Available online at www.sciencedirect.com

SCIENCE DIRECT®

Chemometrics and Intelligent Laboratory Systems 74 (2004) 49–60

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemolab

# Part IV: 50 years of sampling theory—a personal history

Pierre Gy

*Res. de Luynes, 14 Avenue Jean de Noailles, F-06400 Cannes, France*

Received 22 August 2003; received in revised form 6 April 2004; accepted 28 May 2004
Available online 15 September 2004

**Abstract**

This last part, which is a rather personal history of the development of the theory of sampling, is written in the first person singular—for a reason. For a long time already, I have been asked to tell how I became interested in sampling and how I developed the theory. I don't like to speak of myself and I have hitherto refused to do so. I have always been reluctant to accept such an undertaking, at least as long as I thought that my work was not completed. It would now appear that this is no longer the case and so, when the editor requested also the present Part IV as part of the series, he originally invited me to write for the *SSC6* proceedings (see Introduction to this issue), I finally ran out of excuses and obliged (in point of fact, it took much more than a mere "request"). Upon reflection, I am very grateful for offering me this opportunity.

The development of the theory of sampling has been a solitary work from the very beginning. With the exception of the "variogram", a mathematical tool borrowed from geostatistics and Matheron [23] in 1962, I did not use any pre-existing scientific work. On the other hand, no one or no body such as university, school of mines, research organization or industry, even my own employers, *ever* asked or encouraged me to search in this direction and nobody ever paid for my research work (with an exception concerning the theory of "bed-blending", which was sponsored in 1978 by a blending equipment manufacturer—exception duly mentioned in my publications). Unusual.


*Keywords:* Theory of sampling; Bed-blending; Binomial model

## 1. How the seed was planted

From 1946 to 1949, I worked as a Mineral Processing Engineer in a small lead mine, north of the lower course of the Congo River, in the middle of the equatorial bush that covers what was at that time the French Congo (or Congo-Brazzaville). I was in charge of the processing plant and of the laboratories. In 1947, I received a 1-week-old "cable" from Paris asking me to provide the head-office with an estimate of the average grade of a huge heap of lead concentrate of dubious quality, stored in the open since 1940, to study the possibility of its re-treatment. I discussed the question with the mine manager and I soon realized that:

- □ I was asked "to sample" a batch of some 200,000 tons that contained blocks weighing anything between several tons down to microgram particles.
- □ I knew nothing about "sampling".
- □ The available literature (very scarce at the equator), was mute, naive or vague at best.
- □ I had to *improvise*, which I did as best I could—which was not very good.

The seed was planted, but I did not realize it was the starting point of a lifetime's work (Fig. 1).

## 2. State of sampling theory, anno 1949

Back home in France in 1949, as I was in charge of a mineral-processing laboratory in Paris, our team worked on a huge variety of ores and minerals from all over the world. I soon found out that I had to solve sampling problems practically *every day*. The literature available in Paris, though more comprehensive than the one I had access to in the Congo, did nevertheless still not provide me with any satisfactory answers. All authors on sampling (few and far between) had dedicated their work and energy to answering the question "how much", i.e. "what is the minimum sample weight necessary to achieve a certain degree of reliability".

0169-7439/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2004.05.014

Fig. 1. 1947 (aged 23). First job, near M'Fouati lead mine, Middle Congo (200 km from Brazzaville), in the by then French Equatorial Africa.

For instance, more or less arbitrary formulas were proposed according to which the minimum sample weight had to be proportional to the *cube* of the top particle size (Brunton, 1865) or, which makes a serious difference with coarse materials, to its *square* (Richards, 1908).

Brunton's formula was based on very reasonable considerations of geometrical similarity: the idea was that, at different sizes, the same number of fragments was required. That of Richards was based on (quote) "...the fact that the quantities proposed by Brunton's formula were much larger than those accepted in practice"... "the most satisfactory rule must be based on habits acknowledged by the trade of minerals" (in this year 2000, i.e. nearly a century after Richards, the same philosophy is implemented by ISO Technical Committees.)... "by adopting the rule that the sample weight should be proportional to the square (sic) of the top particle size, one should obtain figures that have every chance of being approved by sampling operators" (end of quote). For a famous M.I.T. Professor, this can hardly be called a *scientific* or a *theoretical* approach.

As far as Brunton's formula is concerned, I was worried by the fact that the *constant proportionality factor* did not allow for the other physical or mineralogical properties of the ores and minerals involved, especially their variations in grade and density.

In the 1930s, the trade of coal comprised very large tonnages, as well as huge amounts of money, which were computed on the basis of *assays* (ash, sulphur, etc.) carried out on "samples". Various teams of researchers, mostly in the UK and the USA, realizing that sampling actually generated errors that could have a financial impact, had launched *experimental studies* with the purpose of disclosing relationships between the properties of coal (especially percentage of ash, top particle size, etc.) on the one hand, the sample mass and the sampling variance on the other. Thousands over thousands of data were compiled but no clear conclusion could be drawn and no result could be extrapolated to other minerals, which supports Albert Einstein's statement: "a theory can be checked experimentally but there is no way to derive a theory from experiments".

In the1940s, the Mineral Processing Engineer's Bible was the "Taggart" (first edition 1927; second revised edition 1945, *John Wiley, New York*). In the latter, I found a chapter on sampling, written by Prof. Hassialis, Columbia University, New York, that included a theoretical section based on a statistical multinomial model. This model was sound but involved a very large number of parameters that were never known. For obvious reasons, it could not be *practically* implemented. Fifty-five years later, I have never met anyone who did implement it.

In 1949, the French Mining Engineer R. Duval, searching the handbooks of statistics for a ready-made solution, proposed to approximate a batch of ore with a population of black and white balls (*binomial model*) representing pure valuable mineral and pure gangue, respectively. The model attributed *the same statistical weight* to the "balls", which implied that they had *the same physical mass*. This implicit assumption was so far from reality, where fragment masses could vary in a ratio of 1 to $10^{18}$ and where the minerals were seldom "liberated" from one another, that it was practically worthless. Its results were dangerously misleading. This triggered a reaction from me: for want of any available solution adapted to the problem, I decided to study the question from a purely *theoretical* standpoint... and the seed, planted in 1947, began to germinate in 1949.

## 3. The 1950 theoretical approach

Sampling is always necessary for a single, simple reason: in most cases and for a question of cost, analysis can of course not be carried out on the *entire bulk* of the object, the "lot $L$", to be valued. The practical purpose of sampling is therefore to reduce the mass $M_L$ of lot $L$ to the mass $M_S$ of the "sample $S$" that will represent $L$ in further operations and ultimately in analysis; the analytical result pertaining to the entire lot $L$ is to be estimated on sample $S$ without altering the composition "too much". This mass reduction must be realized by *selecting* a certain number $N_S$ of "constituents" or "elements" (fragments in the case of particulate solids) from the population of $N_L$ elements making up the lot $L$. The theory deals with a single sampling stage: the reduction of a certain lot $L$ to a certain sample $S$. The analysis carried out on the ultimate sample, or assay-portion, $S$ concerns a certain component of interest, $A$, which is called the "critical component". The objective of the interrelated sequence "sampling+analysis" is to *estimate* its proportion in sample $S$. The proportion of $A$ in $L$ and $S$ is called the "critical grade", denoted by $a_L$ and $a_S$, respectively. The objective of my initial research was to study the statistical distribution of $a_S$ and that of the "total sampling error (TSE)" (more details have been presented in Parts I, II and III above).

Specifically, the idea was to develop a mathematical model with the purpose of devising a relationship between:

1. The variance of the sampling error (a random variable),
2. The physical properties of the material being sampled, assumed to be known, and...
3. The lot and sample masses.

From this relationship, the minimum sample mass to be extracted from the lot in order to achieve a given degree of reproducibility (characterized by a given *sampling variance*) could be derived. I was not yet interested in the distribution mean, i.e. in the *sampling bias*: at that time, nobody was. Sampling was universally regarded as a simple *handling technique*, the tools of which were an assortment of shovels, scoops, spoons and containers: the *theoretical question* "how" had never been posed by anyone. I did feel that this point was very important however, but I was not able to deal with it until 20 years later: *the reader should know that I had to carry out my research work in my spare time only, for I was not paid by my employer to carry out this kind of research*.

The theoretical model I first devised was derived for particulate solids of mineral origin such as ores, concentrates and much later for feed to cement factories, etc. irrespective of their nominal particle size. Later again, I was also able to formulate the following further developments:

- ☐ In a first generalization step, the theory was also applicable to *solids of vegetable origin*, such as, e.g., cereals or sugar beets, as well as *solids of animal origin* such as bones imported from India and Pakistan by the gelatine industry—and indeed *any* particulate solid.
- ☐ In a second generalization step, it was applicable also to *liquids and gases*, such as those to be controlled in the chemical, pharmaceutical, oil or hydrometallurgical industries.
- ☐ More generally, with the development of environmental control, the theory was found valid also for sampling of the *rejects* of all kinds of human activities: household or industrial refuse, polluted soils, nuclear materials, etc.
- ☐ Matter is discrete, or discontinuous, by essence: with particulate solids the discontinuity appears at the scale of fragments (sizes expressed in centimeters, millimeters or micrometers). With liquids and gases, it is observed at the scale of molecules or ions (sizes expressed in Angstroms). The difference between particulate solids and liquids is thus not one of essence but rather one of scale—as far as sampling is concerned of course. The general sampling model is valid irrespective of the component size(s); it would therefore appear applicable to all material "objects", irrespective of their physical state (Fig. 2).

Fig. 2. OECC Mission (Europe), 1953. Meeting on: "The beneficiation of low-grade ores". As one of France's two delegates (aged 29, second from right), I am lazily listening to some lecture (the memory is not quite up to the photographic documentation; I have forgotten where the meeting actually took place...).

In the abstraction of the mathematical model of sampling, this theory seems therefore to have some form of universal validity. This point is attested by Richard Bilonick [25].

## 4. The Formula

The 1950 sampling model assumed that the number $N_L$ of elements (fragments) making up the lot, *a number usually very large, unknown but defined unambiguously*, was reduced, in one way or another, to a (much) smaller number $N_S$ of fragments making up the sample $S$. My approach was to compute the mean and the variance of a population made of the grades $a_S$ of *all* possible samples of $N_S$ fragments, i.e. all combinations of $N_L$ objects by groups of $N_S$ units.

To remain as close to reality as possible, I had decided:

- ☐ In a first step, to take into account *all* parameters (unknown but well defined physically) characterizing *all* fragments $F_i$: i.e. the grades $a_i$ and masses $M_i$, as well as the numbers $N_L$ and $N_S$ of fragments making up lot and sample respectively, and to devise strict, indisputable, mathematical relationships, based on simple algebra.
- ☐ In a second step, to introduce *simplifications* and *approximations* in order.
- ☐ In a third step, to devise *practical formulas*, approximate but easy to implement.

At the end of the first step, I had devised strict formulas for the mean and variance of the population of "equally probable samples of $N_S$ fragments". With today's notations:

$$\sigma^2(TSE) = \left[\frac{1}{N_S} - \frac{1}{N_L}\right] \frac{\sum_{(i=1 \text{ to } N_L)} h_i^2}{N_L} \quad \text{with}$$
$$h_i = \frac{(a_i - a_L)}{a_L} \times \frac{M_i}{M_i^*} \tag{1}$$

According to its definition, the TSE is relative. The variance of any relative error is *dimensionless*. In this expression, $h_i$ is what we today call "the contribution of $F_i$ to the constitutional heterogeneity of $L$" and $M_i$* is the average mass of all $F_i$. The role of $h_i$ appears of great importance: it is the link between the concept of quantified heterogeneity I introduced, and later developed, and the sampling variance. Indeed, heterogeneity lies at the root of all sampling errors: the sampling of a strictly homogeneous material would be an exact operation. The theory of homogeneity and heterogeneity was presented in its definitive form since 1975 [15–20].

This basic fomula (1) involves a sum extended to the $N_L$ values of $a_i$ and $M_i$ that are well defined but remain always unknown. It is strict but *cannot* be directly implemented in practice. Today, with the computing facilities at our disposal, it would for example be possible to *simulate* all kinds of distributions of $a_i$ and $M_i$ and to compute $\sigma^2$(TSE) according to Eq. (1). Theory shows that $\sigma^2$(TSE) as expressed in Eq. (1) is a strict *minimum* (see below: Section 6). This is why I termed the corresponding error the "fundamental sampling error (FSE)" (formerly FE).

At the end of the second and third steps, I had indeed obtained an *approximate formula* for the sampling variance, which is often referred to (by others) as "Gy's formula" (here referred to more simply as "The Formula").

It can be expressed as follows (=~ "approximately but practically equal to"):

$$\sigma^2(FSE) = \sim \left[\frac{1}{M_S} - \frac{1}{M_L}\right] c\beta f g d^3$$
$$= \sim \frac{c\beta f g d^3}{M_S} \quad (\text{when } M_S \ll M_L) \tag{2}$$

- ☐ $M_L$ and $M_S$ represent the masses of $L$ and $S$, respectively (expressed in grams).
- ☐ $c$ is a "constitutional parameter". It takes into account the average grade of the material as well as the densities of all components. It has *the physical dimension* (*but not the meaning*) of specific gravity (always expressed in g/cm$^3$). This parameter can vary very widely. The smaller the average content, the larger the parameter $c$. For instance, with an alluvial gold ore containing 1 g of gold per metric ton of ore (1 g/t=1 ppm=$10^{-6}$), its value is *1 million times* the density of gold (19 g/cm$^3$). With the feed to a cement factory, it is only a *fraction* of the density of limestone (ca. 2.7 g/cm$^3$).
- ☐ $\beta$ (or $\lambda$ or $l$ for certain authors) is a dimensionless "liberation parameter", which varies between 0 and 1 according as the components are thoroughly associated—or completely liberated—from one another. The estimation of the liberation parameter is often tricky, especially with gold ores. Francois-Bongarcon proposes the expression $\beta=(d_{\text{lib}}/d)^{1.5}$ (see Literature survey in Part V of this series).
- ☐ $f$ is a dimensionless "particle shape parameter", also varying between 0 and 1 that, with most materials, is practically equal to 0.5 (flakes and needles are exceptions).
- ☐ $g$ is a dimensionless "size range parameter" again varying between 0 and 1. It has a general value of 0.25 with *uncalibrated* mineral populations, tending toward 0.75 with naturally *calibrated* materials such as cereals. It would equal 1.00 with, for example, high-quality bearing balls of strictly identical diameters.
- ☐ $d$—the "top particle size" (expressed in centimeters for dimensional homogeneity) is defined as the size of the aperture of the square-mesh screen that would retain exactly 5% of the material (passing 95%). The determination of $d$ must be very precise *as it is raised to the third power*. A quick visual estimation is not always as precise as necessary.

Thanks to simplifications and approximations which are not supposed to alter the order of magnitude of the variance, I had succeeded in transforming a sum extended to a *multitude of unknown terms* into a product of factors, which can, in most cases, be estimated with a good degree of precision. This "Formula" has seen an unexpected, but pleasing very wide use.

Formulas (1) and (2) were developed in 1950 (now more than 50 years ago), then proposed in an internal, unpublished note (Refs. [1,2], Part V). Contrary to what I recommend in the foreword to Part I of this series, I had in fact answered the second question "how much?" *before* knowing the answer to the first and foremost question "how? " As already mentioned, this fundamental question had never been clearly posed by anyone at that time and I did not answer it before the beginning of the 1970s.

## 5. Conditions of validity of The Formula [2]

The formula (2), which expresses the fundamental variance $\sigma^2$(FSE), is still valid today. However, in the books I have published since 1979, a much more elegant and general demonstration has been given, which the reader finds in Part II. It is now based on the "probabilistic sampling model", whereby each element $U_m$ of $L$ is submitted to the selection process with a certain selection probability, $P_m$. This generates the TSE. In the most general case of the probabilistic sampling model, TSE is the sum of two terms: the *correct sampling error (CSE)* and the additional *incorrect sampling error (ISE)*:

$$TSE = CSE + ISE \tag{3}$$

If the sampling is correct, i.e. if $P_m = P =$ constant, then ISE=0 which entails:

$$TSE = CSE \tag{4}$$

If the sampling is correct, and if the elements are selected *individually and independently*, TSE and CSE boil down to the FSE:

$$TSE = CSE = FSE \tag{5}$$

It entirely falls upon the user to implement a correct sampling but except in tests or computer simulations we cannot select the elements making up the sample *individually and independently*. In practice, therefore, condition (2) is never fulfilled. The best we can do is to extract multi-elemental increments, *I*—i.e. *groups of neighboring elements*—with a uniform selection probability *P*. In this case, the existence of a distributional correlation between selected elements generates a new additional error, namely the *grouping and segregation error (GSE)*, which entails:

$$TSE = CSE = FSE + GSE \tag{6}$$

When using The Formula, the reader should never forget that it is valid *only* when *both* conditions are fulfilled. It is meaningless, and dangerous, to answer the question "how much?" by means of the formula which governs the sampling *variance* only—without first answering the question "how?", which governs the much more influential sampling *bias*.

## 6. Experimental check of theoretical results—first publications

In order to check the validity of my approach, I had, in 1950–1951, organized an experiment (described in Section 23.3 of Ref. [18]), which consisted of splitting a lot *L* (a few kilograms of lead ore) into 16 samples obtained at the end of 4 stages of riffle splitting (divisions into twin fractions). These 16 samples were weighed, carefully pulverized and assayed for Pb. The variance of the population of 16 results (i.e. the variance of the global estimation error, GEE) was computed; the variance of the total analytical error (TAE) was subtracted and thus a first experimental estimate of the variance of the TSE was obtained. It was several times larger than the variance computed according to The Formula, but it was of the same order of magnitude at least. This suggested that the error taken into account by the model was the *minimum of the total sampling error* TSE. For this reason, I decided to call it "fundamental sampling error" (formerly FE).

The error FSE was only *one component* of the CSE and another error, itself the sum of several components, resulted from the fact that actual sampling did not respect the second condition assumed by the model, which generated the GSE. In addition, the ISE made their appearance. It was only much later that these errors were logically analyzed and their components identified. This experiment also showed, as an unexpected by-product, that a perfectly symmetrical riffle splitter could introduce a *sampling bias* when the sampling operator did not follow a certain number of rules, unheard of at the time, which were later formulated as an answer to the question "how?".

The internal *standard* [2] and the results of the splitting experiment were first presented publicly at the occasion of the *Second International Mineral Processing Congress*, held in Paris in 1953 and *published* in the proceedings in 1954 [3].

My next step was to devise a certain number of charts making it easier to implement the 1950 formula. These were presented at the occasion of the *Third International Mineral Processing Congress*, at Goslar, Germany in 1955 and published in Erzmetall [4] and R.I.M. [5]. In the same spirit, I designed a "sampling nomogram", a circular cardboard calculator, produced by *Minerais et Metaux* in 1956 (French [6F], English [6E] and German [6G] versions). This nomogram was presented in Japan (Bull. of the Tohoku University) in 1960 [7], the title in Japanese (Kana) can be found in the literature survey of [18]. The sampling nomogram was followed by a "sampling slide rule" operating on the same principle (French [8F] and English [8E] versions) in 1965. For technical and cost reasons, this slide rule could not be manufactured earlier.[1]

Back in 1956, however, I showed that the 1950 formula could easily be transposed to the case where component *A* was a given *size fraction* and the critical content $a_L$ therefore corresponded to the proportion of this size fraction in lot *L* [9]. I obtained another formula that was first presented in a French magazine in 1956, and then under the title "The Sampling Error Committed on Size Distribution" at a Mining Congress in Jamshedpur, India in 1957; this was published in the Indian Mining Journal the same year [10]. I had then already shown that The Formula was also applicable to the moisture content of a lot of wet material.

It was in 1957 that I first presented the formula in English at an annual meeting of the Society of Mining Engineers of the American Institute of Mining Engineers (SME of AIME) in New Orleans, LA [11]. It was not presented in the UK—at an annual meeting of the Institution of Mining and Metallurgy (IMM) in London—until 1965 [12].

[1] Information to *young* readers (who studied *after* ca. 1970), who known only about electronic calculators and modern computers, students and engineers of the 1940s and 1950s calculated by means of *slide rules* based on the properties of *logarithms*, transforming an equality $AB=CD$ into a sum of the type:

$\log A + \log B = \log C + \log D$.

Nomograms and sampling slide rules worked on this principle.

## 7. Sampling of flowing streams

In 1960–1962, I concentrated my attention on the problem posed by incremental sampling of flowing streams, a problem of paramount importance in mineral processing and ship-loading facilities for example. One can imagine three ways of reducing the mass of a flowing stream of particulate solids, liquids or multiphase media (e.g. pulps of finely ground minerals in water):

- ☐ Taking *the whole* of the stream during *a fraction* of the time (by means of cross-stream samplers taking increments at usually uniform intervals),
- ☐ Taking *a fraction of the stream during the whole* of the time,
- ☐ Taking *a fraction* of the stream during *a fraction* of the time.

In the mineral industries, the first method is often implemented (with exceptions) whereas in the chemical, pharmaceutical, oil, food industries, etc. the second and third methods, cheaper in the short term, are always preferred. I was conscious of the existence of two problems yet *unsolved*:

1. Cross-stream samplers should respect certain rules regarding, for instance, the cutter velocity and the cutter opening, shape and width. These rules remained to be defined scientifically, which was achieved only in 1977, after a campaign of experiments carried out on bauxite blocks [22]. But at that time I was interested in the mathematical problem posed by the second point (next paragraph).
2. The increments extracted from the stream, usually at a constant interval, are *not independent from one another*. There is a correlation between the composition of a slice of matter and the instant it passes through the sampling area. In this case the statistical laws designed for "populations of independent units" are no longer valid. It remained to develop the statistics of auto-correlated time series.

I had already collected many series of experimental data and was studying them when I heard of *Georges Matheron's* work and his recent creation of a new science called *geostatistics* [24], presented in English by *Michel David* [25]. From a theoretical standpoint, there is no difference in essence between the *spatial* correlation along drill cores, for example, and the *time* correlation along a flowing stream. It was on this occasion that I *borrowed* from Matheron the "variogram" as a function characterizing the autocorrelation of flowing streams. This opened new fields of research (later called "chronostatistics") that I explored during the years 1961–1965. I presented my first publication on this subject under the title "Variography" at another annual meeting of SME of AIME in Denver, CO, in 1962 [24] and at the Institution of Mining and Metallurgy (IMM) in London in 1965 [12].

## 8. First book in French—synthesis of the quantitative approach

In 1962, I felt the need (perhaps the *urge*—experienced authors will understand what I mean) to write a book gathering my experience of sampling, both as a theoretician and as a troubleshooter. I was then employed by *Minerais et Metaux* in Paris. I worked in excellent harmony with the CEO, my friend *Roger Testut*, but my time was more and more dedicated to management problems, less and less to scientific matters—the development of a sampling theory had no priority in the objectives of *Minerais et Metaux*. This left me in reality with no time to write such a book. I had to make a choice: I could not be both a manager and a scholar. Were I to stay in the first, very comfortable, position I had to abandon sampling theory. This soon came to a crossroad. I therefore opted for the second option... and for a *random income*.

I became a freelance *sampling consultant*, probably the first of this kind in the World, and I moved out of Paris with my family to the city of Cannes on 1st January 1963. I was now free to write my book and I started right away. Since then, soon 40 years ago, in spite of some difficult years of tightrope walking, I have never regretted this choice.

For the years to come, my time was shared between numerous forms of activity always overlapping each other in time and space: theoretical research, consulting, troubleshooting, lecturing, teaching regular courses in various schools and universities, teaching privately organized short-courses and, last but not least, writing magazine articles and books. My activities, limited to France at the beginning, led me all over the world as soon as my articles and lectures in English helped the mining and metallurgical industries realize the importance of scientific sampling. I now had the opportunity to work on practically all kinds of mineral materials, from coal or cement raw materials to diamond, gold or platinum ores by way of uranium.

So far, I had dealt only with the quantitative approach to the sampling problems where I thought I had proposed adequate solutions for both zero- and one-dimensional objects. I endeavored to gather all the results already obtained in the first of a two-volume book to be published by "Societe de l'Industrie Minerale" (SIM). This project met with more objections than I had anticipated, from one member of SIM scientific committee. The publication of the book, ready in 1965, was delayed until 1967 [13]. It was published in its original version thanks to *Lucien Vieilledent's* and *George Matheron's* friendly help. Their support was decisive in my struggle to have this book published.

This first volume was followed, in 1971, by its second part [14], in which I developed solutions to specific problems such as studies of spatial distributions, the

Fig. 3. 7th International Mineral Processing Congress, Praha, 1970 (aged 46, center). The secretary general of the congress was a friend who has asked me to help him solve a delicate problem—in the middle of the cold war: The Czech ice hockey team had just defeated the Russian team (4–3) at the end of a murderous match. The following evening, the downtown Aeroflot branch office, next door to my hotel, was destroyed. Now, you will understand my friend's problem: At the last minute, five Russian professors demanded to present papers which had not been selected by the appropriate committee—all of them in the opening session "Crushing and Grinding". My friend needed a chairman to take this responsibility away from him. When he asked me—I accepted (I like sports). Reluctantly, an American and an Italian professors accepted to preside together with me (witness their none-too-enthusiastic faces above). I gave each would-be speaker exactly 2 min after which I switched the microphone off. Anyway, nobody understood anything, because every sentence was first translated into Czech and secondly into the four official languages of the Congress (English, French, German, Russian). From what my friends in the room later told me, I was introduced as something like (as related by the French translation): "The President . . . of . . . France" (De Gaulle was still alive!)—and the rest was double Dutch to everyone.

sampling of coal and precious metal ores, sampling for size analysis or for moisture estimation, study of sampling errors resulting from the practical implementations of the model, etc. A large part of this book deals very *practically* with the question "how?", but in a non-structured way. I did not yet introduce the concepts of *probabilistic* and *correct* sampling. In the meantime I had gathered an important number of references and this book contains a 769-reference literature survey (Fig. 3).

## 9. A new theoretical and practical synthesis—first definition of the concept of correct sampling

In 1972 [23], I tried for the first time ever to propose a qualitative approach to the sampling theory and to answer publicly the question "how?", neglected so far. I presented the concepts of probabilistic, *non-probabilistic, correct and incorrect sampling* to the annual meeting of SIM.

- ☐ A sampling was then said to be *probabilistic* when it was based on the notion of *selection probability*. In 1979 [16], this definition was refined and its new formulation is still valid today: a sampling is said to be probabilistic when *all* fragments have a non-zero probability of being selected.
- ☐ A sampling was then, and still is, said to be *non-probabilistic* when this condition is not fulfilled—for instance when it results from "picking", or from a deliberate choice, by the sampling operator, of the fragments making up the sample.

  Today's reader may be sceptical but the "hammer and shovel method", which is based on such a deliberate choice, was then described by most standards, including ISO. In 1988, i.e. sixteen years after [23] and thirteen years after the book [15], ISO proposed a text (ISO/DIS 6153) still describing the (non-probabilistic) "hammer and shovel method" for the sampling of chromium ores.

- ☐ A probabilistic sampling is then, and still is, said to be *correct*:
  - ■ With zero-dimensional objects: when *all* fragments have a *uniform probability* to be selected.
  - ■ With one-dimensional objects: when the density of selection probability is *uniform throughout* the one-dimensional domain occupied by lot $L$.
- ☐ A probabilistic sampling was said to be *incorrect* when the pertinent condition is not fulfilled.

The idea that sampling could be treated as a science was new and shocked some distinguished members of the audience. One of them favored a definition whereby, if sampling was at all to be thought of in terms of probability, the selection probability of each fragment had to be *proportional to its mass*. Readers may easily judge for themselves the pertinence of such a definition. *Arthur Koestler* is right when he says (in "The Sleepwalkers"): "As with contagious diseases, new ideas need long incubation periods before their effects are observed". According to my own experience, I would say between one and two generations.

As soon as the 1971 book was published, I felt the need to write a new book. I had acquired a quarter of a century of experience as a theoretician, consultant and troubleshooter and this book was to be full of practical experience. For personal reasons, I decided to be my own publisher. The writing, typing (by a professional typist), printing and binding of the new book took about 4 years and the first copy of the book was handed over to *Roger Testut*, to whom it was dedicated, in 1975 [15]. No more than a few hundreds copies of this book were ever sold.

For the first time in a book, I was able to distinguish between the a priori selecting conditions—*on which we can act to some extent*—and the a posteriori properties of the sampling error, which *result* from the selecting conditions and which we can but observe, usually too late. This amounted to distinguishing between the possibilities of the sampling equipment and the qualities the users of this equipment could expect or demand from manufacturers of this equipment. In short, I had to build up a logical and mathematical bridge between the selecting conditions and the sampling errors. For the first time —I had in the meantime overcome the wry opposition—the concepts of *probabilistic, non-probabilistic, correct, incorrect* sampling were presented in a book:

- ☐ *Properties of the selection process*: it can be probabilistic or non-probabilistic; if probabilistic it can be correct or incorrect.
- ☐ *Properties of the sampling errors*: these are *random variables* that can be characterized by their statistical distribution and the properties of their moments: a sampling can be *accurate or biased* (properties of the mean), *reproducible or not* (properties of the variance), *representative or not* (properties of the mean-square).

Some of the definitions used today (part I) are *slightly* different from those of 1975 but the overall philosophy of this approach was set then and has not changed since.

## 10. First book in English—first presentation of the double Student's *t*–Fisher test

During congresses, or on the occasions of lectures in English-speaking countries, I had been asked to write a book in English, but nobody had volunteered to translate my latest book. On the other hand, since 1974, I had been working in cooperation with *Elsevier Publishers*, who had asked me to create the "International Journal of Mineral Processing" and to become its Editor-in-Chief. They asked me to write an updated version of my 1975 book [15] in English. I accepted what was a challenge, without realizing the kind of work expected of me: *Elsevier* had asked me to provide *a camera-ready copy* of the text. This entailed that the presentation of each page had to be definitive when it left my office. Digital word processing techniques were as yet totally unavailable and my only choice, *excellent* at that time, was the well-known IBM "golf ball" typewriter, which had already been used for the typing of [15].[2]

I vainly tried to hire in Cannes the services of a professional typist capable of typing an *English* text full of *mathematical* expressions, of *Greek* letters and other symbols, Alas, I had to type it all myself. Due to the difficulty of correcting typing mistakes, I first had to write the entire text *by hand* and then to have it corrected for the language. It is one thing to present a 20-min lecture in a Congress where nobody remembers what you said, far less your language mistakes anyway, and quite another to write several hundred pages in a foreign, not completely familiar language.

I decided to make a test. I would write, as best I could, what was to be an introductory chapter and submit it to an American newsman living in Cannes, whom I had met and who was willing to help me. A few mistakes considered minor by the American reader were corrected and I went on with his benediction. Since then, I have had some doubts as to the reliability of his advice. *Elsevier* had the text reviewed again and I retyped entire pages or paragraphs. The first edition of this imperfect book was available in 1979 [16]. It received a rather satisfactory review and the very decent reviewers were kind enough not to insist on the language deficiencies.

I still vividly remember the winter 1978–1979 when I worked over 10 h a day, 6.5 days a week to type the 431-page manuscript, doing nothing else. At the average rate of 3 pages/day, I spent 4 months on the typing. It very seldom snows in Cannes, but it was one of these rare snowy winters and, on Sunday afternoon, I would walk around the nearby mountains, in knee-deep snow, for a wonderful change.

The book was, for a large part, a translation from existing texts in French. Its most original feature probably was a statistical chapter presenting a *double* Student–Fisher test eliminating the risk of drawing a wrong conclusion using the alternative *single-sided* test. Many people, including authors of bias tests recommended by ISO Standards did (and still do) persistently mistake the "absence of proof of bias" (rightful conclusion of a Student–Fisher test) for the "proof of absence of bias" (wrong, *biased* conclusion of the same test). ISO standard 3086 is entitled: "Iron Ores—Experimental Methods for Checking the Bias of Sampling". As far as I know, its latest version was published in 1986 (7 years after the publication of Ref. [16]) and it still makes the same mistake. Most standards on sampling of iron ores were revised in 1998, but this bias test 3086 was not (19 years after the publication of Ref. [16]), and is still the one on record.

Chemometricians also should be very careful with the Student–Fisher test as it is presented by these and other standards.

As early as 1981, the first edition was nearly out of print and the publishers asked me to prepare a second revised edition, which was available in 1982. The major revision concerned the statistical chapter 31, which was refined and became definitive.

## 11. Complete textbook in French—new developments derived from sampling theory

My latest book in French [15] was 7 years old when the second edition of Ref. [16] was released. In the meantime, I had developed several ideas leading to new applications of the theory, namely:

- * Point by point computation of the auxiliary functions of the variogram,

[2] I still treasure this typewriter together with the collection of six golf balls I used at that time.

* Theory and practice of proportional sampling,
* Theory of bed-blending derived from the theory of sampling and industrial implementation.

On the other hand, from hundreds of missions carried out in more than 80 countries in a time span of 40 years, I had gathered a respectable amount of practical experience illustrating my theoretical conclusions and I thought it useful to publish them. The French Publisher, *Masson, Paris*, was willing to publish such a book. The latter was ready in 1988 [17].

It took me 6 years to achieve this because I had a number of new subjects to incorporate into former texts and I also had to work as a consultant and short-course teacher, my only sources of income (authors' royalties paid by publishers are ridiculously low, but having a book published by a well-known, respected publisher usually generates a certain amount of consulting work.). My troubleshooting activities had started with base metals such as lead, zinc and copper, but with the developments of the uranium industry in France and abroad, the latter had become one of my major sources of work in the 1980s. Ever since the 1950s, I had worked in close cooperation with the French "Commissariat a l'Energie Atomique" (CEA) and, when they were created, with its mining and metallurgical subsidiaries COGEMA and COMURHEX in France, Gabon, Niger, Canada and South Africa. These companies became my major clients, as kindly recalled by my friend *Robert Bodu* in "Les Secrets des Cuves d'Attaque" [27], the history of uranium ore processing in France. I had to alternate writing, teaching and consulting. I led a busy life...

During these years, in addition to illustrations or refinements of the existing theory, I had also developed two new subjects, which similarly needed to be presented in a textbook. I believe that they have an enormous industrial potential:

1. Mass and volume measurement by proportional sampling,
2. Theory, and industrial implementation, of bed-blending.

## 12. Proportional sampling

In 1954, I was confronted with my first problem of "metallurgical balance reconciliation" in a group of North African lead and zinc flotation plants. A metallurgical balance is nothing other than the application of the *Lavoisier's* stochiometric principle at the scale of an entire mineral processing plant—it can be summarized easily enough "whatever comes in must ultimately come out, one way or another". When this is not observed there per force *must* be either *measurement biases* or *unsuspected losses*—and with a single exception in 45 years of consulting, what came out was always *less* than what came in. The mine owner had observed that there was a persistent 2–3% deficit of lead and zinc produced and he suspected shortcomings of his sampling systems (he had just read my first publication [3]). After a visit to the plants and 1 year of remote monitoring work in cooperation with the staff and a check of all measurement devices involved, I reached the conclusion that sampling was only a minor source of bias and that the biggest bias was to be attributed to the *conveyor belt scales*, in spite of the fact that they were calibrated once a day every day. In fact, the bias was due to this calibration.

Over the years, I discovered that all kinds of scales could be found operating on conveyor belts (all types of mechanical scales, nuclear scales), and they all suffered from a structural lack of reliability. It is one thing to carry out an easy electrical measurement and quite another to convert it accurately into a tonnage of ore. This opinion was reinforced when I read *Hendrik Colijn's* "Weighing and Proportioning of bulk Solids" [28]. The following is a quote of chapter 7, confirmed by *Colijn* when we later met.

The actual plant performance of belt scales, unfortunately, does not always measure up to the claims of the manufacturer or to the expectations of the operator. Instead of the ½ percent accuracy, some plant personnel have claimed that 10 percent is a more realistic figure and on a large number of installations, they may be correct.

This is true also of nuclear scales, ibid. (chapter 9).

When developing the theory of sampling, I had reached the mathematical, indisputable conclusion that, when sampling was carried out correctly with uniform selection probability $P$, the sample mass $M_S$ was a random variable with a mean equal to $P$ times the lot mass $M_L$.

$$m(M_S) = PM_L$$

In addition to this property, when the number of increments in the sample is "large enough" (which is nearly always the case) the confidence interval of $M_S$ is very small. The sample $S$ can be weighed, $M_S$, by means of conventional *static* scales (*very reliable*), with the consequence that for correct sampling, and when the uniform selection probability $P$ can be estimated accurately, the quantity $M_S/P$ is an excellent, unbiased estimator of the lot mass $M_L$.

$$M_S/P = \text{ unbiased estimator of } M_L$$

According to my experience, this unbiased estimator is much more reliable than *any* that, e.g., can be obtained by means of the belt or nuclear scales available from existing specialized manufacturers. This is the basis of "proportional sampling" (PropSamp). In 1980, I recommended this new technique to the South African *Rustenburg Platinum Mines* and, since then, it has been used routinely to calibrate the nuclear scales that had been installed originally and provided unreliable results, which I have been able to check on the occasions of later visits to South Africa.

A variant of proportional sampling was implemented at the mineralurgical pilot plant of the *Bureau de Recherches Geologiques et Minieres* (*BRGM*) of Orleans-La Source, France. It consists in implementing a proportionality ratio (the selection probability $P$) that may not be known with great precision but that, *by construction, is strictly the same for all streams* (feed, concentrates, tailings). The sets of samples and lots masses are then "homothetic", which makes it very easy to compute the metallurgical balance.

In practice, the rules of sampling correctness are always applied with a safety factor so as to make sure that the mass estimation will be as highly reliable as desired, with the consequence that this proportional sample is also "perfectly" *representative* of the lot. After appropriate drying and weighing, the sample is reduced and assayed for the critical component(s) in the conventional manner. The same sample thus provides all qualitative and quantitative information needed to compute the metallurgical balance of an entire plant.

*Proportional sampling* is so simple that some people would not believe its efficiency. Nowadays, simple techniques, especially when they do not use sophisticated, preferentially computer-controlled equipment, do not inspire confidence and I was often required to *prove* the adequacy of PropSamp. I was challenged to check its reliability in an existing pilot plant against a weighing system involving a 10-$m^3$ tank mounted on strain gauges, a centrifugal pump, a correct sampler and a water meter. The results of this experiment have been described in my books since 1988, e.g. chapter 29 of Ref. [18] and chapter 13 of Ref. [20]. Interestingly, instead of supposedly proving the unreliability of PropSamp, this experiment in fact helped disclose fundamental inadequacies of both the strain gauges and of the water meter, with which the pilot plant was equipped, and which had hitherto been considered to work "to everybody's satisfaction".

## 13. Bed-blending, derived from sampling theory

Since the 1960s, Lafarge Cements had realized the necessity of an accurate sampling of the feed to their cement kilns. Together with the sampling equipment manufacturer MINEMET-INDUSTRIE (a reincarnation of my former employer Minerais et Metaux), we designed and installed highly reliable sampling plants in their Cement Works.

Cement kilns, like metallurgical furnaces, are known for their severe lack of flexibility. They require to be fed with material as uniform as possible—the ultimate, very costly penalty is the loss of a kiln. To achieve this purpose, in a first step, Lafarge plants feed their raw materials to what is known as a "bed-blending system", which ensures an imperfect form of one-dimensional homogeneity. The feed to this system is sampled in a MINEMET sampling plant coupled with an X-ray analyser capable of assaying a sample for its major components in a few minutes. The whole system, assisted by a computer which calculates the average composition of the pile being formed, works in such a way that at the end of the constitution of a blending pile, its average composition is very well known and, when properly managed, is practically equal to the ideal feed to the kiln.

One of Lafarge subsidiaries had installed a bed-blending system manufactured by PHB-SOMERAL (now MBH) of Mulhouse, France. The blending was adequate but the technical manager of PHB had observed that the blending system did not work in agreement with Gerstel's theory, published in 1977 [29] and generally accepted. He asked my advice in 1978 and I offered to develop a theory of bed blending, which, I realized, could easily be derived from the existing sampling theory. This new theory of *bed-blending* was developed right away and presented to PHB-SOMERAL in June 1978. This was a wonderful but, unfortunately unique experience: an equipment manufacturer wanted—and was ready to pay the services of a consultant—to understand how *his own* equipment really worked.

In order to convince potential clients, I was further asked to carry out a full-scale check of the theory, which was realized about 6 months later at the Heming Cement Works (Lorraine). Lafarge carried out the X-ray assays. To everybody's satisfaction, these experiments showed that the new theory was in perfect agreement with experience, contrary to Gerstel's.

This theory and the experimental check were published for the first time in 1981 [30,31] and can now be found in the books [17–20]. The Canadian mineral industries were interested and invited me to present the philosophy of blending the feed to a plant at the occasion of the 100th anniversary of the Canadian Institute of Mining and Metallurgy (CIM) in Montreal (1998) [32,33].

## 14. Complete textbook in English—presentation of proportional sampling and bed-blending

As soon as my latest book [17] in French was released, *Elsevier* asked me to write its English version. This was ready in 1992 [18] and is practically a translation of Ref. [17]. It contains nothing original worth mentioning.

## 15. Summarized versions in French and English

The voluminous 700-page textbooks [17,18] contained complete, updated mathematical demonstrations and I had to write them as reference books, but they were simply too heavy and too costly to reach a wide audience. There was a need for much shortened (on the order of 150-page) versions. My French publisher *Masson* was ready to publish it, which was achieved in 1996 [19]. My British friend *Allen*

*Royle* kindly offered to translate the French text into English, which I accepted with gratitude. For the first time, my work was available in *excellent* English. *John Wiley* offered to publish this book, which was released in 1998 [20]. It was well accepted by the public and a second edition was marketed in 1999. I am very grateful to "Le bon Royle".

With this, my account has reached 50 years, indeed between one and two generations.

## 16. What does the future hold?

First of all—always—there is the family. I am not a man to talk of my family in a public context such as this. The editor of this account, however, is very persistent. Thus, one *sample* picture from the family Gy will have to suffice (Fig. 4).

## 17. The theory of sampling at the year 2000: 50 years—and beyond

I was originally writing this series with the intention that it was to be part of the proceedings of the *SSC6* conference in 2000 (see Editors' introduction)—in fact, I took pains to be able to finish the text on Christmas Eve 1999, which I considered an appropriate goal: Fifty years to the mark! Editorial events outside my influence later made it expedient to augment this series with several other related papers and to publish this interesting lot (*L*) *altogether* as a Special Issue. It was to be rather severely delayed however.[3] The present amicable journal serves the same scientific community to which I have never before catered, so the delays incurred are hopefully forgiven when the result is now finally at hand.

Fig. 4. 1996, Pougnadoresse (Gard). 50th wedding anniversary with wife Sylvia, daughter Caroline and grandson Stanislas. The whole family was together, altogether 22 persons—and there are now six more great-grandchildren since then. I am not worried about the future of the Gy family. But none of them bears the name Gy, which goes back—at least—to the Norwegian Vikings in the Normandie.

## 18. Is this the end of the story?

I don't know, but the intensity of this work must soon be reduced. It may be the beginning of the end for me but certainly not for the Theory of Sampling and its applications. I am very glad to have reached beyond the mythical "Year 2000"—Now I have begun to hand the relay to a new generation of professors and engineers, students and industrialists—to all *proper samplers*. . .

## Acknowledgements

I would like to mention my Scandinavian friends who have played an important role in disseminating my work in the later years. *Pentti Minkkinen*, Professor at the Lappeenranta University of Technology, Finland, was, to the best of my knowledge, the first to teach courses of the theory of sampling on a regular basis and to develop a computer program (SAMPEX) to estimate the fundamental sampling variance by means of "The Formula" as early as 1986. Thanks to Pentti, I later made the acquaintance of *Kim H. Esbensen*, then Chemometrics Professor at the Telemark Institute of Technology (HIT) in Porsgrunn, Norway. He was the enthusiastic Chairman of the *Sixth Scandinavian Symposium on Chemometrics*(*SSC6*), in which capacity he invited me to present a synopsis of the Theory of Sampling at this occasion, August 15–19, 1999, and from which was developed the present tutorial series. Thanks to Kim, I made the acquaintance of the late (April 2002) Professor *Sunil de Silva*, Head of the Powder Science and Technology Research Group (POSTEC), also of HIT, who in addition invited me to present my work to the audience of the Third Symposium on the Reliable Flow of Particulate Solids (Relpowflow III), of which he was Chairman, also held at HIT, August 11–13, 1999. I thus had the opportunity to reach out to two completely new scientific communities in 1999, to my intense satisfaction.

[3] The author was originally writing this piece as part of a larger tutorial series with the intention to be part of the proceedings of the SSC6 conference in 1999 (due for publishing in late 2000), but it was decided to opt for a whole independent Special Issue on sampling. Sadly, only the proceedings of SSC6 made it into print, while the planned tutorial issue met with surprisingly severe, indeed hostile reactions. Various manoeuvres by highly placed non-to-TOS-interested chemometricians intervened, and the tutorial series was never published in the planned journal. It took another chemometric journal and the foresight the editors Massart and Minkkinen when accepting the proposal for the present Special Issue to eliminate this opposition [Editor's comment].

There has been a steadily growing sampling activity in Scandinavia since, and equally pleasing. I am not worried about the future of sampling in Northern Europe.

In France, the *International Sampling Institute*(*ISI*) was created by a group of French Consultants in 1999 to perpetuate the national tradition of interest in sampling. Back in 1979, *Selected Annotated Titles* wrote in their review of my first book in English [16]:

The French have made a speciality of statistical applications in the earth sciences and this contribution only serves to underline their dominance.

ISI is active in organizing short courses in France (and abroad), in French and in English. Here again, the future of the sampling theory is in good hands. It is ably led by Denis Thirouin.

The newest offspring of organized sampling activity concerns the *International Sampling and Blending Forum* (*ISBF*), the embryo of which was founded by two close colleagues and friends *Kim H. Esbensen* and *Dominique Francois-Bongarcon* at an early 2001 spring day encounter in Copenhagen airport, Kastrup. ISBF will operate on the worldwide scale. ISBF will be lead by a *virtual board* of international directors; the first board was selected at WCSB1. ISBF will make it its objective to reach out primarily to the world university communities—including technical universities of course—on all matters of proper sampling, teaching, research, experimental work in collaboration with industry and other users of TOS.

It is also most appropriate here to acknowledge the active help in discussing, publishing, co-writing many of my later papers, which has been given to me by *Dominique Francois-Bongarcon*.

Last but not least, I would like to express my deep gratitude to my very good friend, to my excellent editor *Kim H. Esbensen* for the huge amount of work he fed into what primarily was a “run-of-the-pen” manuscript pile so as to transform it into a perfectly edited series of tutorial articles that, I hope, will interest many new professions.