Pierre Gy's development of the Theory of Sampling: a retrospective summary with a didactic tutorial on quantitative sampling of one-dimensional lots

R.C.A. Minnitt[a] and K.H. Esbensen[b]

[a]School of Mining Engineering, University of Witwatersrand, Private Bag 3, WITS. 2050, South Africa. E-mail: Richard.Minnitt@wits.ac.za

[b]Consultant, adj. professor (www.kheconsult.com). E-mail: khe.consult@gmail.com

This paper presents an overview of the pioneering work of Pierre Gy on the Theory of Sampling (TOS) over a period of 66 years—a monumental legacy of dedicated work to the science of sampling. It starts with the early years, beginning in 1949, when Gy worked tirelessly, often in isolation, and not without resistance from other scientists, to create a systematic, mathematically based framework within which every error arising from sampling of a heterogeneous material can be identified (named), analysed and explained. Gy is widely remembered for "The Formula", which expresses the variance of the fundamental sampling error (*FSE*) in terms of the mass, fragment top-size and various other compositional attributes of the minerals of interest. But as early as 1947 the seeds of two related questions were planted: "How... to sample" and "How much... material should one take", which eventually evolved to become the Theory of Sampling, essentially completed in 1975. Presentation of this famous equation to the scientific community led to a period of intense practical experimentation on stationary lots that continues to this day. Gy also addressed the challenge of sampling from fluxes and here identified the variability due to autocorrelation between samples taken from flowing streams that led ultimately to the discipline of *chronostatistics*, a study of variability in one-dimensional streams. The scope and depth of Gy's research endeavours grew during the 1960s and led to his encounter with metallurgical balance and reconciliation. As a result, Gy developed the idea of *proportional sampling* that allows several material streams to be sampled according to the same selection probability with very efficient reconciliation benefits. The TOS was also further extended to include the problems that producers faced in regard to *bed blending*. Gy was a prolific writer and published over 250 different articles and books, with his last four papers published in the *Proceedings of the First World Conference on Sampling and Blending*, WCSB1 (2003), Esbjerg, Denmark. In these papers, Gy left us a personal history of the development of TOS as well as three fundamental tutorials, with practical examples, summarising how to manage both the qualitative and quantitative aspects of sampling of discrete materials. This tribute focuses especially on Pierre Gy's pioneering applications of *variography* to understanding the large-scale variability in process plants and process control from as early as the 1950s, and he devoted a major part of the development period of TOS to this critically important subject. The *variogram* allows one to identify sources of variability and provides valuable insight into correlations between successive samples. Poor understanding of the analytical capabilities of the variogram mean that it has not been widely applied in process control. Failure to address the concept of stream heterogeneity means that conventional statistics and statistical process control (SPC) fail to identify and distinguish the *sources of variability* in a process stream. For each type of heterogeneity, there is a matching variety of process variability. Although the method is powerful in terms of the insights one is able to gain in regard to plant performance and management, there are surprisingly few examples of its application in the literature, although there has been greater acceptance of the method. The authors believe the concluding didactic presentation of Pierre Gy's approach to process sampling may be a useful starting point for newcomers to TOS.

Introduction

The fascinating story of Pierre Gy's interest in sampling and development of the Theory of Sampling is presented in the Proceedings of WCSB1, a collection of papers as a tribute to his work and personal history.[1]

Pierre Gy began his career in French Equatorial Africa (Congo) working on the small M'Fouati lead mine as the Mineral Process Engineer in 1946, where he was in charge of the processing plant and associated laboratories. In 1947 the Paris-based head office asked Pierre to estimate the grade of a 200,000 t, apparently low-grade stockpile that had been dormant since 1940. He soon recognised i) that fragments on the stockpile varied from several tonnes

to fine dust, ii) he knew *nothing* about sampling, iii) there was no meaningful literature available and iv) that he would have to improvise. This request planted the seed of life-long interest in his mind.

On his return to Paris in 1949, his work in a mineral-processing laboratory also constantly brought issues of "sampling" to his attention, in particular the question of "the minimum sample weight necessary to achieve a certain degree of reliability".[1] In his search through the available literature, such as there was, Gy found that Brunton[2] claimed that the minimum sample weight was proportional to the cube of the top particle size, while Richards[3] suggested that the square of the particle size was important. Brunton[2] based his ideas on

the "constant proportionality factor", meaning that for samples with different fragment top sizes, the same number of fragments was required, but Gy[1] was concerned that variations in grade or density had not been properly incorporated.

It was the magnitude of financial transactions in the coal trade based on assays for ash and sulphur in "coal samples" that promoted much of the early research into sampling. Gy tells about UK- and USA-based researchers that "realised that sampling actually generated errors that could have a financial impact", and so began the interest in investigating coal properties in regard to particle top size, sample mass and sample variance. He mentions a Professor Hassialis from Columbia University, New York who

wrote a chapter on sampling based on a statistical multinomial model in the *Mineral Processing Engineer's Bible*, first published in 1927. The number of influencing parameters that were never known meant that this approach could not be practically implemented. A French Mining Engineer, R. Duval proposed a binomial model (in which the world is made up of white and black balls only) representing pure gangue and pure mineral, in which *all* fragments were considered to have the *same* physical mass. While Gy understandably found aspects of this model "dangerously misleading" (*sic*), it germinated the seed of interest sown by his earlier experiences, leading to his 1949 decision to study the theoretical issues around sampling in earnest.[1]

Gy expressed his intention to develop a mathematical model relating the variance of the sampling error to the mass of the lot, the sample mass and the *knowable* physical properties of the material being sampled. Such a relationship would allow the minimum sample mass needed to achieve an acceptable sampling variance to be determined. Gy's hopes of addressing the question "how much" had to be pursued in his own off-work time, as his employer provided neither time nor resources for this research. Surprisingly, even with these obstacles, he devised and wrote up the formula and the basic tenets of the TOS in two internal, unpublished notes for his company *Ste Minerais et Metaux*, entitled: "A formula for the minimum sample mass" and "Minimum sample mass required to represent a batch of ore" as early as 1950. Historically the TOS was born in 1950.

This endeavour led to a first theoretical model specifically for *particulate solids*, but generalised models for solids of animal and vegetable origin, types of domestic and industrial waste, liquids and gasses were also developed. By this stage Gy recognised that the models had *universal* validity and that it was *scale* rather than physical state that differentiated between the range of applications.

The Formula established

The progression in Gy's logic in regard to formulating the variance model, as early as 1950, is fascinating. He first identified all the unknown, but physically well-defined parameters, including the number of N_L fragments making up the lot, the corresponding N_S fragments making up the sample, a_S the grade of the sample, F_i the number of fragments and M_i the individual fragment mass. From these he devised strict, algebraically simple mathematical relationships into which he introduced *simplifications* and *approximations* to produce easily implementable and practical formulae. He first devised formulae for the mean and the variance of a population of "equally probable samples of N_S fragments". At some point in this work, he realised that he needed to educate himself more properly in statistics—and only a few years later he was awarded his second PhD—et voilà! This was necessary in order to able to work more stringently with the crucial approximation simplifications of the full mathematical descriptions.

In this context, among his most germane ideas was the concept of a quantitative measure of *heterogeneity*, which "lies at the root of all sampling errors", which in Equation 1 can be introduced as:

$$h_i \frac{(a_i - a_L)}{a_L} \times \frac{M_i}{M_i^*} \qquad (1)$$

where h_i represents the constitutional heterogeneity carried by one fragment F_i in the lot L, M_i and M_i^* are the individual and the average mass of all fragments. The importance of h_i is that it expresses the *heterogeneity contribution* carried by each fragment in the lot, which crucially can be summed up and, when divided by the number of fragments in the lot and given an appropriate statistical weight, leads to the desired approximate measure of the variance of the total sampling error (*TSE*), as shown in Equation 2.

$$\sigma_{TSE}^2 = \left(\frac{1}{N_S} - \frac{1}{N_L}\right) \bullet \frac{\sum_{i=1}^{N_L} h_i^2}{N_L} \qquad (2)$$

After a very long process of trial and error (many years) in which he tested out a plethora of simplifications and approximations for correlating the sampling error to the physical properties of the lot material (this is where Pierre Gy decided his second PhD in statistics was essential), he arrived at the by now well-known general equation which is publically referred to as "Gy's formula" (but which he strongly preferred to call the "Formula" only),[1] shown in Equation 3.

$$\sigma_{FSE}^2 =\sim \left(\frac{1}{M_S} - \frac{1}{M_L}\right) \bullet cfg\ell d^3$$
$$=\sim \frac{cfg\ell d^3}{M_S} \quad \text{when } M_S \ll M_L \qquad (3)$$

Here c is the *mineralogical composition factor* having units of (but not the meaning of) specific gravity (g cm^{-3}), taking the average grade and densities of all components 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 grade of the increasingly more heterogeneously distributed phase. ℓ is a dimension-less *liberation factor*, defined by Francois-Bongarcon[4] as $\ell = (d_\ell / d_N)^{0.5}$ that varies between 0 for completely liberated components and 1 for non-liberated ores. f is a dimension-less *particle shape* factor having a general value of 0.5 (a kind of early "mission impossible" trying to input quantitative information in the formula representing the principal *form* of the mineral phase of interest with a simple number in the interval [0,1]) and g is a dimension-less *size range factor* with a general value of 0.25; this latter is a measure of the *sorting* of the material being sampled. d is the top particle size in centimetres through which 95% of the material passes, aka d_{95} (the *cube* of the top particle diameter won out—not the square, see re. Brunton vs Richards above).

This early explanation was followed by a more elegant proof in Gy[5] in which *TSE* is generated by each element of the lot being submitted to the selection process (extraction), the *sampling*, with a certain selection probability P_m. In this case *TSE* is now to be understood as the sum of i) the correct sampling error (*CSE*), only related to material composition and heterogeneity, and ii) the incorrect sampling error (*ISE*), specifically related to sample extraction and materials handling (i.e. the errors produced by the sampling process itself, if not effectively mitigated), such that $TSE = CSE + ISE$ from which two cases arose.

If sampling is *correct*, $P_m = P = \text{constant}$ and $ISE = 0$, meaning that $TSE = CSE$, and if in addition elements are selected individually and independently, then $TSE = CSE = FSE$. However, in practice the condition that samples be collected individually and independently is never satisfied, the best we can do is extract *groups* of neighbouring fragments (groups = increments) with uniform selection probability P. The likelihood of a spatial correlation between selected fragments in the spatial

volume of the lot generates a new error referred to as the grouping and segregation error (*GSE*) where:

$$TSE = CSE = FSE + GSE$$

Historically all manner of adverse problems have since manifested themselves whenever people applying Gy's formula do not take proper account of *GSE*—scores of frustrated *rapporteurs* lament that the level of total sampling errors estimated are usually larger (not seldom much larger) than what is indicated by var(*FSE*). This is all due to missing out on understanding and managing (reducing, eliminating as much as possible) *GSE*.

Further Gy[1] also makes the key point that using the formula to determine the sampling variance without appreciation of the possible crucial sampling bias (i.e. the gamut of all incorrect sampling errors) would render straightforward application of the formula *meaningless*. This is a point very well remembered. This is where all serious sampling starts, lest all possibilities of representativity are lost.

Practical experimentation with the Formula

Gy[1] tells the story of how he originally attempted to validate the formula by calculating the variance of a lead ore using 16 "equally split" samples of pulverised material, splitting here taking the role of sampling. His experimental *TSE* was several times *larger* than the theoretical value, something he interpreted as indicating that the *FSE* was only *one* of several components in the game. The other components of sampling error he suggested were the *GSE* as well as the sampling bias, see above, introduced through incorrect use of the riffle splitter. His research in the mid-1950s then led to the development of a circular cardboard sampling nomogram and later a sampling slide rule. The formula was first presented in English to the Society of Mining Engineers of the American Institute of Mining Engineers (SME of AIME) in 1957. However, it was only in 1965 that his research was presented in London at a meeting of the Institution of Mining and Metallurgy (IMM).

Sampling of flowing streams

Pierre Gy's 1960–1962 research into flowing streams of materials on conveyor belts and liquid launders brought to his attention the importance of sampling the "whole stream" for a fraction of the time, i.e. any increment must be a physical full slice of the stream. He identified the key issues in regard to cross-stream sampler operations, namely that the cutter velocity through the stream, the width of the cutter opening and the shape of the cutter are all-important, but it was first in 1977 that these issues were scientifically resolved.

He also recognised that increments extracted at constant intervals from a flowing stream are not independent of one another, but that some level of *auto-correlation* exists between most time series sample data. As early as 1962 Gy started published his work on *chronostatistics*, as it later became known, by borrowing the idea of spatial correlation between samples using concepts and data from the semi-variogram feature proposed by Matheron,[6] and later by David,[7] within geostatistics and transferring it to linear auto-correlation of time series data.

At this stage of his life, Gy made the choice to dedicate himself to writing and further research around the theory and practice of sampling, rather than to continue in his rather comfortable managerial position at Minerais et Metaux in Paris. This led him to what history now recognises as a grand 40-year period of theoretical research, consulting, trouble-shooting, lecturing and teaching regular courses at schools and Universities, and writing articles and books gradually being disseminated all over the world. The reader is encouraged to peruse Pierre Gy's complete bibliography.[1]

Theory of Sampling introduced—and challenged

This time of progressive successes was not without serious challenges, however, as some parties and individuals strongly opposed Gy's ideas and objected to his 1967 publication in French, "Sampling of Particulate Materials".[8] No story is only about success—it is a sad historical fact that the response from ISO standards committees has been less than unanimously accepting of the work and insights of Pierre Gy (although this situation has begun to be significantly turned around since 2003 by a dedicated effort by the sampling community). The world now has at its disposition a first standard dedicated to the universal principles of representative sampling, DS 3077.[9] This last part of the history of TOS can be followed in detail in the proceedings from the WCSB conferences and in *TOS Forum*.

Interestingly, the notion of correct sampling and its linkages to probabilistic sampling were only first proposed by Gy in 1972. In modern parlance, the fundamental tenet is that a sample is correct if and only if each lot fragment has the same statistical probability of being selected for the sample as every other fragment in the entire lot. Under any other circumstance, the sampling procedure is said to be incorrect and will therefore result in unrepresentative lot "samples" (better designated "specimens" for optimal distinction).

About this time Gy found that some members of the scientific community resisted his ideas about sampling as a scientific endeavour. His 1971 book entitled *Sampling of Particulate Materials, Volume 2* was soon followed by another book *The Theory and Practice of the Sampling of Particulate Materials* in 1975, but only a few hundred copies were ever sold. In this particular book, Gy made a very significant step in that he built "the mathematical bridge between selecting conditions and sampling errors". He identified for the first time, the distinction between *a priori* conditions of sample selection (conditions we can do something about before taking the sample), and *a posteriori* conditions (conditions we can observe, but about which we can do very little after the fact). The selection process itself can further be either probabilistic or non-probabilistic—and even if probabilistic, it can be correct or incorrect. Sampling errors are random errors, characterised by their statistical distribution and moments. Sampling can be accurate or biased (property of the *mean*), reproducible or not (property of the *variance*), and representative or not (property of the mean-squared error).

Pierre Gy also tells of the difficulties he faced in 1978–1979 writing his first book in English, a translation of this seminal 1975 text. The book, published in 1979, was followed by a second edition in 1982.

Between the release of the Second Edition of the 1979 text and his latest book in French,[10] Gy developed a number of new applications of his theory including the computation of auxiliary functions of the variogram, the ideas underlying proportional sampling and a theory of bed-blending.

Proportional sampling

Gy's first encounter with metallurgical balance reconciliation was in some North

African lead-zinc flotation plants where he summarised the idea of balance saying that "whatever comes in must ultimately come out, one way or another". He noted that if this principle of balance is not observed, then there must be "measurement biases or unsuspected losses", and that with a single exception in his 45 years of consulting, what came out was always *less* than what went in.[1] Eventually, after checking every sampling and measurement device, he reached the conclusion that the principal culprit for the 2–3% deficit was the calibration of the conveyor belt scales. After observing numerous conveyor belts over the years, Gy concluded that they suffer from a structural lack of reliability, the main problem being the conversion of an electrical current into an accurate measurement of tonnes of ore. Rather than the 0.5% accuracy *claimed* by manufacturers, plant personnel found a more realistic figure to be about 10% deviation from accuracy.

In view of the importance of proportional sampling to metallurgical balances and based on the excellent exposé of this subject provided by Wavrer,[11] the explanation that follows here is somewhat detailed. During his development of the theory of sampling Gy concluded that if the probability of selection P is a uniform distribution in time, then sampling is correct, and the mean of the sample mass M_S is a random variable equal to P times the mass of the lot M_L (Equation 4).

$$m(M_S) = P \times M_L \qquad (4)$$

The corollary is that accurate estimation of P means that M_S/P is an unbiased estimator of the mass of the lot, M_L. If the number of increments is large, the minimum and maximum values that the sample mass could take with given sampling equipment is very accurately known, and confidence limits for these sample masses, M_S, are very small. Proportional sampling must not only be correct, but the mass and volume of the sample must be proportional to the mass and volume of the lot. Thus all concentrates, tailings and feed streams, sampled according to the same proportional ratio (selection probability), make the proportionality factor constant. According to Gy,[1] M_S/P is a much more reliable *unbiased* estimator of the mass of the lot M_L, than any that can be obtained by weightometers, and this became the basis of his revolutionary idea of *proportional sampling*. Wavrer's

simple and elegant explanations[11] are presented in Table 1. Gy defined a time sampling ratio and a mass sampling ratio, for which equations are presented in Table 1. All feed streams, concentrates and tailings are to be sampled according to the *same* proportional ratio (selection probability).

In this way sampling from all material streams are now completely *comparable*, making the calculation of the material balance a simple and very accurate task. The critical success factor is that the selection probability is kept constant under all circumstances.

Bed blending

Perhaps the most important aspect of feeding a metallurgical furnace is to blend the raw materials in such a way that the average composition of the feed will be more-or-less uniform and homogenous in the one dimension of the ingoing material stream. Gy's work on *bed blending* began with a study of material processed in cement kilns. The lack of flexibility and sensitivity of cement kilns is such that feed materials must be as uniform as absolutely possible to avoid costly damage. For this reason, a large cement company introduced a bed blending system in order to homogenise, as best as possible, the ingoing raw materials. Good sampling equipment aided by on-line analysers allowed major components in the cement to be determined every few minutes. Computerised assistance to calculate the average composition of the stockpiled kiln feed allowed the composition of the blending pile to be known with accuracy, providing an almost ideal feed to the kiln.[1]

On one occasion the failure of the blending system to comply with Gerstel's theory,[12] Gy was asked to advice on the process and

found that the bed blending theory was easily derived from existing sampling theory. In this case the manufacturer received an excellent explanation of how his equipment actually worked. Other cement producers approached Gy sometime later when it was proved that his new theory on bed blending was in perfect agreement with practice. The theory and practical aspects were published in 1981,[13,14] with a presentation on the subject to the Canadian Institute of Mining and Metallurgy (CIM) in Montreal.[15,16] Theories that Gy published over the years have consistently proved to be correct, and were easily adapted to the science of bed blending.

Gy's publications

It is not possible to tell Gy's story of discovery without at the same time telling what and where he published over 250 scientific books and papers on the TOS. His last textbook publication: *Heterogeneite, Echantillonnage, Homogeneisation* (Heterogeneity, Sampling, Homogenising), published in 1988 in French, was immediately translated into English and was published in 1992. It was the French version of this book that Dr Francis Pitard digested and shortened to produce his volume entitled: *Pierre Gy's Sampling Theory and Sampling Practice, Heterogeneity, Sampling Correctness and Statistical Process Control*. It is the Second Edition of this latter volume that has become a world famous publication used by many practitioners and now also taught in leading universities.[17]

In 1999, Allen Royle performed an English translation of Gy's 1996 text originally entitled *Echantillonnage pour Analyse Granulometrique* (Sampling for a size analysis). Royle's translated 150-page, 1998 English

Table 1. Wavrer's derivation for proportional sampling.[11]

Explanation	Equation	Parameters
"Time sampling ratio" of a lot, τ'	$\tau' = \dfrac{QT_i}{T_L}$	T_L is the flow time of the lot L Q is the number of increments between $t=0$ and $t=T_L$ T_i is the time to take one increment
"Mass sampling ratio" of a lot, τ	$\tau = \dfrac{M_S}{M_L}$	M_S is the mass of the sample S M_L is the mass of the lot L
If sampling is to be correct	$\tau' = m(\tau)$	$m(\tau)$ is the mean of τ
Rearranging	$\tau' = \dfrac{m(M_S)}{M_L}$ $\qquad M_L = \dfrac{m(M_S)}{\tau'}$	

version was entitled *Sampling for Analytical Purposes* and was published by John Wiley, with a second edition in June 1999.[18] This has become a cherished avenue for newcomers into the sampling world. Pierre Gy was forever grateful to Royle for his help at a critical time, so much so that Gy *insisted* that the first Pierre Gy Sampling Gold medal was to be awarded to him. Allen Graham ("Bon") Royle was honoured by the TOS community in an obituary in *TOS Forum* Issue 1.

The WCSB1 in 2003 and the Proceedings of that meeting[19] were in honour of and a tribute to Pierre Gy's whole life and his work on the TOS. This volume contains 24 articles covering a very broad swath of the breadth and depth of the TOS as of 2003, five of which Pierre Gy wrote himself. (The 137 attendees to this first ever sampling world conference represented pretty much 90% of the then active individuals in the whole world.) This volume was to be the first in a series of Proceedings, the eighth will be from the Eighth World Conference on Sampling and Blending (WCSB8) being held in Perth in May 2017. The series of WCSB Proceedings is indispensable for anyone wanting to get into the theory and practice of sampling.

Summary of Pierre Gy's last contributions

Pierre Gy's last technical contributions to the sampling fraternity came in the form of a three-part theory and practice tutorial entitled "Sampling of discrete materials I, II, III" in the WCSB1 Proceedings.[19] While it was not possible for him to cover all aspects of sampling, these three papers nevertheless represent a formidable distillation of more than 50 years' work. Together with the accompanying feature article,[1] "Theory of Sampling—a personal history", these papers rank amongst some of the most important contributions from Gy, and deserve our full attention a.o. because he took the pains to rectify the admittedly bewildering, often changing three-letter-abbreviations (TLA) that have always characterised TOS, while laying down his final resumé of logical development history of the *theory* of sampling. These three tutorials specifically address the two fundamental questions that inspired Gy's original dedication in 1949, questions which cannot be dissociated from one another: "*How* should one select a representative sample? and *How much* material should be selected for this purpose?". Part I covers the foundation for the qualitative approach, and Parts II and III the quantitative approach to answering these two questions.

Part I: A new introduction to the TOS—qualitative approach

Part I is an up-to-date introduction to the TOS in which Gy explains the seriousness of good sampling science and technique and how without it, not only money, but also lives could be at stake. He set forth the basic definitions and notations, drawing a distinction between samples (representative) and specimens (worthless), showing how the sampling conditions and the proper definitions are strongly related. Explanations of the concepts of constitutional and spatially distributed heterogeneity (*CH*, *DH*)

are followed by outlining the processes and methods of mass reduction, such as grab sampling, splitting and incremental sampling that again distinguish samples from specimens. Gy[8] summarised Part I by saying that sampling is a science, and while the TOS, which explains the generation of errors and proposes practical solutions, *may* be contested by those who wish, it can never be ignored. The inadequacy of most standards dedicated to or containing sections pertaining to sampling even today is because they fail to recognise TOS, while bias in other standards is deliberate (*sic*). TOS is the only scientifically consistent and exact means of extracting correct, unbiased, accurate and reliable samples. Non-probabilistic or incorrect probabilistic sampling will always produce biased and unreliable specimens, making meaningful decisions impossible.

The taxonomy and nomenclature of sampling errors is admittedly somewhat complex, but Gy[8] presents ultimate definitions of errors arising from sampling in Part I of his tutorials. Sampling consists of progressive stages of comminution and representative mass reduction of the original lot *L* to produce sub-samples. The primary sampling from any lot is very often accomplished by problem-specific composite sampling (often comminution is not possible at the primary stage). Errors at this stage are the primary sampling errors (*PSE*). Secondary sampling of the primary sample typically occurs in the laboratory, and here typically consists of progressive comminution and representative mass reduction of sub-samples, ending up with an *aliquot a_L* for analysis. Errors at this stage are secondary sampling errors (*SSE*). The combination of *PSE* and *SSE* errors give rise to the total sampling error ($TSE = PSE + SSE$). Combining *TSE* with the total analytical error (*TAE*), produces what Gy refers to as the global estimation error (*GEE*).

Part II: Quantitative approach—sampling of zero-dimensional objects

Parts II and III cover the quantitative approach and the sampling of zero-dimensional and one-dimensional objects. Part II introduces the concept of the sampling dimensions of the lot, stockpiles being typical of 3D lots, open-pit operations being typical of 2D lots, material on a conveyor belt being typical of a 1D lot, while discrete, independent, easily accessible objects lending themselves to easy *manipulation* (mixing, sub-sampling) are described as zero-dimensional lots. Gy[20,21] only deals with sampling of zero and one-dimensional lots.

Gy theoretically defines a lot made up of discrete material, *L*, as consisting of a set of units (potential sampling increments), the sets being either a population of non-ordered, zero-dimensional units (as in a stockpile), or as ordered, chronological cross-sections of one-dimensional units in a flowing stream of material on a conveyor belt (one-dimensional lot). In general, the one-dimensional units are characterised by a correlation between the position of the unit in the series and their composition. Such autocorrelation requires a very different mathematical approach to identifying and quantifying sampling errors compared to zero-dimensional lots, and hence the sub-division into Part II and Part III of Gy's papers.[20,21] While Part I covered the general concepts of heterogeneity and homogeneity in detail, Gy devotes Part II[21] to explaining the mathematical relationships between sampling errors as a result of the salient aspects of heterogeneity.

Gy provides a progressive mathematical basis for a model describing the contribution of heterogeneity to the lot by an unspecified unit, U_m, and uses this to define constitutional heterogeneity, CH_L in the case when U_m is a single fragment F_i. He then expands

this by alternatively viewing the unit, U_m, now as a *group*, G_n, of neighbouring fragments F_i to define the distributional heterogeneity DH_L of lot L, and further describes the relationship between CH_L and DH_L. In these steps Gy mathematically quantifies the various forms of heterogeneity and then expresses the sampling errors in terms of their mean, variance and mean squared errors. Part II[22] also describes the heterogeneity invariant HI_L before examining the notion and nomenclature around zero-dimensional probabilistic sampling and the related topics of correct and incorrect sampling.

Gy then describes a method of experimental estimation of what is now commonly referred to as the *heterogeneity test*, providing three practical examples

of implementation of this technique from industry using lateritic iron ores, fragments of precious metal ore and sphalerite flotation concentrate pellets. Gy[21] finally makes a number of important summary statements about sampling in zero-dimensions iterating the most fundamental understanding that "sampling errors are the consequence of one form or another of heterogeneity". He notes that sampling of homogenous materials would be an exact operation according to the definition of homogeneity, but that it is *never* observed in the real world. The simplest way to convey the progression in Gy's layout of the quantitative approach to sampling zero-dimensional objects is to use the summary that he provided (p. 37), describing

the relevant equations in support of these concepts, shown in Table 2.

Gy defines the *TSE* as the sum of only two components, i) the *CSE* and ii) the *ISE*, for which the probability of selection can be uniform or non-uniform.

Selection with uniform probability will only incur the correct sampling error (*CSE*) arising from the constitutional and distributional heterogeneity of the material, which give rise to the fundamental sampling error (*FSE*) and the grouping and segregation errors (*GSE*), respectively. Selection with non-uniform probability will incur additional, bias-generated incorrect sampling errors (*ISE*), which are a combination of the incorrect delimitation error (*IDE*), the incorrect extraction error (*IEE*) and the incorrect preparation error

Table 2. Concepts and supporting equations for the quantitative approach to sampling of zero-dimensional objects according to Gy.[21]

Relevant concepts	Supporting equations
The contribution h_m of a given unit U_m to the heterogeneity of the set L of units. Unit U_m can be either a single constituent F_i or a group G_n of constituents such as an increment I. The heterogeneity contribution h_m is a function of the mass and composition of unit U_m and lot L.	$h_m = \dfrac{(a_i - a_L)}{a_L} \times \dfrac{M_m}{M_m^*}$
The constitutional heterogeneity, CH_L, of lot L considered as a population of single elements. CH_L is the variance of the corresponding population of h_i.	$CH_L = \sigma^2(h_i) = \dfrac{1}{N_F}\sum_i h_i^2$ and $\sum_i h_i = 0$
The heterogeneity invariant, HI_L, derived from CH_L for practical purposes and usage.	$HI_L = CH_L \dfrac{M_L}{N_F}$
The distributional heterogeneity, DH_L, of lot L considered as a population of groups of neighbouring elements. DH_L is the variance of the corresponding population of h.	$DH_L = \sigma^2(h_n) = \dfrac{1}{N_G}\sum_n h_n^2$ and $\sum_n h_n = 0$
The *TSE* generated when selecting constituents in a probabilistic way (non-probabilistic sampling cannot be analysed theoretically).	$TSE = \dfrac{a_S - a_L}{a_L}$
TSE is disaggregated into the sum of two components, *CSE* and *ISE*. Additional *ISE*s are observed when the sampling is incorrect.	$TSE = CSE + ISE$
The *CSE*s, observed when the sampling is correct; there is a first and second approximation.	$m(CSE)_1 = 0$ $m(CSE)_2 = -\dfrac{1-P}{P}\sum_m \dfrac{a_m - a_L}{a_L} \times \dfrac{M_m^2}{M_L^2}$
The *FSE* is the *CSE*, observed in ideal conditions, when the constituents are selected correctly, one by one and independently.	$\sigma_{FSE}^2 = \left(\dfrac{1}{M_S} - \dfrac{1}{M_L}\right)\bullet cfg\ell d^3$
The variance of *FSE* is proportional to the CH_L, and, in practical applications, to the HI_L.	$\sigma_{FSE}^2 = \left(\dfrac{1}{M_S} - \dfrac{1}{M_L}\right)\bullet HI_L$
A practical, experimental method to estimate HI_L and hence the variance of *FSE*.	$HI_S = \sum_i \dfrac{(a_i - a_S)^2}{a_S^2} \times \dfrac{M_i^2}{M_S}$
The *GSE* is the additional error generated when selecting constituents with a uniform probability P, by groups (increments) of non-independent constituents. The variance of *GSE* is proportional to the DH_L.	$\sigma_{GSE}^2 \propto DH_L$

(*IPE*) (sometimes also including an incorrect weighing error, *IWE*). *IDE* results from incorrect geometrical delimitation of the physical three-dimensional increments, *IEE* result from incorrect extraction of material increments from the delineated increments and *IPE* results from incorrect preparation and handling of material increments and samples after having been extracted (correctly or incorrectly). All these latter errors arise from ill-informed and/or poor mechanical design, or improper use, or poorly maintained sampling equipment, or improper handling of the samples after extraction. In this case the *TSE* is given by:

$$TSE = CSE + ISE$$
$$= FSE + GSE + ISE + IEE + IPE.$$

Quantitative approach—sampling of one-dimensional objects

Pierre Gy pioneered applications of *variography* to understanding the large-scale variability in process plants and process control from as early as the 1950s and devoted a major part of the remainder of his TOS development period to this subject. The variogram allows one to identify *sources of variability* and provides valuable insight into correlations between successive samples. Poor understanding of the data analytical capabilities of the variogram means that it has not been widely applied in process control up until this day, except in the industry sectors which have embraced TOS (mining, cement and certain parts of process industries). Failure to address the concept of stream heterogeneity means that conventional statistics and statistical process control (SPC) fail to identify and distinguish the *sources of variability* in a process stream.[22] For each type of heterogeneity, there is a matching variety of process variability. Although the method is powerful in terms of the insights one is able to gain in regard to plant performance and management, examples of the application of this particular method is surprisingly absent in the literature.

Process variability

Large-scale variability, usually in the form of product composition cycles and chemical cyclic variations may be manifest throughout a sampling system or a process plant, but this type of variability is most easily discernible in one-dimensional process streams. In terms of the nomenclature of errors arising from the sampling of one-dimensional objects, Gy[21] considers a flowing stream of material or lot *L* as a sequence of zero-dimensional increments with adjacent potential increments centred on point increments. One-dimensional sampling is a two-step process, which may be correct or incorrect. It consists, first, of sample selection at a correct or incorrect *position* on a time axis of immaterial, point-increments, giving rise to the point selection error (*PSE*). Second, the *extraction* of material increments at the point increments of materialisation (the sampling event by cross-stream or cross-belt sampling), giving rise to the materialisation selection error (*MSE*). *MSE* is in turn the sum of only two components, the correct sampling error (*CSE*) and the incorrect sampling error (*ISE*), whereas the *TSE* is:

$$TSE = PSE + CSE + ISE = PSE + (FSE + GSE) + (IDE + IEE + IPE).$$

The overall contribution of the three main sources of variability, random error *V*[0], process error *V*[1] and cyclical error *V*[cyclic] can conveniently be represented, for example, in a pie diagram.

Examples of one-dimensional lots include moving aggregate mixtures on conveyor belts, liquids, suspensions and slurries in pipes or chutes. Pitard[23] explained that the

components of variability around a targeted average are the integrated accumulation of heterogeneity arising from three sources, and expressed as: $h_T = h_1 + h_2 + h_3$, where:

Heterogeneity h_T = total heterogeneity, Heterogeneity h_1 = random, discontinuous heterogeneity that is a property of the materials, Heterogeneity h_2 = non-random, continuous heterogeneity that is a function of time, Heterogeneity h_3 = cyclic, continuous heterogeneity that is a mechanical function of the system.

The variability, and by analogy the heterogeneity, in a moving stream of material can best be represented by a variogram, a model for which is shown in Figure 2.

Different sample intervals in the stream provide the data for calculating the variability between samples. In the first step, the variance between the closest consecutive samples is calculated and averaged according to Equation 5.

$$\gamma_1^2 = \frac{1}{2N} \sum (a_i - a_j)^2 \qquad (5)$$

This provides the first data point on the variogram. The second step is to calculate and average the variance between every second sample, and then every third sample and so on, giving rise to a typical variogram in Figure 2. The relative variogram simply divides the average variance for each point in the variogram by the mean
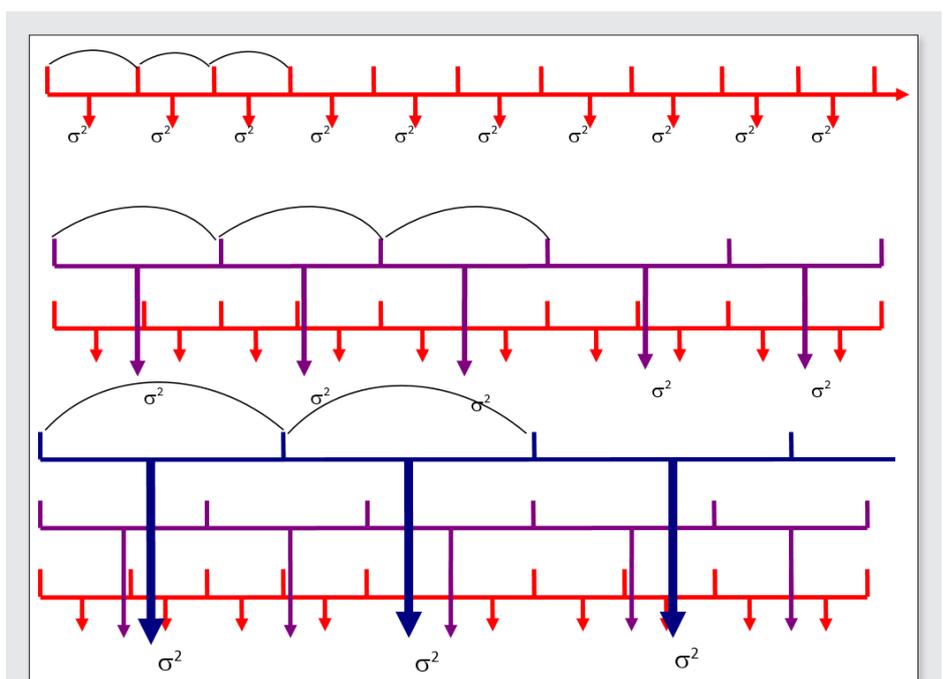

Figure 1. Top row: A series of sample points separated by distance *j* (the lag). Second row: a series of points separated by distance 2*j*. Third row: a series of points separated by distance 3*j* etc.
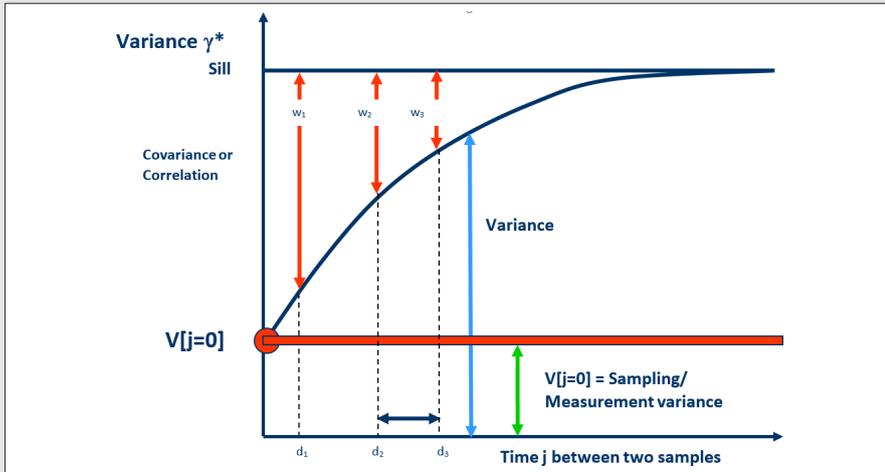
Figure 2. The components of the variogram indicating $V[j=0]$, commonly referred to as the nugget effect.

grade squared. The same effect, but also mass-corrected, is achieved by basing the entire variogram calculation on heterogeneity contributions. The difference between the absolute and relative variograms is only in the way they are calculated.

Absolute and relative variogram based on concentration values:

$$V(j) = \frac{1}{2(N_u - j)a_L^2} \sum_m \left[a_{m+j} - a_m\right]^2$$

Relative variogram based exclusively on heterogeneity contributions:

$$V(j) = \frac{1}{2(N_u - j)} \sum_m \left[h_{m+j} - h_m\right]^2$$

A modified set of 445 normally distributed data from a process plant is used below as an example of the principles and application of variography in process control.

The *moving average* allows one to identify small- and large-scale variability as well as cyclical behaviour in the process streams, while the variogram is a custom-built tool that allows components of variability to be identified and resolved. The components of variance are read off the variogram at specific line intersections. Variances are converted to standard deviations and plotted on the control charts, allowing an operator to manage the upper and lower control and specification limits.

Hydrometallurgical process plant example

Chemical variations in plant systems shown on the variability plot in Figure 3 provide perfect understanding in hindsight, but this information is by itself of only little use to a superintendent trying to stabilise plant variability.

At settings of 8.7% X and 9.28% X, several samples lie outside the upper and lower specification limits (Figure 4), indicating the difficulty of maintaining plant stability. Widening specification limits may help, but the erratic variation suggests the system is easily subject to overcorrection.

Large-scale variability: the moving average

Large-scale, cyclical variability is identified in the sampling data using a wide moving average window. Such cycles may have a regular period, but irregular amplitude, as in the case of metal% X (Figure 5). A 30-point moving average shown in Figure 5 emphasises the strong 118-hour cycle with numerous smaller superimposed cycles.

A 5-point moving average over the same 445-data (Figure 6) illustrates the much smaller scale cyclical behaviour in the data. These cycles have an average period of 14 h, but periods could vary from 10 h to 16 h.

Absolute variogram

The absolute variogram in Figure 7 was calculated for a 445-hour period although only 357 lags (hours) are shown on the x-axis. Three distinct cycles, each having cycles of 118-hour periods are evident in the variogram.

Although the periods are strongly regular, the amplitude of the cycles in the variogram is irregular.

Sources of variability

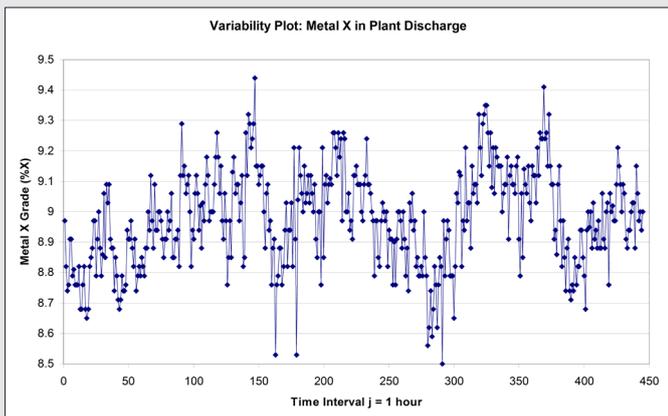The following points are evident in the variogram (Figure 7) and are summarised in Table 3.

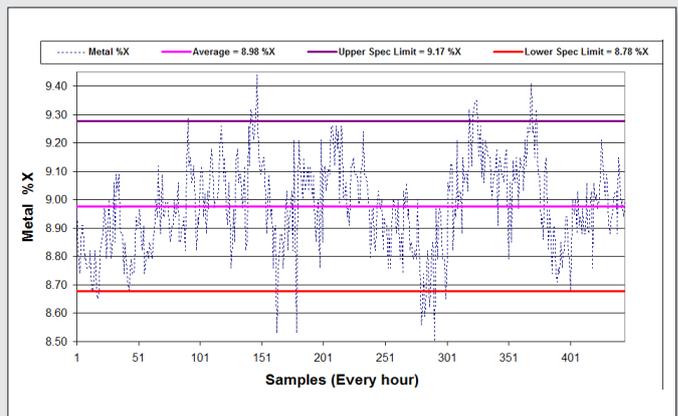**Figure 3.** Variability plot of metal% X with highly variable sample data.

Figure 4. Variability plot of metal% X showing the average, as well as upper and lower specification limits.
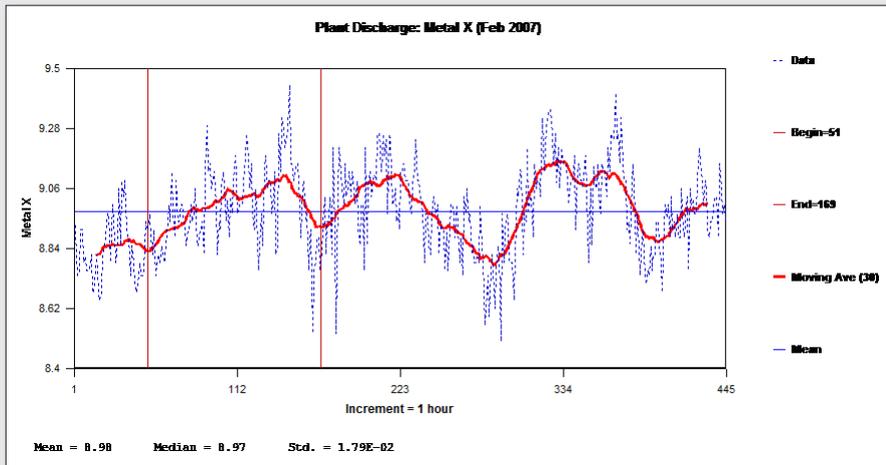
Figure 5. 30-point moving average showing four major cycles in the 445-data, with a 118-hour period (5-day period) superimposed on the data.
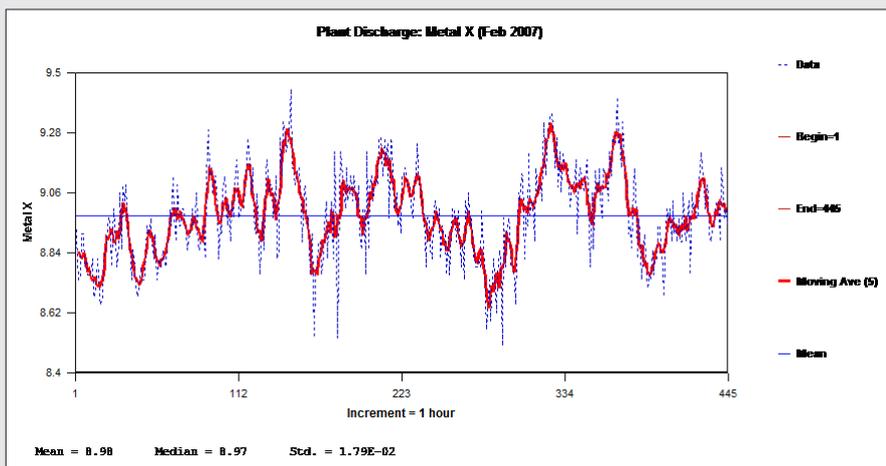

Figure 6. 5-point moving average showing the fine detail in the large scale 118-hour cycles.
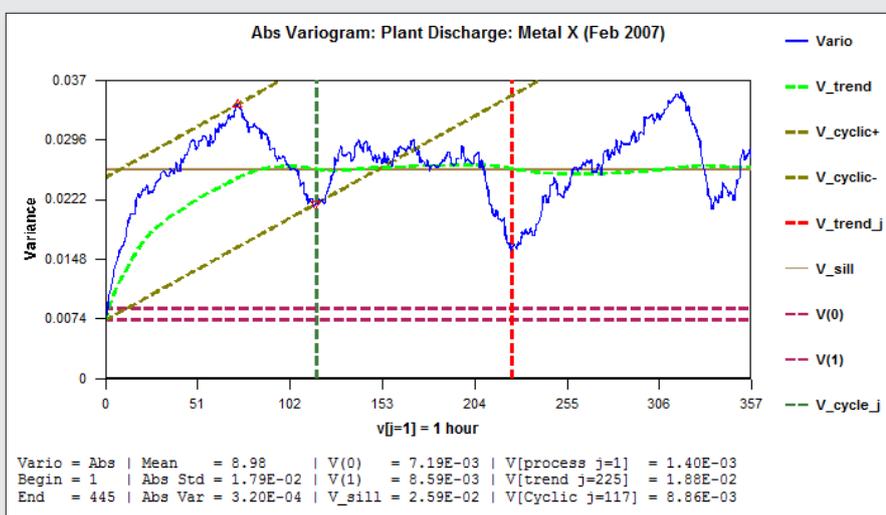

Figure 7. Absolute variogram for metal% X showing three distinct cycles.

The information derived from the variogram is transferred to the control chart and a variogram-based hierarchy of control limits is defined in Table 4, which shows how the various control Limits are defined and calculated.

The data listed in Table 4 can now be plotted on the control chart for metal X as a guide to what the sample variability is likely to do within the given constraints shown in Figure 8.

Analysing the sources of variability

Dr W. Edwards Deming clearly demonstrated in several experiments that one should not react to variations within $\pm 3\sqrt{V[0]}$. Indeed, there is no logical reason why anyone should react to some variability that does not exist in the process (i.e., *UCL* and *LCL*).

This concept allows us to optimise the sampling/measurement protocol. An active on-going reaction philosophy should be established to correct for all true process movements, when they are large enough to show through the random noise. For these movements, we must be ready to apply corrective methods.[22]

The variations characterised by $V2[j=1]$ are true process movements, predictable over a certain time frame, and potentially correctable. Therefore, we do not want to let variations become larger than $\pm\{(3\sqrt{V[0]}) + \sqrt{V2[j=1]}\}$ without reacting (i.e., *UCL′* and *LCL′*).

However, adding $\sqrt{V2[j=1]}$ to the limits *UCL* and *LCL*, to obtain *UCL′* and *LCL′*, makes a calculated allowance for process movements that are outside our control. Indeed, we are unable to react at time intervals shorter then $j=1$. This concept allows us to optimise the sampling/measurement interval.

Adding $\sqrt{V3}$ to the limits *UCL′* and *LCL′*, to obtain *UCL″* and *LCL″*, makes a calculated allowance for a process cycle beyond our control. It is very difficult to effectively correct a cycle. If this addition becomes too large, it is necessary to investigate the causes of the cycle. Sometimes, engineering modifications of the process are necessary to eliminate a cycle. Sometimes the way we organise our work results in a cycle. Understanding the cyclic nature of variability provides an opportunity to minimise a visible or invisible cost.

Table 3. The symbol, source, and amount of variability in the absolute variogram.

	Mean	**8.98% X**
	Absolute relative standard deviation	**0.0179% X**
	Absolute relative variance	**0.00032% X**
$V[0]$	**Short-range random variability** $V[0]$: this short-range random or irrelevant variability is shown by the lowest horizontal, dash/dotted line and corresponds to the total sampling, sub-sampling and measurement variability; it is not related to the plant process. It is a combination of inherent heterogeneity (random variability), fundamental error (*FE*) and grouping and segregation errors (*GSE*) and all uncontrolled sampling errors arising from a poor sampling procedure, and is a function of the total sampling and measurement variability. In the rest of this text $V[0]$ is simply referred to as the random variability or "sampling variability". If $V[0]$ is large, other errors are also likely to be large. The nugget effect accounts for a certain proportion of the total sill which is a relative measure of the overall variation. This component of error could be reduced if appropriate attention is paid to the sampling protocol. If the $V[0]$ to sill ratio is high relative to the overall variation, anything above 65%, it suggests that the sampling errors associated with short-range random variability such as the grouping and segregation error (*GE*) and the fundamental error (*FE*) of the sampling protocol, are a significant problem. Upper and lower control limits (*UCL* and *LCL*) are applied to the control charts by adding three times the standard deviation of $V[0]$ to the mean or the target average, in order to cover the 99.7% confidence interval.	$V[0] = 0.00719$ $\sqrt{V[0]} = 0.0848$ $S = \pm0.0848$ $UCL = \text{Mean} + 3S = 9.23\% \text{ X}$ $LCL = \text{Mean} - 3S = 8.73\% \text{ X}$
$V[1]$	**Total process variability** $V[1]$: is the typical value of $V[1]$ at the first lag point in the variogram. It is the total non-random variation that occurs in the plant between any two consecutive analyses. This is a combination of the total sampling and measuring variability of the process, as well as the daily sampling interval (or process) capability. This variability cannot be controlled unless the routine sampling interval is reduced. The non-random component of variability is due to bias in the sampling process related to the delimitation error (*DE*), the extraction error (*EE*), the preparation error (*PE*) or the analytical error (*AE*) and can be eliminated through implementing an optimised sampling protocol. When $V[1]$ is placed around the target mean (*TA*) it provides an indication of the upper and lower sampling capacity.	$V[1] = 0.00859$ $\sqrt{V[1]} = 0.0927$ $UCL' = 8.98 + 3S + \sqrt{V[1]} = 9.32\% \text{ X}$ $LCL' = 8.98 - 3S - \sqrt{V[1]} = 8.64\% \text{ X}$
$V[\text{process}]$	**Process variance** $V[\text{process}]$: is the continuous, non-random variability in the plant between any two consecutive analyses; it is simply the difference between $V[1]$ and $V[0]$ at the first lag point. It is variability due to the sampling process related to the delimitation error (*DE*), the extraction error (*EE*), the preparation error (*PE*) or the analytical error (*AE*) and can be eliminated through establishing a sound sampling protocol. Process variability cannot be controlled unless the sampling interval is reduced. If it is relatively high in relation to $V[0]$, almost three times, and it is likely that process variations are responsible for this. UCL' and LCL' include both the random variability due to the random, and the non-random process variability that takes place between two consecutive samples. These upper and lower control limits combine three standard deviations of the random variability $V[0]$ and the process variability $V[\text{process}]$, to give a 99.7% level of confidence in the control charts. The upper and lower control limits (*UCL'* and *LCL'*) are set in the control chart by multiplying the standard deviation of $V[0]$ by 3 in order to cover the 99.7% confidence interval, and then adding the contribution from $\sqrt{V[\text{process}]}$. The position of the sampling, measuring and process variability relative to the product stream is shown in the control chart as *UCL'* and *LCL'*.	$V[\text{process}] = V[1] - V[0]$ $= 0.00859 - 0.00719$ $= 0.0014$ $\sqrt{V[\text{process}]} = 0.0374$

V[cyclic]	**Cyclical variability** V[cyclic]: the value for V[cyclic] is half the total amplitude of the process cycle, between the highest and lowest points on the variogram, usually associated with the first cycle. It is a non-random variable related to specific activity in the process. This variability is introduced as a direct consequence of interventions on, or interactions with, the process stream. It may be related to mechanical or human interventions, but is usually due to periodic changes in the diurnal performance of or maintenance interference with equipment, or due to changes in manpower behaviour or material inputs on the plant. The regularity of both the period and the amplitude of the short-range cycles in the variogram suggest that this effect is introduced by mechanical equipment. The reason for the cycles should be identified and adjustments made to the sampling equipment.	V[cyclic] $= 0.00886$ $\sqrt{V}$[cyclic] $= 0.0941$ $UCL'' = +3S = 9.41\% \ X$ $LCL'' = -3S = 8.55\% \ X$
V[sill]	**Average variability of the process measured across the total data set** V[sill]: V[sill] is measured across the entire variogram and is a measure of the total variance in the data set, whereas V[trend] is the value of the sill at a given lag. This should be the same as V[trend] since they measure the same thing.	V[sill] $= V$[trend] $= 0.0188\% \ X$ at $j = 225$ lags
V[trend]	**First order integral of the variogram at any given lag** V[trend]: this indicates what the average variability of any individual sample might be at any given lag. Is the difference between V[0] and the first integral of the variogram at any lag and is specifically for the purpose of extrapolating the variogram back to the y-axis to obtain a value for V[0]. V[trend] can also be measured at any lag distance, but usually at the specific lag point where the variogram reaches a maximum. The trend, being the first order integral of the variogram, rises sharply in the early lags of the variogram and flattens out as it approaches the range of influence. Beyond the range of influence, V[trend] is more or less constant. It is in fact the geostatistical dispersion variance for a sample of a given support. The trend provides an explanation of the variogram behaviour during the period of the large-scale cycle. This component of variability is due to some mechanical or human intervention that takes place approximately every two to three days and introduces variability into the system. Generally, the trend of the variogram is upwards until a point (the range of influence) is reached, beyond which the variogram is level or declines.	V[trend] $= 0.0188\% \ X$ at $j = 225$ lags

Components of variability

The relative variogram allows the components of variability to be standardised and compared against one another as shown in the pie diagram of Figure 9.

The main contributor to the variability is the overall trend (or sill) that accounts for 51% of the variability. This is the height of the sill at the end of the third cycle shown in Figure 7. The reason for the high contribution of V[trend] may be a possible strong autocorrelation between one sample and the next because of the mixing of liquids.

The next major component of variability is due to the cyclical nature of the process, which accounts for about 24%. The source of this variability lies in the incorrect delimitation error (*IDE*), the incorrect extraction error (*IEE*), the incorrect preparation error (*IPE*) or the total analytical error (*TAE*), and should be identified and eliminated.

The variability due to the differences between one sample and the next (V[1]−V[0]) only accounts for about 4% of the variability. This is a relatively small value

and it tends to suggest that there is not much in the way of correlation between one sample and the next.

The value of 20% for V[0] suggests that there may be issues in regard to the sampling protocol that could be improved to
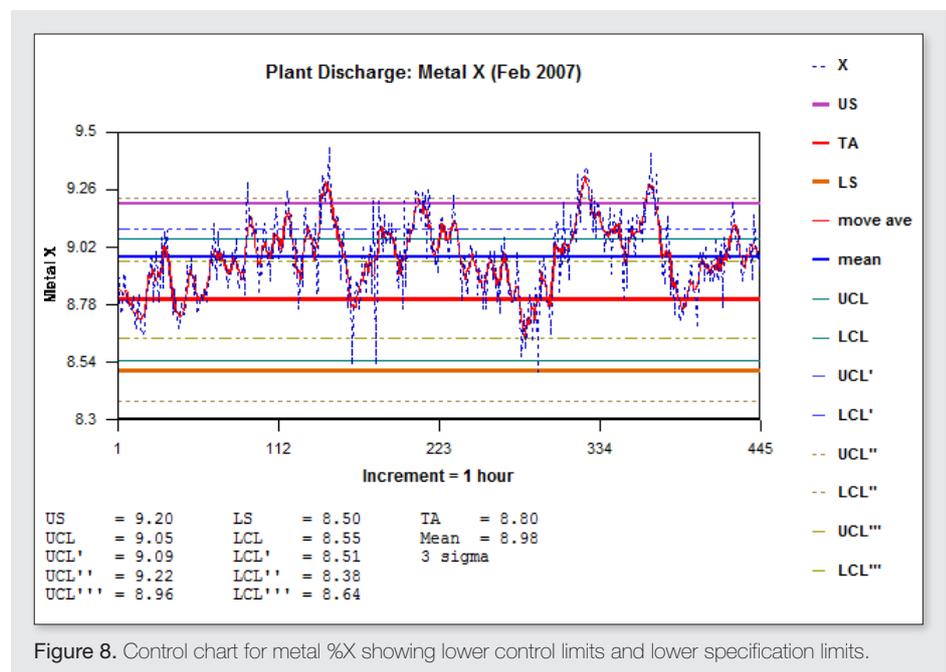

Figure 8. Control chart for metal %X showing lower control limits and lower specification limits.

Table 4. Hierarchy of control limits derived from the variogram.

Component of variability	Variances	Control limits
$V[0]$, the combined *NE*, *FE* and *GE*: *UCL* and *LCL* = Mean $\pm 3\sqrt{V[0]}$	$V[0] = 0.00719$ $S = \sqrt{V[0]} = 0.0848$ $3S = 0.2544\%\ X$	$UCL = 8.98 + 3S = 9.23\%\ X$ $LCL = 8.98 - 3S = 8.73\%\ X$
+Plus process allowance: *UCL′* and *LCL′* $= UCL \pm \sqrt{V[1]}$ and $LCL \pm \sqrt{V[1]}$ $= \pm\{(3\sqrt{V[0]}) + \sqrt{V[1]}\}$	$3 \times \sqrt{V[0]} = 0.2544\%\ X$ $+$ $\sqrt{V[1]} = 0.0927$ $= 0.3471$	$UCL′ = +3S + V[1] = 8.98 + 0.3471 = 9.33\%\ X$ $LCL′ = -(3S + V[1]) = 8.98 - 0.3471 = 8.63\%\ X$
+Plus cyclicity allowance: *UCL″* and $LCL″ = \pm\{(3\sqrt{V[0]}) + \sqrt{V[1]} + \sqrt{V[cyclic]}\}$	$3 \times \sqrt{V[0]} = 0.2544$ $+$ $\sqrt{V[1]} = 0.0927$ $+$ $\sqrt{V[cyclic]} = 0.0941$ $= 0.4412$	$UCL″ = 3 \times \sqrt{V[0]} + \sqrt{V[1]} + \sqrt{V[cyclic]} = 8.98 + 0.441 = 9.42\%\ X$ $LCL″ = 3 \times \sqrt{V[0]} + \sqrt{V[1]} + \sqrt{V[cyclic]} = 8.98 - 0.441 = 8.54\%\ X$

reduce $V[0]$. The main contributors to this sampling error are the nugget effect (NE), the fundamental sampling error (*FSE*), and the grouping and segregation errors (*GSE*), each of which could be investigated to identify the main contributor to this error.

Conclusions from the control chart

A control chart that compares the variability of metal% X against the sources of sampling error and control limits is shown in Figure 8. The main conclusion from the chart is that the lower specification limit is set appropriately. The sampling interval is more than adequate and management may even think of reducing the sampling frequency. Even though the sampling procedures and protocol may require some more attention, the plant operators should be able to control the process well. It is essential that the dynamic characteristics of the process be acknowledged, and that the reason for the cyclical behaviour be identified and eliminated. Under present conditions, it is unlikely that large invisible financial losses are occurring.

The legacy of Pierre Gy

While history turned out so that Pierre Gy only attended the first of what was to become the biannual WCSB Conferences (but happily in the company of his wife Sylvia, see *TOS Forum* Issue 5), these meetings of sampling practitioners, engineers, scientists and enthusiasts have acted to consolidate the research and application of sampling theory and practice at a global level and have presented a regular stream of high quality research since 2003. The global sampling fraternity owes much to those who initiated, contributed and have continued to organise and support these outstanding conferences.

Pierre Gy passed away on 5 November 2015, but his legacy lives on with force. His tireless work, depth of insight and completeness of scope of sampling issues from 1949 to 2015, a period of 66 years, established a formidable foundation that ensures that sampling of materials will remain a fertile field of research—and of a superior informed practise—for ever. The Theory of Sampling (TOS) provides the only complete scientific basis for sampling, including a basis for relevant standards and guidelines, and will prove itself to all who will take the time to investigate the solutions that this logical, modular approach to sampling problems offers. The ideas and approach taken by Gy in adopting Matheron's variogram concept in its application to problems associated with process control is a special feature of TOS and led to the field of study referred to as chronostatistics and, as an extension, to the logical, coherent theory of bed blending. A seminal paper in which Matheron himself assessed
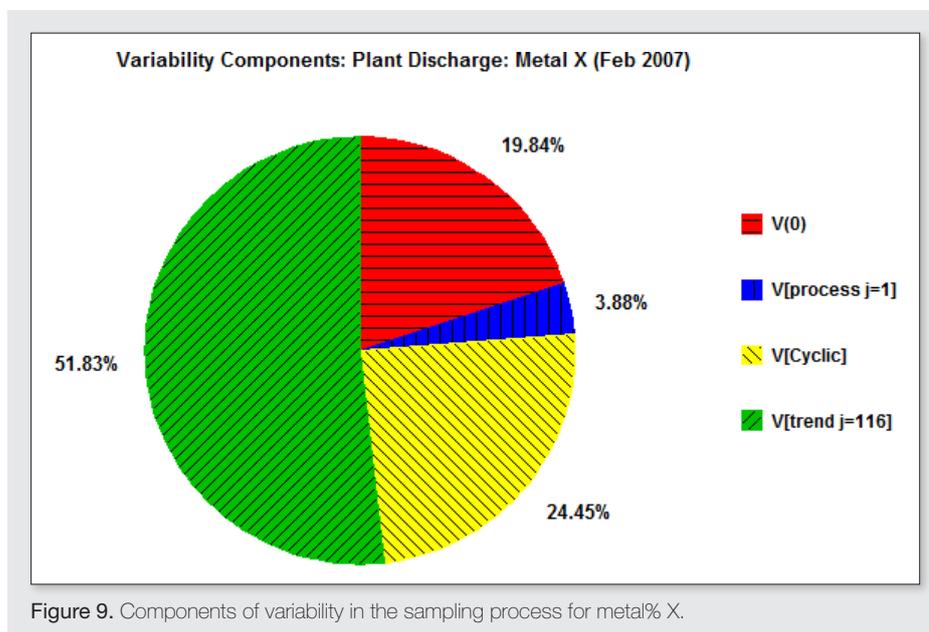

Figure 9. Components of variability in the sampling process for metal% X.

the validity of Gy's application of the vari-ogram approach to linear sampling was a highlight of WCSB7, Bordeaux, 2015 and can be found in the Proceedings of the meeting.

References

1. P. Gy, Pierre, "Part IV: 50 years of sampling theory—a personal history", *Chemometr. Intell. Lab. Syst.* **74,** 49–60 (2004). doi: https://doi.org/10.1016/j.chemolab.2004.05.014

2. D.W. Brunton, "The theory and practice of ore-sampling", *Trans. AIME* **XXV,** 826–844 (Atlanta Meeting of October 1895) (1896).

3. R. Richards, *Ore Dressing*, Volume 2. McGraw-Hill (the sampling table is on page 850) (1908). https://www.brainshark.com/malvern/vu?pi=192277776&text=M021507

4. D. Francois-Bongarçon, *Extensions to the Demonstration of Gy's Formula* (January 1998).

5. P. Gy, "L'Echantillonnage des Minerais en Vrac" (Sampling of Particulate Materials), *Revue de Industrie Minérale* (St Etienne, France) Numéro Special (Special issue) (15 Janvier 1967).

6. G. Matheron, *Les Variables regionalisees et leur estimation* (Regionalized Variables and their Estimation). PhD thesis, Masson, Paris (1965).

7. M. David, *Handbook of Applied Advanced Geostatistical Ore Reserve Estimation*. Elsevier, Amsterdam (1988).

8. P. Gy, "Sampling of discrete materials-a new introduction to the theory of sampling. I. Qualitative approach", *Chemometr. Intell. Lab. Syst.* **74,** 7–24 (2004). doi: https://doi.org/10.1016/s0169-7439(04)00167-4

9. DS 3077, *DS 3077. Representative Sampling—Horizontal Standard*". Danish Standards (2013), www.ds.dk

10. P. Gy, *Heterogeneité, Echantillonnage, Homogeneisation* (Heterogeneity, Sampling, Homogenising). Masson, Paris (1988). Complete textbook in French

11. P. Wavrer, "An automatic linear proportional sampler based on the principles of the Theory of Sampling", *TOS Forum* **Issue 6,** 25–27 (2016). doi: https://doi.org/10.1255/tosf.85

12. R. Wohlbier, "Stacking, Blending, Reclaiming", in *Bed Blending*, Ed by A.W. Gerstel. Trans-Tech Publications, Clausthal, Germany, Ch. 24 (1977).

13. P. Gy, *Theorie de la Prehomogeneisation derivee de la Theorie de l'Echantillonnage* (Theory of Bed-Blending derived from the Theory of Sampling). Industrie Minerale-Les Techniques (1981).

14. P. Gy, "A new theory of bed-blending derived from the Theory of Sampling—development and full-scale experimental check", *Int. J. Miner. Process.* **8,** 201–238 (1981). doi: https://doi.org/10.1016/0301-7516(81)90013-2

15. P. Gy, *Optimisation de la Strategie d'Exploitation d'un ensemble Mine-Metallurgie ou Carriere(s)-Cimenterie* (Optimising the Operational Strategy of a Mine-Metallurgy or Quarry-Cement Works Complex). Mines et Carrieres-Les Techniques, pp. 118–124 (1998).

16. P. Gy, "Optimizing the operational strategy of a mine metallurgy or quarry-cement works complex", *Explor. Min. Geol.* **7(1&2),** 175–180 (1998).

17. F.F. Pitard, *Pierre Gy's Sampling Theory and Sampling Practice: Heterogeneity, Sampling Correctness and Statistical Process Control*, 2nd Edn. CRC Press (1993). ISBN: 0-8493-8917-8

18. P. Gy, *Sampling for Analytical Purposes*, Translated by A. Royle. John Wiley, Chichester (1998); 2nd Edn (1999).

19. K.H. Esbensen and P. Minkkinen (Eds), WCSB1 Proceedings, Special Issue of *Chemometr. Intell. Lab. Syst.* **74(1),** (2004).

20. P. Gy, "Sampling of discrete materials II. Quantitative approach—sampling of zero-dimensional objects", *Chemometr. Intell. Lab. Syst.* **74,** 25–38 (2004).

21. P. Gy, "Sampling of discrete materials III. Quantitative approach—sampling of one-dimensional objects", *Chemometr. Intell. Lab. Syst.* **74,** 39–47 (2004).

22. F.F. Pitard, *The Theory and Practice of Sampling*. A short course presented through the Geostatistical Society of South Africa (GSSA) in the School of Mining Engineering, University of the Witwatersrand (2006). See also Reference 17.