it represents the minimum relative sampling variance. Gerlach et al. [4] explained this discrepancy by the fact that the use of the nominal diameter, d_N, in Equation [4] assumes that this is the upper bound of a continuous distribution of particle sizes, while the distribution of their sand is distinctly discrete and bimodal, due to the addition of a large proportion (33%) of well-screened coarse particles. This is another application of the standard formula in a situation in which this is logically unsound; the formula should simply not have been invoked.

Using the appropriate Equation [11] provided an estimated RSD of 25%, which is smaller than the measured RSD of 35% reported and a much better estimate than that obtained by Gerlach et al. [4] using the mineral model. The fact that Equation [11] does not rely on assumptions regarding the particle size distribution provides it with a better predictive capacity of HI_L when the sampled materials do not follow the assumptions of Equation [4].

5.1.3. Case 3: sampling a mixture of analyte-enriched and analyte-coated particles [4]

Table 3 presents $s^2(FSE)$ values calculated by Gerlach et al. [4] using Equation [5] compared to our recalculated values using Equations [1] and [12] for a mixture of salt-coated sand particles of different sizes. The RSD value calculated in the present study were generally at the same level as that calculated by Gerlach et al. [4] and smaller than the empirically measured RSD. This result shows that Equation [12] can be used with good reason for analyte-coated particles, provided the analyte content in each size fraction is quantifiable. Moreover, contrary to Equation [4], Equation [12] does not require the determination of the number of particles in each size fraction, which, in all likelihood, is an impossible task with real-world soils.

5.2. Comparing the "mineral model" and the "SDCD model" for real-world contaminated soils

Initial $s^2(THEO)$ values calculated with Equation [5] in Boudreault et al. [6] and Dubé et al. [7] largely overestimated the measured $s^2(EXP)$. As mentioned earlier, the use of the "mineral model" in these studies served as an example of how the difficulty in characterizing its parameters for contaminated soils can lead to an ill-advised use of "blanket" values for many of them, i.e. f, g, and l. Moreover, the mineralogical factor, c, as defined in the "mineral model" (i.e. as in Equations [4] and [5]) is extremely difficult to calculate for contaminated soils within any practical horizon. Most particularly, the density of the analyte is defined as that of a pure mineral, ρ_M, a concept extremely difficult to reconcile with the many possible different states of contaminants in soil (e.g. precipitates, complexes, exchangeable ions). Moreover, at the very low grades of contaminants in polluted soils compared with ores, the mineralogical factor becomes very large indeed. Combined with the suggested "blanket" value of $l = 1$ [25], the resulting $s^2(THEO)$ estimate takes on nonsensical high values when compared to experimental variances. To 'control' this inflation of c and, as a result, of $s^2(THEO)$ estimates, l would have to take on a very small value.

This is precisely what Boudreault et al. [6] and Dubé et al. [7] reflected upon by investigating the use of Equation [6] for estimating l, but without conclusive results as to its applicability for contaminated soils. These early reflections lead to the developments presented in the present treatment.

At the heart of these developments is the hypothesis that in contaminated soil the analyte (i.e. the contaminant) content must be viewed as correlated with both particle size and density. This fundamentally changes how the practical equation stemming from Equation [2] is derived (i.e. Equation [8] and its variants Equations [9]-[12] vs Equation [4]). Sample calculations made with the "SDCD model" using data from Boudreault et al. [6] and Dubé et al. [7] are presented in Appendix B. The following will refer to this appendix were appropriate.

An important feature of the "SDCD model" is the term $(a_\alpha - a_L)$ in Equations [11] and [12], which reflects the heterogeneity carried by the analyte to its distribution amongst size fractions. This is crucial for a proper understanding of the constitutional heterogeneity of contaminated soil and its relation to sampling.

Assuming that all incorrect sampling errors have been eliminated (an absolute must before any considerations of correct sampling errors can occur – but which is in fact one of the prominent transgressions most often met with), a soil in which the

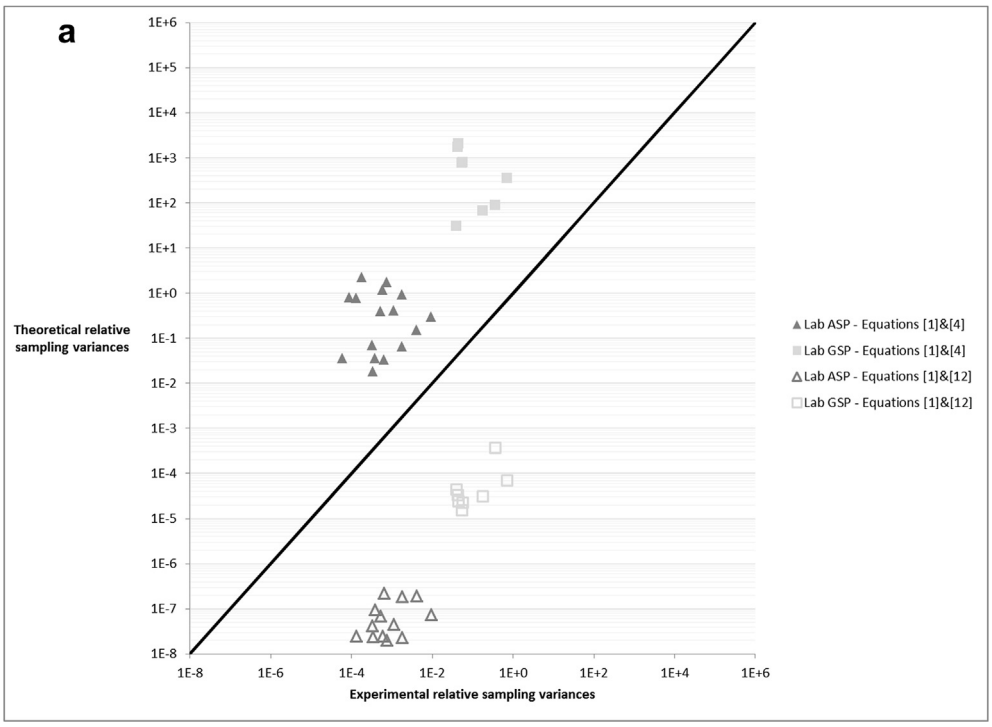


Fig. 3a. Theoretical vs. empirical relative sampling variances from Boudreault et al. [6]. Solid 45-degree sloping line denotes perfect correlation. See text for interpretation. (Lab GSP values were incorrectly reported in Boudreault et al. [6]; but are corrected here, without changing any of the findings illustrated).

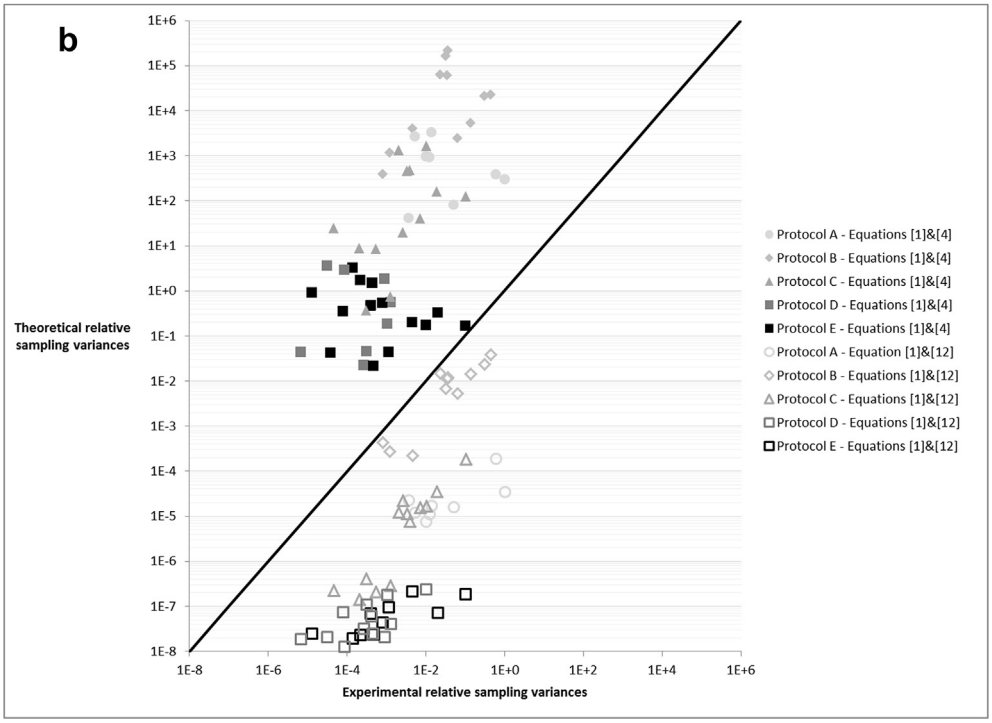


Fig. 3b. Theoretical vs. empirical relative sampling variance from Dubé et al. [7]. Solid 45-degree sloping line denotes perfect correlation. See text for interpretation.

contaminant is more-or-less equally distributed amongst all size fractions will have a lower constitutional heterogeneity and for which it will be “easier” to minimize the correct sampling errors than for more divergent cases.

Therefore at the other end of this spectrum, a soil in which the contaminant is more-or-less prominently (or perhaps only) concentrated in a specific size fraction, will have a much larger constitutional heterogeneity because $(a_\alpha - a_L)$ will be large for this size fraction and null for the others and, therefore, minimizing correct sampling errors will require more attention. Table 4 illustrates this with a sample distribution of trace element concentrations amongst size fractions in one of the soil samples selected for the present study. These data show empirically that all trace metals are found in all size fractions and that their content in each size fraction was close to their average soil content, i.e. $(a_\alpha - a_L)^2$ in Equation [12].

Analyte distributions such as shown by Table 4 were therefore used to recalculate HI_L and $s^2(THEO)$ values for both Boudreault et al. [6] and Dubé et al. [7]. Appendix B presents sample calculations for HI_L values. Fig. 3a and 3b present the $s^2(THEO)$ values obtained from the HI_L values reported in Boudreault et al. [6] and Dubé et al. [7]; calculated with the "mineral model" versus those recalculated in the present study with the "SDCD model" for graphical comparison.

NB. Note that the $s^2(THEO)$ values obtained in Boudreault et al. [6] for GSP at the laboratory sampling stage were incorrectly reported in that publication; they should have been smaller. However, this does not lead to any loss of generalization. These values were corrected in the present study in Fig. 3a, and nevertheless remain much larger than $s^2(EXP)$, which is still impossible. Thus, the aforementioned shortcomings of the "mineral model" for contaminated soil and similar materials stand unaffected.

Table B2 presents average HI_L values obtained with both models. The "SDCD model" resulted in HI_L values several orders of magnitude smaller than those obtained with the "mineral model", as the latter is unable to represent the analyte distribution amongst grainsize fractions because it does not account for $(a_\alpha - a_L)^2$. As shown in Table 4, the typical analyte distribution found in the studied soil resulted in very small $(a_\alpha - a_L)^2$ in Equation [12] and, therefore, in much smaller $s^2(THEO)$ values.

As shown in Fig. 3a and b all $s^2(THEO)$ values recalculated with the "SDCD model" were smaller than measured $s^2(EXP)$, i.e. in the geometric setup of this plot they all fall below the line of perfect correlation. This implies that, contrary to the $s^2(THEO)$ values obtained with the "mineral model", the recalculated $s^2(THEO)$ values are in de facto agreement with the general principles of TOS since $s^2(THEO)$ reflects the irreducible minimum Fundamental Sampling Error, while $s^2(EXP)$ also encompasses other sampling errors, such as not fully eliminated incorrect sampling errors or residual grouping and segregation error effects. Therefore, it should always be expected that $s^2(THEO) < s^2(EXP)$. In this context, in the present study a perfect correlation between theoretical and experimental relative variances was not to be expected, as this would have implied that absolutely no sampling error other than the FSE was made during the sampling experiments of Boudreault et al. [6] and Dubé et al. [7]; which they likely were not - despite significant efforts.

From these results, it must be presumed that the classical "mineral model" has a poor predictive capability for distinguishable material classes characterized by a mixture of classical fragments (analyte enriched particles) and coated particles, such as the contaminated soil studied in Boudreault et al. [6] and Dubé et al. [7,34]. On the other hand, the "SDCD model" provides a more practical approach for more meaningful estimates of HI_L and $s^2(THEO)$.

6. Discussion

In the examples above, it was noted that very often the calculated $s^2(THEO)$ was actually larger than its empirically estimated counterpart $s^2(EXP)$. This is distinctly inconsistent with the logic of TOS, at least if the theoretical model for HI_L, or its modifications, are taken at face value to mean that $s^2(FSE)$ is indeed the incompressible absolute minimum sampling error. In very many applications of the "mineral model" found in the literature this traditional understanding is prevalent. Thus and therefore, any $s^2(EXP)$ found to be smaller than its theoretical counterpart presents an illogical empirical finding in need of a deeper explanation.

We have here investigated various possible explanations for such "impossible" state of affairs. First, one could argue that the "mineral model" is not at fault, but rather the values attributed to its parameters are. This has been discussed in previous papers [6,7] where it was shown that it is indeed possible to adjust the values of parameters, such as the liberation factor, to match empirically measured sampling variances. By doing so, the aforementioned studies attempted to highlight the discrepancy between the blanket value of $l = 1$ and the maximal value l could have taken if the measured sampling variances were due solely to the fundamental sampling error. Since other sources of sampling errors are always embedded in measured sampling errors, the real l value was deemed to be smaller than the adjusted value, and much smaller than 1 for contaminated soil. However, looking at Equation [5], one could have also chosen to adjust c, f or g instead of l. Hence the inverse problem of adjusting theoretical sampling variances to experimental sampling variances becomes insoluble, indeed a travesty. These factors were categorically never intended as such fudge factors by Gy himself, see introduction above.

The parameters c and l, which are constitutive of the "mineral model" for HI_L (i.e. Eq [4]) and $s^2(THEO)$ (i.e. Equation [5]), are difficult (if at all possible) to estimate a priori for contaminated soils, because they lack mathematical definitions based on measurable characteristics of the soil and the distribution of the contaminants it contains. Therefore, even if the values attributed to the parameters of the "mineral model" could be adjusted to experimental measurements of sampling variances, they would still not be substantiated by independent physical measurements. Conversely, the "SDCD model" for HI_L presented herein with Equations [8] to [13] provides the means to design sampling protocols based on measureable parameters, most importantly the distribution of the analyte between the size classes in Equations [11] to [13].

Second, one could also argue that the soil sampled in Boudreault et al. [6] and in Dubé et al. [7] is a special case that would warrant a specific mathematical derivation $s^2(THEO)$ from Equation [2] and that the theoretical "mineral model" is a general model representing the vast majority of analytes in particulate matter. However, the analysis presented above showed that when the "mineral model" and the "SDCD model" were applied to the same sampling data (i.e. from the Lyn et al. [32] and the Gerlach et al. [3,4] studies), the "mineral model" failed to represent cases where the analyte distribution differed from the case of analyte-enriched particles,

while the "SDCD model" was able to correctly model all cases.

An undeniable strength of the classical "mineral model" is its capacity to provide estimates of HI_L and $s^2(THEO)$ using a priori values of its constitutive parameters. However, for cases other than analyte-enriched particles, this is at the cost of specific ad-hoc assumptions on the analyte distribution in the particulate matter. Therefore, the "SDCD model" is a de facto improvement on the "mineral model" inasmuch as it provides demonstrably better estimates of HI_L and $s^2(THEO)$, while avoiding the limiting assumptions of the latter.

Contrary to the "mineral model", the "SDCD model" is directly unable to provide these estimates using a priori values of its constitutive parameters because it only relies on empirical, measured values for grainsize and analyte distribution. Thus, depending on the specific case, its use may require a preliminary sampling experiment in order to make such measurements as a form of calibration of the model before using it for designing a final sampling protocol. We find this a small price to pay relative to the vastly increased realism provided by the "SDCD model".

If the constitutional heterogeneity of contaminated soil can be represented by that of coated particles (case 1), then $s^2(THEO)$ can be calculated using Equations [1] and [10]. Equation [10] can be simplified when considering that the critical size fraction is larger than the nominal particle size (section 3.1). The latter can also be established from a preliminary observation of the soil. The minimum sample mass and a sampling protocol can thus be derived, and then adjusted if necessary, from measurement of the empirical grain size distribution.

If the constitutional heterogeneity of the soil stems from analyte-enriched and coated particles (cases 2 and 3), the preliminary sampling phase can be conducted as described for a "case 1" application, with the imperative objective to obtain samples representative of all grain size fractions, in order to be able to obtain meaningful estimates of HI_L from Equation [11] or [12] or [13]. Then, using Equation [1] with the estimated HI_L, the minimum sample mass and a first sampling protocol can be readjusted to ensure that the measurements of analyte concentration meet the required degree of representativeness.

A last, very important, general point: It should be noted that all models for HI_L reported herein, i.e. the "mineral model" as well as all variants of the "SDCD model", are based on the assumption that the analyte is randomly distributed in the lot. This assumption goes all the way to Gy's original derivations. The models only provide a valid estimate of the sampling variance, or of the minimum sample mass, for randomly distributed lots from which samples can be extracted (grab or composite sampling) by collecting particles individually. Of course, such an idealized lot seldom exists in reality. Therefore, the sampling variance estimated by these models will always constitute an absolute minimum sampling variance. In reality, several residual sampling errors often occur that will make sampling variance larger. As mentioned throughout this paper (see also Minkkinen and Esbensen [21]), Incorrect Sampling Errors (ISE) must be fully eliminated before any meaningful analysis and comparison of sampling variance can be attempted. Even when/if this is the case, the reality is that in almost all instances, it is impossible to collect samples from individual particles from the lot - and the distribution of the analyte is almost always affected by some form of segregation, ibid. except in marginal, often artificial laboratory cases from which no generalization is merited. Therefore, as thoroughly demonstrated by Minkkinen and Esbensen [11]; for such cases there will very nearly always be a residual grouping and segregation error effect, which will increase variance and make impossible valid determination of the sampling variance based on HI_L only. Sampling variance will invariably include terms resulting from short-range and long-range effects from one or several forms of segregation (original, induced, pouring. mixing … segregation), or such as autocorrelation characterising the distribution of the analyte in stationary or moving 1-D lots (the latter in the form of process lots). This realization is particularly important for primary sampling, e.g. field sampling on contaminated sites, where it is obviously impossible to reduce segregation before sampling lest the whole site must undergo forceful mixing, and clearly transgressing any-and-all realistic sampling cost considerations.

In the studies recalled in the present study, especially those of Gerlach et al. [3,4] in which the lot to be sampled was purposefully segregated, segregation could explain the differences between the theoretical estimates of sampling variance recalculated with the "SDCD model". It should also be noted that when segregation and short/long-range autocorrelation patterns are present in the lot, sampling variance will also depend on the sample mass and especially on the sampling mode (systematic, stratified random or random sampling), Minkkinen and Esbensen [11,21]. Some studies reported above involved different sampling modes, such as Gerlach et al. [3]; Boudreault et al. [6] and Dubé et al. [7] and it could clearly be seen that difference between experimental sampling variance and its theoretical estimation from the "SDCD model" depend on the mode of sampling used, grab sampling corresponding to the largest differences.

Finally, it should be noted that minimum sample masses were not calculated using the "SDCD model" and compared to masses used in the comparative studies because, as explained above, contributions to sampling variance due to segregation are not accounted for by any HI_L model presented here.

7. Conclusions

At the heart of TOS is the well-known often-called "Gy's equation", or "Gy's Formula" (Equation [5] above), which gives the irreducible relative sampling variance due to the Fundamental Sampling Error, $s^2(FSE)$. The latter is considered to be an estimate of the absolute minimum sampling error variance, provided all other sampling errors have been properly minimized and/or eliminated ("Gy's mandate", see introduction). It cannot be over-emphasized that adequate competence with respect to TOS along the full lot-to-analysis pathway is a critical success factor for this endeavor. Ill-reflected, voluntary, complacent parameter plug-in actions have no place in the science of representative sampling.

Gy's equation is based on a practical implementation of the heterogeneity invariant for lot materials made up of aggregates of assemblages of distinct fragments from different size fractions and density phases. Equation [5] is based on parameters which for many particulate materials, but certainly not all, can be estimated a priori (i.e. before actual sampling), which makes it a powerful tool for designing practical sampling protocols. However, research has shown that Equation [5] does not adequately represent the constitutional heterogeneity for example of contaminated soils and similar mixed materials. In fact, estimates of the theoretical relative sampling variance with Equation [5] for typical real-world contaminated soils were several orders of magnitude larger than experimentally measured relative

sampling variances. This is due to the misuse of Equation [5] for granular matrices having a constitutional heterogeneity not adequately represented by the model of constitutional heterogeneity upon which this equation is based. The model of constitutional heterogeneity in Equation [5] corresponds to a granular material made of analyte-enriched particles distinct from analyte-free particles, while in contaminated soils for example, the analyte can also, or wholly, be present as a coating on soil matrix particles.

Following this realization, this paper returns to the concept of size-density classes to show how one can obtain a model of constitutional heterogeneity capable of representing mineral-like as well as soil-like materials. This "size-density-class derived" or "SDCD model" is based on the assumptions that the analyte content amongst the soil fragments is correlated both to size fraction as well as to phase density, whereas the effect of the correlation of analyte content to size fraction is not represented in Equation [5]. According to these assumptions, an analyte uniformly distributed amongst size fractions has a smaller heterogeneity than an analyte found mainly in a specific size fraction, or only. Importantly, there undoubtedly exists a gradual spectrum of real-world situations between these end-member scenarios.

Using the "SDCD model" with published data reproduced the relative sampling variances calculated for mineral-like matrices (as in the Gerlach et al. studies), but more importantly also corrected the relative sampling variance calculated for real contaminants by several orders of magnitudes. In all cases, the recalculated relative sampling variances were decreased below their corresponding experimental measurements, thus now fully as expected from TOS, i.e. relative sampling variance due to constitutional heterogeneity is indeed the minimum irreducible variance when all sampling errors are properly minimized or eliminated. Therefore, the "SDCD model" has the capacity of producing estimates of relative sampling variance due to the fundamental sampling error, which are fully meaningful with respect to all TOS conceptual understandings and principles.

Further empirical validation of the "SDCD model" is highly desirable, and strongly encouraged, but upon its first merits presented here, it shows a very promising potential of increasing TOS' applicability to contaminated soils, coated particular aggregates and similar mixed material systems. Further work on the "SDCD model" will need to address the many facetted forms and effects of grouping and segregation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge valuable financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) of Canada to studies referenced herein, as well as the present study through Discovery grants (grant numbers RGPIN 249734-13 and RGPIN 2019-06227) and a Canada Research Chair awarded to Professor Dubé.

KHE acknowledges the privilege of becoming a friend of Pierre Gy (1999–2015), not only gaining an inside scientific induction to TOS, but which also included many personal, social and cultural enrichments in life. Gy's life and oeuvre is a legacy [1], indeed a role model never more relevant than in these times of faster, easier, more … in many walks of science, technology and industry, perhaps not always wanting to make the full efforts needed.

The authors are most grateful to one ardent and insisting reviewer who kindly pointed out missing issues of key historical interest in the first version of our manuscript. We cannot thank this reviewer enough for his scholarly intervention and help.

Appendix A. Influence of analyte content on particle density

A thought-experiment was conducted to investigate the relationship between the concentration of a trace metal in soil and the density of the soil particles it is attached to through common retention mechanisms. Metal-enriched particles liberated from soil particles were not considered as both types of particles are then considered to have their specific density.

Cadmium (Cd) and lead (Pb) were considered for this experiment as well as possible geochemical species of these two elements that could be found on soil particles, namely Cd- or Pb-carbonate, hydroxide, oxide, and sulphide.

Table A.1 presents the density and the molar mass of each geochemical species, while Table A.2 presents the resulting particle density as a function of the concentration of each geochemical species.

For example, for a Pb concentration of 10 000 mg per kg of dry particulate matter (or 1%) and $PbCO_3$ as the geochemical species, the calculation of the resulting density is as follows.

First, the concentration of $PbCO_{3(s)}$, i.e. $[PbCO_3]_s$, is calculated as,

$$[PbCO_3]_s = [Pb]_s \times M_{Pb} \Big/ M_{PbCO_3} = 1 \times 10^4\, mg\, kg^{-1} \times \left(207.2\, g\, mol^{-1} \Big/ 267.2\, g\, mol^{-1}\right) = 1.29 \times 10^4\, mg\, kg^{-1}$$

Then, the resulting density of a soil particle having the above concentration $PbCO_{3(s)}$ is calculated as the weighted mean of the respective densities of soil particle and $PbCO_{3(s)}$,

$$\rho = \rho_{soil} \times \left(1 - a_{PbCO_{3(s)}}\right) + \rho_{PbCO_{3(s)}} \times a_{PbCO_{3(s)}} = 2.65\ g\ cm^{-3} \times \left[1 - \left(1.29 \times 10^4\ mg \Big/ 10^6\ mg\right)\right] + 6.6\ g\ cm^{-3} \times 1.29 \times 10^4\ mg \Big/ 10^6\ mg = 2.70\ g\ cm^{-3}$$

Fig. A.1 shows the relationship between the concentration of the geochemical species in the soil and soil particle density. Since soil particle density is expected to vary between 2.6 and 2.8 g cm^{-3}, trace metal species attached to it will not change it significantly for concentration below 10 000 mg kg^{-1} or 1%.

Table A.1
Density and molar mass of geochemical species

Geochemical species	Density (g cm^{-3})	Molar mass (g $mole^{-1}$)
Pb	11.3	207.2
$PbCO_3$	6.6	267.2
$Pb(OH)_2$	7.41	225.2
PbO	9.64	223.2
PbS	7.6	239.2
Cd	8.65	112.4
$CdCO_3$	4.26	172.4
$Cd(OH)_2$	4.79	146.4
CdO	7.0	128.4
CdS	4.82	144.5
Soil particle	2.65	N/A

Table A.2
Calculated densities of a soil particle coated with a precipitated Cd or Pb species

Lead (Pb)									
Total Pb concentration (mg kg^{-1})	Total Pb concentration (%)	Geochemical species concentration (%)				Density of soil particle coated with Pb species (g cm^{-3})			
		$PbCO_3$	$Pb(OH)_2$	PbS	PbO	$PbCO_3$	$Pb(OH)_2$	PbO	PbS
1	0.0001	0.000129	0.000109	0.000115	0.000108	2.65	2.65	2.65	2.65
10	0.001	0.00129	0.00109	0.00115	0.00108	2.65	2.65	2.65	2.65
100	0.01	0.0129	0.0109	0.0115	0.0108	2.65	2.65	2.65	2.65
1000	0.1	0.129	0.109	0.115	0.108	2.66	2.66	2.66	2.66
10 000	1	1.29	1.09	1.15	1.08	2.70	2.71	2.74	2.71
100 000	10	12.90	10.87	11.54	10.77	3.16	3.26	3.55	3.29
Cadmium (Cd)									
Cd concentration (mg kg^{-1})	Cd concentration (%)	Geochemical species concentration (%)				Density of soil particle coated with Cd species (g cm^{-3})			
		$CdCO_3$	$Cd(OH)_2$	CdS	CdO	$CdCO_3$	$Cd(OH)_2$	CdO	CdS
1	0.00	0.000153	0.000130	0.000129	0.000114	2.65	2.65	2.65	2.65
10	0.00	0.00153	0.00130	0.00129	0.00114	2.65	2.65	2.65	2.65
100	0.01	0.0153	0.0130	0.0129	0.0114	2.65	2.65	2.65	2.65
1000	0.10	0.153	0.130	0.129	0.114	2.66	2.66	2.66	2.66
10 000	1.00	1.53	1.30	1.29	1.14	2.71	2.72	2.76	2.73
100 000	10.00	15.34	13.03	12.85	11.42	3.26	3.38	3.72	3.41

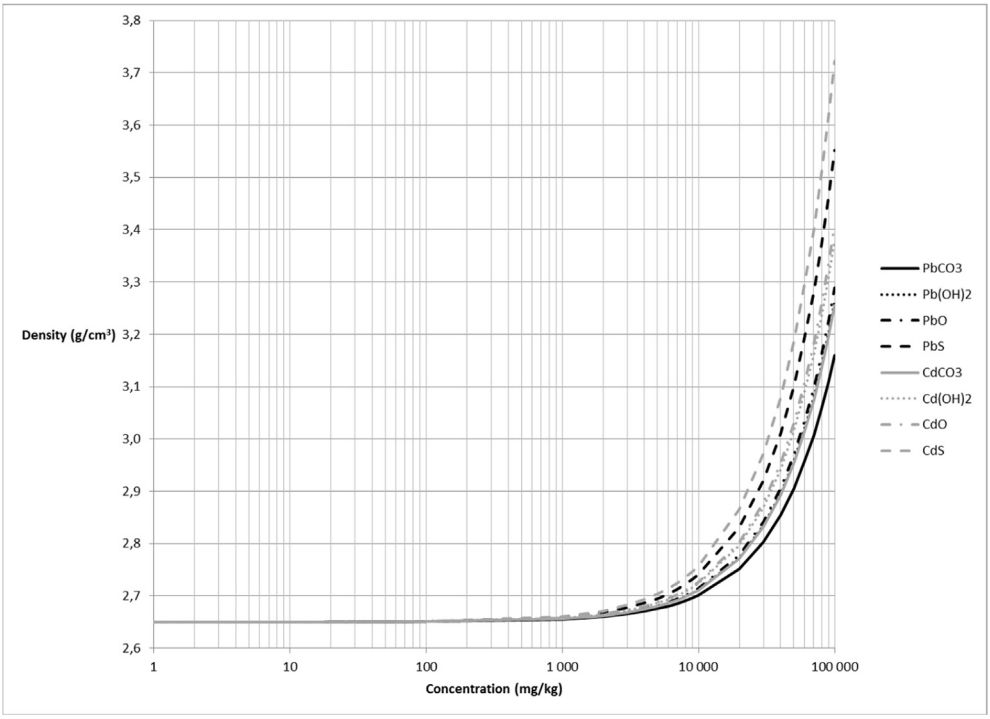


Fig. A.1. Particle density as a function of the concentration of precipitated Cd or Pb species.

Appendix B. Sample calculations of HI_L values

This appendix present hypotheses and sample calculations for the determination of characteristic HI_L values with Equation [12], for trace metals in the samples randomly selected from the Boudreault et al. study [38] (see section 4.2.1).

The characteristic HI_L values were used for recalculating the $s^2(THEO)$ calculated with Equation [5] and presented in Boudreault et al. [6] and Dubé et al. [7]. As explained in section 4.2.1, Equation [12] requires a sample to be separated into size fractions and the analyte content to be determined in each fraction. Table B.1 presents such data obtained for Cd content in one of the samples. For each size fraction, it provides the average particle diameter, d_α, the analyte (i.e. Cd) content, a_a, the size fraction proportion, $\frac{m_{L_\alpha}}{m_L}$, as well as the heterogeneity carried by the analyte (i.e. Cd), $f_\alpha \rho_\alpha d_\alpha^3 \frac{m_{L_\alpha}}{m_L} \frac{(a_\alpha - a_L)^2}{a_L^2}$.

The initial tertiary field samples brought to the laboratory (see section 4.2.1) had a nominal particle diameter, $d_N = 1$ cm. Then, depending on the subsampling protocol, some were ground in two comminution stages at $d_N = 0.1$ cm, and then at $d_N = 0.0212$ cm, while other were sieved at 0.2 cm prior to being ground at $d_N = 0.1$ cm, and then at $d_N = 0.0212$ cm. All possible states of a sample are presented in Table B.2, but, first, Table B.1 presents data

for three such distinct states of a sample, which were chosen to illustrate typical calculations that were performed.

Therefore, for the initial state of a sample, i.e. unsieved and unground, with $d_N = 1$ cm, a sample calculation for the heterogeneity carried by Cd in size fraction F1 would be:

$$f_\alpha \rho_\alpha d_\alpha^3 \frac{m_{L_\alpha}}{m_L} \frac{(a_\alpha - a_L)^2}{a_L^2} = 0.5 \times 2.65 \frac{g}{cm^3} \times (1cm)^3 \times (0.28) \times \frac{\left(5.10 \times 10^{-6} - 7.48 \times 10^{-6}\right)^2}{\left(7.48 \times 10^{-6}\right)^2} = 1.13 \times 10^{-2}\ g$$

Note that for all calculations, f = 0.5, and $\rho_\alpha = 2.65$ g cm^{-3}.

For protocols which involved sieving to remove grainsize fractions larger than 0.2 cm, a_α remained unchanged, while $m_{L_\alpha}/\ m_L$ was recalculated with m_L as the total mass of grainsize fractions <0.2 cm (see sample calculation in footnote "a" under Table B.1). For the grainsize fractions removed by sieving, $m_{L_\alpha} = 0$.

For protocols which involved grinding, both a_α and $m_{L_\alpha}/\ m_L$ were recalculated. Therefore, depending on the aforementioned nominal particle size obtained by grinding, i.e. either 1 mm or 0,212 mm, a_α was recalculated as the weighted sum of trace metal mass fractions down to the grainsize fraction immediately smaller than the new nominal particle size (see sample calculation in footnote "b" under Table B.1). Moreover, m_{L_α}/m_L was recalculated for the size fraction immediately smaller than the new nominal particle size as the sum of all m_{L_α}/m_L above this new nominal particle size (see sample calculation in footnote "c" under Table B.1). m_{L_α}/m_L values remained the same for the other size fractions.

Several HI_L values were thus obtained for each trace metal corresponding to each possible state of a sample under the different sampling protocols. These calculations were repeated for each of the nine samples and averaged to obtain mean HI_L values for each trace metal. These are presented in Table B.2 for each possible state of a sample depending on the subsampling protocol. It must be noted that these calculations were made a posteriori using data from previous studies and therefore, required hypotheses on m_{L_α}/m_L and a_α to account for modifications to the state of the sample due to sieving and grinding. In further studies, these hypotheses can be avoided by measuring these parameters a priori.

Table B.1
Sample grainsize data, Cd content and sample state for HI_L calculations for a given sampling location

		Sieve openings (cm)	d_α (cm)	No sieving or grinding, $d_N = 1$ cm: a_a (x 10^{-6})	$\frac{m_{L_\alpha}}{m_L}$	$f_\alpha \rho_\alpha d_\alpha^3 \frac{m_{L_\alpha}}{m_L} \frac{(a_\alpha - a_L)^2}{a_L^2}$ (g)	Sieving and no grinding, $d_N = 0.2$ cm: a_a (x 10^{-6})	$\frac{m_{L_\alpha}}{m_L}$	$f_\alpha \rho_\alpha d_\alpha^3 \frac{m_{L_\alpha}}{m_L} \frac{(a_\alpha - a_L)^2}{a_L^2}$ (g)	Sieving and grinding, $d_N = 0.1$ cm: a_a (x 10^{-6})	$\frac{m_{L_\alpha}}{m_L}$	$f_\alpha \rho_\alpha d_\alpha^3 \frac{m_{L_\alpha}}{m_L} \frac{(a_\alpha - a_L)^2}{a_L^2}$ (g)
Size fraction	F1	1.00 0.335	0.67	5.10	0.28	1.13 x 10^{-2}						
	F2	0.335 0.200	0.27	9.80	0.15	3.67 x 10^{-4}						
	F3	0.200 0.085	0.14	8.60	0.10	8.50 x 10^{-6}	8.60	0.17[a]	3.15 x 10^{-6}			
	F4	0.085 0.043	0.06	7.80	0.14	8.99 x 10^{-8}	7.80	0.25	8.22 x 10^{-8}	3.48[b]	0.42[c]	4.74 x 10^{-5}
	F5	0.043 0.021	0.032	6.20	0.05	6.28 x 10^{-8}	6.20	0.09	1.99 x 10^{-7}	6.20	0.09	1.99 x 10^{-7}
	F6	0.021 0.015	0.018	6.30	0.10	1.88 x 10^{-8}	6.30	0.17	6.28 x 10^{-8}	6.30	0.17	6.28 x 10^{-8}
	F7	0.015 0	0.008	9.40	0.18	6.57 x 10^{-9}	9.40	0.31	4.97 x 10^{-9}	9.40	0.31	4.97 x 10^{-9}
HI_L (g)						1.16 x 10^{-2}			3.50 x 10^{-6}			4.77 x 10^{-5}

[a] e.g. $\frac{m_{L_{F3}}}{m_L} = \frac{0.10}{(0.10 + 0.14 + 0.05 + 0.10 + 0.18)} = 0.17$

[b] e.g. $a_{F4} = [(0.17 \times 8.60) + (0.25 \times 7.80)] \times 10^{-6} = 3.48 \times 10^{-6}$

[c] e.g. $\frac{m_{L_{F4}}}{m_L} = 0.17 + 0.25 = 0.42$

Table B.2
Average HI_L (g) values calculated with the mineral and generalized models (Equations [4] and [12] respectively)

Trace metal	Initial d_N (cm)											
	1						0.2					
	d_N after grinding (cm)						**d_N after grinding (cm)**					
	1 (no grinding)		0.1		0.0212		0.2 (no grinding)		0.1		0.0212	
	HI_L											
	Eq. [4]	Eq. [12]	Eq. [4]	Eq. [12]	Eq. [4]	Eq. [12]	Eq. [4]	Eq. [12]	Eq. [4]	Eq. [12]	Eq. [4]	Eq. [12]
Cd	1.04 x 10^{5}	1.23 x 10^{-2}	1.04 x 10^{2}	1.10 x 10^{-6}	9.88 x 10^{-1}	2.41 x 10^{-8}	8.29 x 10^{2}	1.71 x 10^{-5}	1.04 x 10^{2}	1.18 x 10^{-6}	9.88 x 10^{-1}	2.17 x 10^{-8}
Co	1.29 x 10^{5}	7.03 x 10^{-3}	1.29 x 10^{2}	4.85 x 10^{-7}	1.23 x 10^{0}	2.51 x 10^{-8}	1.03 x 10^{3}	1.23 x 10^{-5}	1.29 x 10^{2}	7.19 x 10^{-7}	1.23 x 10^{0}	2.59 x 10^{-8}
Cr	4.64 x 10^{4}	1.53 x 10^{-2}	4.64 x 10^{1}	1.39 x 10^{-6}	4.42 x 10^{-1}	7.15 x 10^{-8}	3.71 x 10^{2}	1.14 x 10^{-5}	4.64 x 10^{1}	8.44 x 10^{-7}	4.42 x 10^{-1}	6.27 x 10^{-8}
Cu	1.96 x 10^{4}	2.42 x 10^{-2}	1.96 x 10^{1}	4.08 x 10^{-6}	1.87 x 10^{-1}	1.96 x 10^{-7}	1.57 x 10^{2}	3.53 x 10^{-5}	1.96 x 10^{1}	4.14 x 10^{-6}	1.87 x 10^{-1}	2.46 x 10^{-7}
Mn	1.97 x 10^{3}	5.49 x 10^{-3}	1.97 x 10^{0}	5.42 x 10^{-7}	1.88 x 10^{-2}	2.49 x 10^{-8}	1.58 x 10^{1}	2.27 x 10^{-5}	1.97 x 10^{0}	1.04 x 10^{-6}	1.88 x 10^{-2}	3.26 x 10^{-8}
Ni	4.65 x 10^{4}	1.12 x 10^{-2}	4.65 x 10^{1}	5.10 x 10^{-7}	4.43 x 10^{-1}	4.58 x 10^{-8}	3.72 x 10^{2}	7.76 x 10^{-6}	4.65 x 10^{1}	5.54 x 10^{-7}	4.43 x 10^{-1}	4.19 x 10^{-8}
Pb	4.63 x 10^{3}	3.94 x 10^{-2}	4.63 x 10^{0}	1.04 x 10^{-5}	4.41 x 10^{-2}	2.26 x 10^{-7}	3.71 x 10^{1}	1.92 x 10-4	4.63 x 10^{0}	5.98 x 10^{-6}	4.41 x 10^{-2}	1.83 x 10^{-7}
Zn	4.06 x 10^{3}	1.44 x 10^{-2}	4.06 x 10^{0}	2.30 x 10^{-6}	3.87 x 10^{-2}	9.85 x 10^{-8}	3.25 x 10^{1}	1.60 x 10^{-5}	4.06 x 10^{0}	2.05 x 10^{-6}	3.87 x 10^{-2}	1.12 x 10^{-7}

Table B.3
HI_L (g) for Cd per size fraction for all sampling location and mean HI_L (g) (Initial d_N = 1 cm, no sieving or grinding)

			Sampling location								
	Size fraction	d_α (cm)	1	2	3	4	5	6	7	8	9
Size fraction	F1	0.67	1.13×10^{-2}	1.66×10^{-2}	1.02×10^{-5}	3.09×10^{-2}	3.55×10^{-4}	4.10×10^{-5}	4.00×10^{-2}	9.61×10^{-3}	1.19×10^{-4}
	F2	0.27	3.67×10^{-4}	4.40×10^{-4}	5.99×10^{-6}	7.58×10^{-7}	1.16×10^{-5}	2.08×10^{-5}	3.25×10^{-4}	2.17×10^{-5}	2.35×10^{-5}
	F3	0.14	8.50×10^{-6}	5.10×10^{-5}	2.25×10^{-9}	1.47×10^{-4}	1.46×10^{-5}	1.52×10^{-6}	2.34×10^{-5}	1.55×10^{-6}	1.07×10^{-5}
	F4	0.06	8.99×10^{-8}	8.77×10^{-7}	7.24×10^{-8}	4.30×10^{-6}	5.58×10^{-8}	7.32×10^{-8}	4.14×10^{-7}	8.90×10^{-7}	1.08×10^{-7}
	F5	0.032	6.28×10^{-8}	1.68×10^{-7}	3.67×10^{-8}	4.98×10^{-8}	2.72×10^{-7}	2.57×10^{-8}	7.14×10^{-8}	1.40×10^{-7}	9.95×10^{-8}
	F6	0.018	1.88×10^{-8}	3.54×10^{-8}	1.23×10^{-8}	1.73×10^{-8}	2.40×10^{-8}	3.25×10^{-9}	5.91×10^{-9}	1.31×10^{-9}	3.72×10^{-8}
	F7	0.008	6.57×10^{-9}	3.29×10^{-9}	1.54×10^{-9}	1.01×10^{-8}	5.83×10^{-9}	2.90×10^{-11}	3.04×10^{-8}	7.86×10^{-9}	7.58×10^{-10}
	$HI_L(g) = f\rho\sum_\alpha d_\alpha^3 \frac{m_{L_\alpha}}{m_L} \frac{(a_\alpha - a_L)^2}{a_L^2}$		1.16×10^{-2}	1.71×10^{-2}	1.63×10^{-5}	3.10×10^{-2}	3.82×10^{-4}	6.35×10^{-5}	4.03×10^{-2}	9.63×10^{-3}	1.53×10^{-4}
	Mean HI_L (g)		1.23×10^{-2}								

Table B.2 shows that there is a large discrepancy between HI_L values calculated with Equation [4] or [12]. It was observed that trace metal content in each size fraction, a_α, is close to the estimated trace metal content of the lot a_L for all size fractions and all metals (see Table 4 for an example). Therefore, trace metals are distributed amongst all size fractions. This is modeled by Equation [12] in which, as a result, $(a_\alpha - a_L)$ is very small. This distribution of the analyte cannot be adequately modeled by Equation [4] without making debatable assumptions about the liberation factor.

Finally, as an example, Table B.3 presents all HI_L for Cd, calculated as above, for all sampling locations and for a specific state of comminution. It can be seen that there was a significant variability in constitutional heterogeneity between sampling locations. The mean HI_L value was used for recalculating $s^2(THEO)$ for Fig. 3a and b. As it can be seen, using the mean HI_L led to larger and more conservative estimates of $s^2(THEO)$ as HI_L values were smaller at some sampling locations.

References

[1] P.M. Gy, Part IV: 50 Years of Sampling Theory – A Personal History 74, 2004, pp. 49–60.
[2] K.H. Esbensen, Introduction to the Theory and Practice of Sampling, IM Publications, Chichester, 2018, p. 327, https://doi.org/10.1255/978-1-906715-29-8. ISBN: 978-1-906715-29-8.
[3] R.W. Gerlach, D.E. Dobb, G.A. Raab, J.M. Nocerino, Gy Sampling theory in environmental studies I: assessing soil splitting protocols, J. Chemometr. 16 (2002) 321–328, https://doi.org/10.1002/cem.705.
[4] R.W. Gerlach, J.M. Nocerino, C.A. Ramsey, B.C. Venner, Gy sampling theory in environmental studies 2. Subsampling error estimates, Anal. Chim. Acta 490 (2003) 159–168.
[5] B. Svensmark, Extensions to the theory of sampling 1. The extended Gy's formula, the segregation paradox and the fundamental sampling uncertainty (FSU), Anal. Chim. Acta 1187 (2021), 339127, https://doi.org/10.1016/j.aca.2021.339127.
[6] J.P. Boudreault, J.S. Dubé, M. Sona, E. Hardy, Analysis of procedures for sampling contaminated soil using Gy's Sampling Theory and Practice, Sci. Total Environ. 425 (2012) 199–207.
[7] J.S. Dubé, J.P. Boudreault, R. Bost, M. Sona, F. Duhaime, Y. Éthier, Representativeness of laboratory sampling procedures for the analysis of trace metals in soil, Environ. Sci. Pollut. Control Ser. 22 (15) (2015) 11862–11876, https://doi.org/10.1007/s11356-015-4447-1.
[8] R. Minnitt, Calibrating K and alpha in Gy's formula: a new approach, Math. Geosci. 48 (2016) 211–232.
[9] D. François-Bongarçon, Introduction and first ever rigorous derivation of the liberation factor, in: Proc. 7th World Conference of Sampling and Blending (WCSB7), IM Publications, 2015, pp. 165–168, https://doi.org/10.1255/tosf.79.
[10] D. François-Bongarçon, The liberation factor and other parameters: are we at the end of the journey?, in: Proceedings 9th World Conference on Sampling and Blending WCSB9), Beijing, China, 2019, pp. 27–32.
[11] 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, Anal. Chim. Acta 1049 (2019) 47–64.
[12] P.M. Gy, Sampling of Heterogeneous and Dynamic Material Systems: Theories of Heterogeneity, Sampling, and Homogenizing, Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1992.
[13] Danish Standards, Representative Sampling – Horizontal Standard, DS 3077, Danish standards foundation, Copenhagen, 2013. approved August 26, 2013.
[14] F.F. Pitard, Pierre Gy's Sampling Theory and Sampling Practive, second ed., CRC Press, Boca Raton, 1993, p. 488p.
[15] F.F. Pitard, Theory of Sampling and Sampling Practice, third ed., CRC Press, Boca Raton, 2019, p. 691p.
[16] P.M. Gy, Sampling for Analytical Purposes, Wiley, Chichester, UK, 1998.
[17] P.M. Gy, Sampling of discrete materials – a new introduction to the theory of sampling, I. Qualitative approach, Chemometr. Intell. Lab. 74 (2004a) 7–24.
[18] P.M. Gy, Sampling of discrete materials – a new introduction to the theory of sampling, II. Quantitative approach – sampling of zero-dimensional objects, Chemometr. Intell. Lab. 74 (2004b) 25–38.
[19] P.M. Gy, Sampling of discrete materials – a new introduction to the theory of sampling, III. Quantitative approach – sampling of one-dimensional objects, Chemometr. Intell. Lab. 74 (2004c) 39–47.
[20] D.M. François-Bongarçon, P. Gy, The most common error in applying 'Gy's Formula' in the theory of mineral sampling and the history of the liberation factor, The Southern Afr. Inst. Min. Metall. J. 102 (2002) 475–479.
[21] 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, https://doi.org/10.1016/j.aca.2009.08.039.
[22] R.C.A. Minnitt, P.M. Rice, C. Spangenberg, Part 1: understanding the components of the fundamental sampling error: a key to good sampling practice, J. S. Afr. Inst. Min. Metall 107 (2007a) 505–511.
[23] R.C.A. Minnitt, P.M. Rice, C. Spangenberg, Part 2: experimental calibration of sampling parameters K and alpha for Gy's formula by the sampling tree method, J. S. Afr. Inst. Min. Metall 107 (2007b) 513–518.
[24] R.C.A. Minnitt, D. François-Bongarçon, F.F. Pitard, Segregation Free Analysis for calibrating the constants K and α for use in Gy's formula, in: WCSB 5 Sampling. Santiago, Chile, 2011, pp. 133–150.
[25] R.W. Gerlach, J.M. Nocerino, Guidance for Obtaining Representative Laboratory Analytical Subsamples from Particulate Laboratory Samples, U.S. Environmental Protection Agency., 2003, p. 134p. EPA/600/R-03/027.
[26] G. Lyman, R. Schouwstra, Use of the scanning electron microscope to determine the sampling constant and liberation factor for fine minerals, in: WCSB 5 Sampling. Santiago, Chile, 2011, pp. 89–103.
[27] G. Sposito, The Chemistry of Soils, second ed., Oxford University Press, Oxford, New York, United States of America, 2008.
[28] S.D. Young, Chemistry of heavy metals and metalloids in soils, in: B. Alloway (Ed.), Heavy Metals in Soils. Environmental Pollution, vol. 22, Springer, Dordrecht, 2013, https://doi.org/10.1007/978-94-007-4470-7_3.
[29] G.J. Lyman, Application of Gy's Sampling Theory to Coal: a simplified explanation and illustrations of some basic aspects, Int. J. Min. Process. 17 (1986) 1–22.
[30] R.D. Holtz, W.D. Kovacs, T.C. amd Sheahan, An Introduction to Geotechnical Engineering, second ed., Peasron Education Inc., Upper Saddle Rive, NJ, 2011, p. 853p.
[31] J.K. Mitchell, K. Soga, Fundamentals of Soil Behavior, third ed., John Wiley & Sons, Inc., Hoboken, NJ, 2005, p. 652p.
[32] J.A. Lyn, M.H. Ramsey, A.P. Damant, R. Wood, Empirical versus modelling approaches to the estimation of measurement uncertainty caused by primary sampling, Analyst 132 (2007) 1231–1237.
[33] M. Sona, J.-S. Dubé, Sampling particulate matter for analysis - Controlling uncertainty and bias using the theory of sampling, Anal. Chim. Acta 1185 (2021), https://doi.org/10.1016/j.aca.2021.338982.
[34] J.-S. Dubé, M. Sona, J.-P. Boudreault, E. Hardy, Influence of particle size and grinding on measurement of trace metal concentration in urban

anthropogenic soils, J. Environ. Eng. 140 (6) (2014), https://doi.org/10.1061/(ASCE)EE.1943-7870.0000825.
[35] J.-P. Boudreault, et al., Quantification and minimization of uncertainty by geostatistical simulations during the characterization of contaminated sites: 3-D approach to a multi-element contamination, Geoderma 264 (2016) 214–226, https://doi.org/10.1016/j.geoderma.2015.10.019.
[36] Centre d'expertise en analyse environnementale du Québec (CEAEQ), Détermination des métaux : méthode par spectrométrie de masse à source ionisante au plasma d'argon, MA. 200 – Mét. 1.2, révision 7, ministère de l'Environnement et de la Lutte contre les changements climatiques, 2020, p. 18.
[37] Centre d'expertise en analyse environnementale du Québec (CEAEQ), Protocole pour la validation d'une méthode d'analyse en chimie, DR-12-VMC, ministère de l'Environnement et de la Lutte contre les changements climatiques, 2021, p. 35.