A Critique of Gy's Sampling Theory

Dihalu, D.S.
Email: D.S.Dihalu@tudelft.nl

Geelhoed, B.
Email: b.geelhoed@kpnplanet.nl

One of the most used theories for the sampling of materials for physical, chemical or biological testing is the theory developed by Pierre Gy. After a number of scientific publications, including several in the French language (e.g. Gy, 1953, 1964, 1975), he made – in 1979– his entire new theory available to the worldwide sampling community in a book (Gy, 1979) written in English. This book contains a complete description of Gy's sampling theory. Later, Gy has made several refinements, but the essential character of the theory has always remained the same as the theory described in his 1979 book.

The impact of this book (and the entire theory of Gy) has been significant; even nowadays this book is regarded as the number one source of sampling-related information for engineers and process operators. Even though the practical impact and the scientific value of this work are unquestionably strong, several critical points of discussion need to be mentioned here, because the development of new technologies, recent experimental results and novel insights show that parts of Gy's theory need to be updated or revised.

1. Introduction

Since 1979, Gy's theory is increasingly being used in all areas where sampling plays a role. This theory already plays a role in official sampling standards (e.g. some ISO standards, see e.g. Holmes, 2011) and work is in progress to implement it in the near future in even more official standards (see e.g. Esbensen, et al, 2009; Esbensen & Minkkinen, 2011). Of course it is to be welcomed when scientific theories are implemented in standards, if that helps practitioners to incorporate scientifically-based methods in their practice. But at the same time people need to stay aware that implementation in standards can lead to future uncritical and dogmatic application of the scientific theory concerned (Gy's theory in this case). And that would be unwelcome, especially if the theory still needs to be updated or revised.

Moreover, in recent scientific publications there is an increasing trend of implicitly assuming that the term "Gy's theory" would be synonymous for the term "Theory of Sampling (TOS)". After reading such publications, one could get the wrong impression that Gy's theory is already universally accepted by the worldwide academic community, but that is not yet the case.

Throughout the years, many other scientists have also contributed valuable work within the framework of sampling theories (described e.g. by Geelhoed, 2010). This demonstrates that there does not yet exist a universal sampling theory, and that therefore the term "TOS" or "Theory of Sampling" is currently irrelevant if it is related to one of the, yet incomplete, theories. Moreover, Gy has largely worked outside of mainstream academics. Most of his work was published by himself in books outside of the standard peer review process. Gy's theory has therefore not yet been critically examined nor accepted by the worldwide academic community. A factor that may have contributed to bring about this current state of affairs is in the authors' opinion the unnecessary complexity of Gy's theory and the unnecessary long length and complexity of several of the mathematical derivations underlying this theory.

However, all over the world paid experts (for brevity we will call them "consultants", without naming any specifically) are emerging who are consulting, teaching and applying this not yet critically examined and accepted theory. And we see no reason to expect that these consultants would stop doing this if no one puts a stop to it. Instead, we expect that they will implement the theory of Gy wherever they can (get away with it). And once Gy's theory is implemented, these consultants will have more or less secured their job, because they will be needed again and again because no one else within the institute feels to be competent within the area of Gy's sampling theory, and problems are sure to arise. The consultants have no financial incentive to improve or to clarify the theory of Gy, because if this would be done, two things will go against the financial interests of the consultants:

(i) improvement and clarification of the theory to people working in the company would make it unnecessary for the company to hire the consultant in the future because they can then solve problems themselves, and
(ii) improvement and clarification of the theory will also allow people to understand the theory and this will open the theory for criticism. People will then see what the theory really is, instead of just a complex theoretical work outside of their grasp.

In this critique we will attempt stop the premature implementation of Gy's theory, by providing some of the most important of the relevant scientific arguments against this theory in its current state. By doing this, we hope that we stimulate also other sampling scientists to join the combined effort to develop a better, clearer and completer sampling theory.

The main focus of this critique will be on the work that was published by Gy in 1979 and 1982 (Gy 1979; Gy 1982), being the most comprehensive bundle of work shared by Gy in the English language. But the critique is also applicable to the later refinements by Gy and by the adherents of his theory, because the critique focuses on the essential ingredients of Gy's theory, which have never been changed or altered either by Gy or by any of his adherents.

In the next section (section 2), a brief overview of the essential ingredients of Gy's theory is presented. We will make critical comments whenever deemed is necessary. In section 3, we will give a more detailed critique on one specifically important ingredient of Gy's theory, namely the treatment of the fundamental sampling error by Gy.

2. The essential ingredients of Gy's theory

Gy's theory consists roughly of three main parts:

(i) The discrete selection model
(ii) The continuous selection model
(iii) Practical rules for the dimensions and operating speeds of sampling tools.

In the discrete selection model, Gy derives a formula for the sampling variance that is based on a naïve model of the sample drawing process in which independent particle selections are assumed. Gy derives, without giving references to previous work, a basic equation for the sampling variance that was however already in a very similar form derived in 1935 by Kassel and Guy (Kassel and Guy, 1935). Gy subsequently simplifies the basic equation as a product of several factors. But because Gy's model does not have to represent the reality, other factors, which are at the moment considered irrelevant in the theory of Gy and therefore not yet accounted for, may prove to have a strong influence on the sampling variance. Moreover, Gy assigns default values to his parameters, causing the abolishment of performing measurements in many practical scenarios.

In the continuous selection model, Gy's theory attempts to capture or describe the temporal variation of an industrial or natural process. However, Gy only uses the variogram for this purpose, while in the theory of signal processing other tools are more common, notably the Fourier transformation. In order to unify the two completely different models, (that is the continuous and discrete model) Gy assumes that the variogram has a so-called "nugget", which contains the variance components necessary to make the continuous and the discrete model compatible. There is no experimental proof that this is a valid methodology and that both models are indeed compatible. And it has also been argued that the nugget, in one school of thought at least, should always be zero (Clarke, 2009).

With respect to third part of Gy's theory (namely the practical rules for the dimensions and operating speeds of sampling tools) modern results obtained using Discrete Element Modeling (DEM) simulations show that Gy's rules, which are based on approximating the material "flow" by the movement of a single spherical particle, are not necessarily applicable (Cleary et al, 2005; Cleary et al, 2007; Robinson and Cleary, 2009; Robinson and Sinnot, 2011).

3 Gy's Discrete Selection Model

Let's zoom in on Gy's discrete selection model (DSM). Our major point of criticism regarding the DSM lies in the fact that Gy uses empirical correction factors in order to "fudge" the end result towards experimentally obtained values. As an example, we first look at Gy´s treatment of the effect of grouping and segregation on the sampling variance. In his theory, Gy derives a formula (the FSE-equation) with which the variance can be predicted based on the particle properties. But this formula is derived by Gy under the assumption of a model of independent particle selections that does not account for grouping and segregation, because it assumes that each particle is selected independently from the others, whereas in case of grouping and segregation particle selections must be regarded as dependent (see e.g. Geelhoed, 2007).

The formula therefore leads to inaccurate predictions in case of grouping and segregation (see also e.g. Geelhoed et al, 2009; Geelhoed, 2011). In order to still account for grouping and segregation, Gy introduced two factors, namely a "grouping factor" (Y) and a "segregation factor" (Z). Gy then says that the effect of grouping and segregation on the sampling variance is to multiply the variance-prediction based on his formula by the factor (1+YZ). However, Gy provides no method of calculating the factor (1+YZ) in an independent way. That is, in a way other than by comparison of the experimental value with the variance-prediction of the formula. The factor (1+YZ) therefore is a fudge factor.

Gy's formula (FSE-equation below) is a simplification of a more fundamental, but mathematically identical, equation (the latter equation is given on page 232 of Gy's 1979 book). From a mathematical point of view, it is stupefying to observe the similarity of this more fundamental equation with the formula originally developed by Kassel and Guy in 1935 (Kassel&Guy, 1935). We therefore call both equations here the "KG-equation". (In case the minor difference between the Gy-variant and the Kassel and Guy variant becomes important, we here propose the use the names "Gy-variant of the KG-equation" and "Kassel and Guy-variant of the KG-equation". But for the current critique this difference, which is minor, plays no role). The KG-equation furthermore has similarity to a later developed variance estimator (Geelhoed and Glass, 2004) that has the added advantage that it is based on sample information rather than on inaccessible population information.

In his 1979 book, Gy shows how to derive the following result from the KG-equation.

$V = c_{batch}^2 \, f \, g \, L \, c \, D^3 / M_{sample}$ (FSE-equation)

where:

V = the sampling variance
c_{batch} = the concentration of the property of interest in the population
f = the Brunton shape factor
g = the size range factor
L = the liberation factor
c = the mineralogical composition factor
D = the nominal particle size
M_{sample}= the mass (or weight) of a sample.

Because the FSE-equation contains a considerably reduced set of parameters with respect to the KG-equation, one may be tempted to conclude that the FSE-equation requires much less effort to practically apply than the KG-equation (this conclusion is implicitly made by Gy). However, this conclusion would be premature, because it overlooks the effort required to get a numerical value of each parameter. The KG-equation may contain much more parameters than the FSE-equation, but each parameter of the KG-equation can be numerically evaluated with less effort than a typical parameter of the FSE-equation.

The FSE-equation only leads to significant reduction of effort to calculate the sampling variance, when inadmissible procedures to quantify the parameters are applied. And this is what Gy's theory does. For example, typical values –i.e. default values– that the parameters may have, supposedly based on "experience", are given by Gy. But there is no guarantee that a default value is the correct value and there is the risk that the use of default values amounts to the same thing as using fudge factors, if the "default values" are set so that they match a desirable experimentally obtained variance. Gy describes a possible misinterpretation of his work relating to the setting of the L parameter in several publications (Francois-Bongarcon and Gy, 2002), but fails to acknowledge that he is at least partially responsible for the widespread use of inadmissible procedures for calculating the sampling variance and the L parameter.

While Gy pays a lot of attention to the limited selection of parameters in his FSE-equation (namely c_{batch}, f, g, L, c, D, and M_{sample}), he ignores the possibility that other parameters may also have a strong influence on the sampling variance. For example, two major parameters are not included in the FSE-equation: C_{ij} and f_n. From experimental verifications (Geelhoed, et al. 2009) it is concluded that dependent particle selections significantly influence the sampling variance. These dependencies are described by the parameter C_{ij}. The other parameter not considered by Gy is the multi-axial shape factor f_n (Dihalu and Geelhoed 2011). C_{ij} and f_n are just two examples; many alternative parameters could also influence the sampling variance, but are not considered by Gy.

3 Conservatism in Science

The whole setup of Gy's theory, with its lengthy mathematical derivations (which are often unnecessarily complex) and its hidden use of fudge factors, makes this theory unreachable for the average scientist. The number of people who have read and understood the complete sampling theory of Gy, including the basic variance estimator, is small and as a result some of these people have created a nichemarket for themselves on which they can manifest on. Researchers that have to outsource the important but time-consuming topic of sampling will therefore continuously be dependent on the availability and viewpoints of these experts.

From an academic point of view, science should be at least understandable for the majority of interested people in that research area. As Gy pointed out himself, sampling is a multidisciplinary, delicate, but complex topic. In this respect, it is very unfortunate that deep knowledge of his sampling theory is at the moment only present in a small number of people. It has been observed in the past that a conflict of interest arose occasionally whenever a small group of people was hired as expert or consultants for a complex topic. Financial reasons might drive some of these professionals to an unproductive attitude of conservatism, leading e.g. to the preservation of unnecessary complexity of the sampling theory, even when simplifications could be made. Furthermore, this unproductive conservative approach towards the subject will make the whole topic less likeable to be subject to changes or additions due to new insights. But science should always have progressive and dynamic pattern to avoid such stagnation.

4 TOS

Currently, the use of the term "TOS" has been claimed by people who adhere to the theory of Gy, and implicitly use "TOS" as a synonym for "Gy's theory". But "TOS" is not a synonym for "Gy's theory", because Gy's theory is not the only sampling theory of interest. For example, even Gy himself was evidently inspired by the equation presented by Kassel and Guy in 1935. So if "TOS" is not synonym with "Gy's theory", what is TOS?

Utopically, a complete and flawless theory of sampling would be a correct synthesis of the insights of the major scientists that have made their contributions throughout the years. We propose to use the term "Theory of Sampling", or "TOS", for such a perfect and complete to-be developed theory.

5 Conclusions

Our main conclusions are:

- The theory of Gy fails to provide convincing theoretical and/or experimental proof that two of its major theoretical parts, namely the discrete selection model and the continuous selection model, are compatible and unconflicting.

- Gy's rules for the dimensions and operating speeds of sampling tools are based on too simplistic considerations, whereas nowadays more realistic DEM simulation methods are available. When these are applied to the question of sampling tool design, very different conclusions than those of Gy are certainly to be expected.

- The use of fudge factors to tweak the predicted values with the experimental values is a major point of concern in Gy's theory.

- Gy’s theory is at various points in its presentation unnecessarily complex and lack clarity.

- “Theory of Sampling”, or “TOS”, should not be used as a synonym for “Gy’s theory”.

As Pierre Gy indicated himself: “The time when “honest sampling” was made of “good judgment” and “practical experience” is over” (see Gy, 1979, page 8). In that respect, the current aim of sampling scientists should be to develop a sampling theory in which fudge factors –whether they be called “correction factors”, “empirical constants”, or “parameters”– do not play a role.

References

Clarke, I. (2009) Statistics or geostatistics? Sampling error or nugget effect? Fourth World Conference on Samplingg & Blending, Conference Proceedings The Southern African Institute of Mining and Metallurgy, 2009, page 13-18.

Cleary, PW, Robinson, GK, Sinnot, MD (2005) Use of granular flow modelling to investigate possible bias of sample cutters. Second World Conference on Sampling & Blending, conference proceedings, p. 69-81.

Cleary, P, Robinson, GK, Owen, P, Golding, M. (2007) A study of falling-stream cutters using discrete element modeling with non-spherical particles. Third World Conference on Sampling and Blending, conference proceedings, p.17-32. (ISBN 978-85-61155-00-1).

Dihalu DS, Geelhoed B. (2011) A new multi-axial particle shape factor--application to particle sampling. Analyst, 136(18):3783-8.

Esbensen, KH, Paoletti, C, Minkkinen, P. , Pitard, F. (2009) Developing meaningful international strandards – Where do we stand today? The world’s first horizontal (matrix-independent) standard – First foray. Fourth World Conference on Samplingg & Blending, Conference Proceedings The Southern African Institute of Mining and Metallurgy, 2009, page 63-64.

Esbensen, KH, Minkkinen, P (2011) Illustrating sampling standards –How to guarantee complete understanding and TOS-compliance? Fifth World Conference on Sampling & Blending, Conference Proceedings, Gecamin Ltda, p.57-63.

Francois-Bongarcon and Pierre Gy (2002) The most common error in applying ‘Gy’s Formula’ in the theory of mineral sampling, and the history of the liberation factor, The Journal of The South African Institute of Mining and Metallurgy, p.475-479.

Geelhoed, B, Glass, HJ (2004) 'Estimators for particulate sampling derived from a multinomial distribution', Statistica Neerlandica, vol. 58, no. 1, 57-74.

Geelhoed, B, 2007. Variable second-order inclusion probabilities as a tool to predict the sampling variance, Third World Conference on Sampling and Blending, conference proceedings, p.82-89.(ISBN 978-85-61155-00-1), http://arxiv.org/ftp/arxiv/papers/0911/0911.1472.pdf.

Geelhoed, B, Koster-Ammerlaan, MJJ, Kraaijveld, GJC, Bode, P, Dihalu, DS, Cheng, H (2009), An experimental comparison of Gy's sampling model with a more general model for particulate material sampling. Southern African Institute of Mining and Metallurgy, 2009. Symposium Series S59. page 27-38 ISBN: 978-1-920211-29-5.

Geelhoed, B (ed) (2010) Approaches in Material Sampling, Delft University Press, 152 pp.

Geelhoed, B. (2011) Is Gy's formula for the Fundamental Sampling Error accurate? Experimental evidence. Minerals Engineering 2011; 24(2): 169-173.

Gy, P (1953) Erreur commise dans le prelevement d'un echantillonsur un lot de minerai. Congres des Laveries des Mines Metalliques Francaises. R. Ind. Min. 36: 311-345.

Gy, P (1964) Le principe d'equiprobabilite. Ann. Min. (Dec. 1964): 779-794.

Gy, P (1975) Theorie et pratique de l'echantillonage des matieres morcelees. Editions PG, Cannes, France. 597 pp.

Gy P. (1979 & 1982) Sampling of Particulate Material: Theory and Practice. Elsevier: Amsterdam, 431 pp.

Holmes, RJ (2011) Challenges of developing ISO sampling standards. Fifth World Conference on Sampling & Blending, Conference Proceedings, Gecamin Ltda, p.383-392.

Kassel LS, Guy TW (1935) Determining the correct weight of sample in coal sampling. Industrial and Engineering Chemistry Analytical Edition 1935; 7(2):112--115.

Robinson, GK and Cleary, PW (2009) Some investigations of Vezin sampler performance. Fourth World Conference on Samplingg & Blending, Conference Proceedings The Southern African Institute of Mining and Metallurgy, 2009, page 219-229.

Robinson, GK, Sinnot, M. (2011) Discrete element modeling of square cross-belt samplers with baffles. Fifth World Conference on Sampling & Blending, presentation, http://www.sampling2011.com/evento2011/index.php?option=com_content&task=view&id=57