THE POINT SELECTION ERROR INTRODUCED BY SAMPLING ONE-DIMENSIONAL LOTS

Ana Carolina Chieregati; Professor; Mining and Petroleum Engineering Department; University of São Paulo; Av. Prof. Mello Moraes, 2373; 05508-900; São Paulo – SP - Brazil; +55 11 3091-6033; +55 11 9155-0148; ana.chieregati@poli.usp.br.

João Felipe Coimbra Leite Costa; Professor; Mining Engineering Department; Federal University of Rio Grande do Sul; Av. Bento Gonçalves, 9500; Bloco IV, Prédio 75; 91501-970; Porto Alegre – RS - Brazil; +55 51 3308-9480; +55 51 9987-0704; jfelipe@ufrgs.br.

Marcelo Monteiro da Rocha; Professor; Geosciences Institute; University of São Paulo; Rua do Lago, 562; 05508-080; São Paulo – SP - Brazil; +55 11 3091-4204; +55 11 8397-9970; mmrocha@usp.br.

Anastácio Barbosa Pains; System Analyst; Rio Paracatu Mineração S/A; Geology Department; Estrada do Machado, s/n°; 38600-000; Paracatu – MG - Brazil; +55 38 3679-1063; anastacio.pains@rpm.kinross.com.br.

ABSTRACT

A sample is said to be representative when it is taken by a selection method that is both accurate and reproducible. Thus, representativeness is characterized by the absence of bias and an acceptable variance. The sole aim of sampling is to reduce the mass of a lot without significantly changing its other properties. Reducing the mass of a lot of broken ore can be done by: (1) either taking increments, used on flowing streams of material; or (2) splitting, used when the whole of the lot can be handled. Incremental sampling is used on flowing streams of discrete particles (solid fragments, liquids) that can be characterized by one-dimensional time or space models. The calculation of sampling variances of increments taken from moving streams is a problem that must be properly addressed, even though certain authors and standard committees resort to excessively simplistic solutions. One-dimensional heterogeneity of chronologically ordered sets or time series can be characterized with the aid of a very powerful tool, the variogram. This study was performed at a gold mine in Brazil and makes use of variogram auxiliary functions to estimate and compare, for four sampling campaigns, the point selection error produced when sampling one-dimensional lots.

INTRODUCTION

The reliability in sampling results depends on several aspects, *i.e.* the mineralization characteristics, sampling quality, sample analysis and manipulation. This reliability can be evaluated by sample grades variation (precision) and by results accuracy (bias). The variance of the overall estimation error is related to sampling (80%), to sample preparation (15%) and to chemical analysis (5%). Therefore, when planning a sampling program, one must aim at the elimination of all errors or, at least, the elimination of those possible to eliminate, in such a way that it is possible to gain precision and accuracy with a minimum cost (Grigorieff, 2002).

Basically sampling errors have two main causes. The first one is related to the systematic error, which leads to often increased or often decreased mean values of sets of results. The other cause can be related to the variability or heterogeneity of the studied material and is mainly attached to fundamental error, to point selection error and to grouping and segregation error. Apart from such errors one must also consider errors associated to chemical analysis; even being apart from the total sampling error, they are included in the overall estimation error.

When samples or sample increments are collected from a conveyor belt at regular time intervals, a one-dimensional error is introduced named the 'continuous selection error' or 'integration error', recently changed to 'point selection error' (PSE). This error depends on ore heterogeneity and on the variability of particles inside a lot, and can be defined as the difference between the real grade of the lot and the grade of the sample collected along the time interval considered (Grigorieff et al., 2005). After collecting increments from the flow, these increments are seen as zero-dimensional and are affected by the fundamental error (FE) and by the grouping and segregation error (GSE). All of these errors are associated to the heterogeneity between particles, and their variances can be added together, producing the total variance of the sampling error.

This study took place in a gold mine in Brazil and aims to determine and study the point selection error based on four sampling campaigns from four different mining blocks. The results clearly demonstrate the influence of the increment number and the time interval between sampling, which draws upon sampling precision.

METHODOLOGY

This study included four sampling campaigns, in four different mining blocks, which, after mining and crushing stages, were sampled from the conveyor belts that feed the processing plant. The samples collected consisted of 1m-belt material, each increment weighting approximately 50 kg.

The sampling method was the same for all four blocks: we stopped the conveyor belt and collected, using a shovel, material referred to 1 m of the belt. Stopped belt sampling is a popular method recommended by well-known national and international standards (Pitard, 1993). The difference between the campaigns was the number of increments, the time interval between samplings and the total sampling time. Table 1 specifies, for each block sampled, the total number of increments, the total sampling time and the time interval between samplings.

Table 1 - Sampling specifications for each campaign.

block	number of increments	total sampling time	minimum interval between samplings
1	6	2h	20'
2	18	30h	60'
3	50	65h	60'
4	24	14h	30'

STATISTICAL ANALYSIS AND DISCUSSION

Industrial activities are characterized by a constant need to transport materials. The practical implementation of such activities necessarily generates long piles, running materials on conveyor belts, and streams that are all defined as one-dimensional lots. These lots are generated by chronological operations; consequently, they will be affected by fluctuations that are mainly reflecting human activities at the mine. The variability between particles or groups of particles transported by these streams produces a large scale error of observation known as the point selection error (PSE). The following approach introduces the concepts and presents the methodology for the calculation of the point selection error.

The Variogram

A sampling selection is said precise when the sampling error is minimally dispersed around its average. Therefore, precision concerns measuring the variability of a sample around the average of the lot from which it has been collected. This measurement is generally expressed as the variance of the sampling error $s^2(SE)$. The following procedure intends to estimate this variance, with the aid of a very powerful tool: the variogram.

A variogram is a plot of the average differences in a characteristic (such as the values of the various h_m, or the total heterogeneity carried by an increment of a one-dimensional lot) between pairs of increments selected as a function of time or distance. When a variogram is calculated, increments regularly spaced along a one-dimensional lot are considered, and the measurement of a continuity index of their heterogeneity h_m as a function of time or distance is determined. Therefore, the variogram characterizes the sequential heterogeneity of the series.

According to Pitard (1993), in many practical cases, the variogram of the grade a_m can be assimilated to the variogram of the heterogeneity h_m, if the various increments taken into consideration have similar weights (*i.e.*, within a maximum of ±20%). The following study considered increment grades, since all weights remained around 50 kg. The first step, after determination of the increment grades a_q, was the calculation of the experimental variogram, defined as:

$$v(j) = \frac{1}{2N} \sum_q (a_{q+j} - a_q)^2 \quad (1)$$

where j is the lag or time interval between two increments, v(j) is the variogram function for the

time interval j, a_q is the grade of the increment q and a_{q+j} is the grade of the increment separated by j from the increment q. There are several pairs of values separated by j, therefore, N represents the number of pairs.
In order to better understand the use of experimental variograms, we usually express the time interval between two increments by a dimensionless number, named j. This number is calculated by dividing a given time interval θ by the minimum interval θ_{min} between two increments, and we call it 'lag', or time interval that separates two increments:

$$j = \frac{\theta}{\theta_{min}} \quad (2)$$

The main advantage of j over θ and θ_{min} is that j is dimensionless and is therefore easy to use and quick for reference purposes (Pitard, 1993). We define the lower limit j = 1 as the 'basic time interval' or 'basic unit' of the variogram.

Next step is the estimation of v(0), also called 'nugget effect', which in fact is the sum of several components:

- The variance of the fundamental error FE.
- The variance of the grouping and segregation error GSE.
- The variance of all the other components of the sampling error.
- The variance of the different sampling errors generated during the stages of sample reduction from the primary increments to the final selection of the material taken for analysis.
- The variance of the true analytical error AE.

According to Gy (1998), we are not dealing with a series of actual grades – always unknown – but with random estimates introduced by selection, preparation and of analysis. All the errors generated are to be found in the intercept v(0) of the variogram.

There are several methods of estimating v(0), but the simplest one is duplicating the increments. Provided enough material is available, it is easy to divide each increment into two by splitting. This method gives Q pairs of matching increments and Q differences between the estimated or calculated values of a_q. The variance of this population of differences is an excellent estimator of 2v(0). An other option is to extrapolate the points v(3), v(2), v(1) to zero. When the first points of the variogram show a regular behavior, no serious errors are incurred by extrapolating them to zero. Table 2 shows the values of v(0) calculated by extrapolation and by duplicating increments. When possible, both methods were used, with similar presenting results.

Table 2 - Estimates of v(0) for each block.

block	v(0) duplicates	v(0) extrapolation
1	0.0173	0.0163
2	0.0163	-
3	0.0161	-
4	0.0063	0.0071

Figure 1 shows the experimental variograms calculated for each block. The horizontal dashed line characterizes the global heterogeneity of the series and can be represented by the statistical variance s^2. This variance is the model of a variogram where the values of a_q were taken at random from a population with a mean of zero and a variance $s^2(a_q)$.

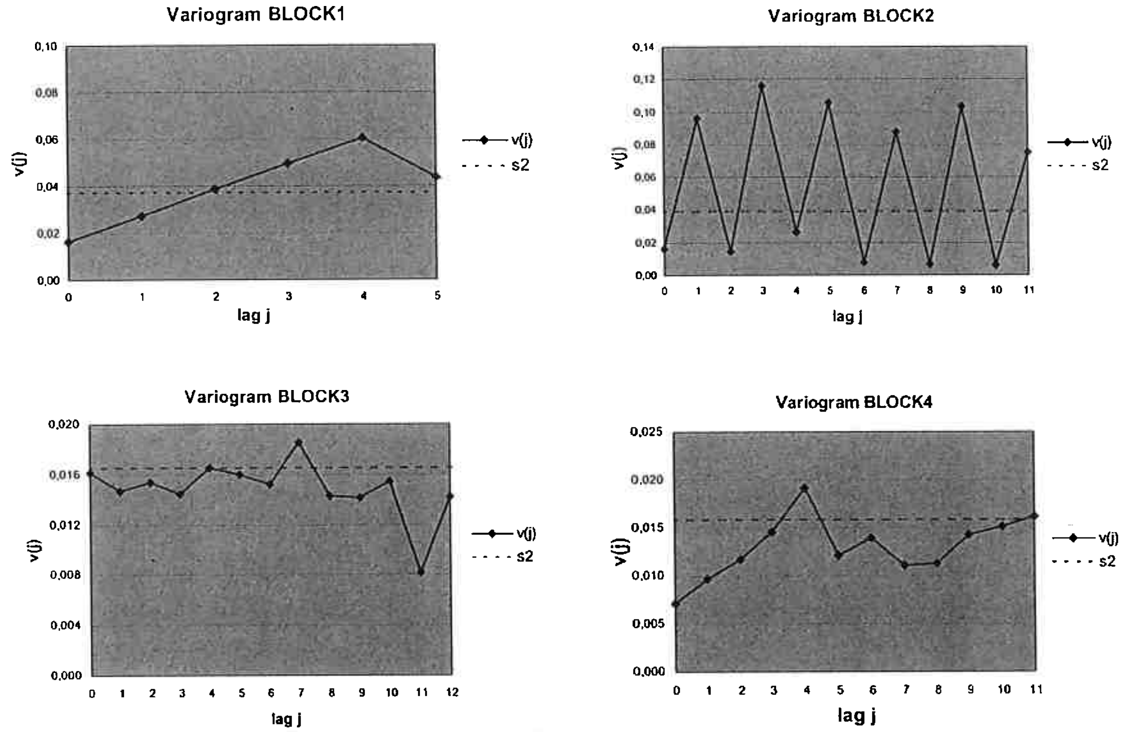

Figure 1 - Experimental variograms for each block.

As can be seen, the only variograms that presented a correlation between sample increments are numbers 1 and 4, with basic time intervals (j = 1) of 20' and 30' respectively. The other two variograms, with a 60'- basic time interval, did not show temporal correlation between samples. This effect is named the 'pure nugget effect' and it represents a typical behavior of spatially independent variables. The calculation of point selection error will be demonstrated only for block 4. The results for the others are presented at the end.

The Auxiliary and the Error Generator Functions

The variogram function can be used to develop other functions that are needed for the estimation of sampling variances. This process begins with the definition of four auxiliary functions:

- S(j): simple integral of the variogram,
- w(j): mean value of S(j),
- S'(j): double integral of the variogram,
- w'(j): mean value of S'(j).

Considering a systematic selection, with regular time intervals, the auxiliary functions can be written as:

$$S(j) = \int_0^j v(j')dj' \quad (3)$$

$$w(j) = \frac{1}{j} S(j) \quad (4)$$

$$S'(j) = \frac{1}{j}\int_0^j dj' \int_0^{j'} v(j'')dj'' \quad (5)$$

$$w'(j) = \frac{1}{j}S'(j) \quad (6)$$

The next step is to calculate the error generator function. According to Pitard (1993), an 'error generator' can be defined as a function of the variographic interval j, directly derived from the variogram and its auxiliary functions, and often reduced to a constant that can be divided by the number of increments Q making up a sample to calculate the variance of the point selection error. The error generator function, W(j), is defined as:

$$W(j) = 2w(j/2) - w'(j) \quad (7)$$

Point-by-Point Calculation of an Integral

There are two methods to calculate the auxiliary functions:

- Mathematical modeling of the variogram and algebraic calculation of its single and double integrals.
- Point-by-point estimation of the integral.

The second method, is superior and much simpler to use (Gy, 1998). This technique was developed by Gy and can considerably simplify the task of the computer software specialist. The technique on which the estimation of the auxiliary functions depends is based entirely on the point-by-point estimation of an integral, and it rests on two hypotheses that appear to be the most realistic and the simplest of any:

- The variogram passes through all the experimental points v(j) and through the estimated value v(0).
- The variogram consists of straight lines connecting its points (as shown in Figure 1).

In keeping these hypotheses, the integral of v(j) is equal to the area enclosed by the lines joining the points v(j) and the abscissa axes. Some elementary geometrical observations show that this area is strictly equal to the area enclosed by the step-wise line joining the points v(0) to v(5) and the abscissa axes, as shown in Figure 2, where T_0 represents the basic time interval of the variogram.

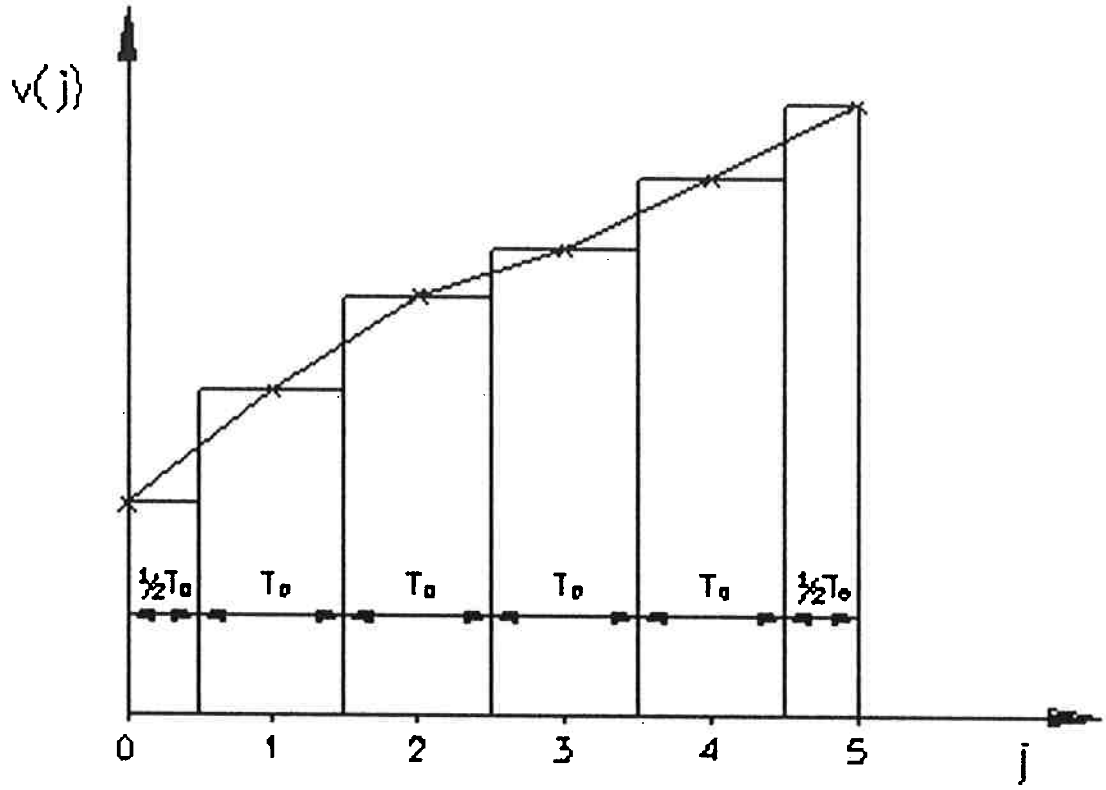

Figure 2 - Point-by-point calculation of an integral.

And so, S(j) e S'(j) can be written as:

$$S(j) = S(j-1) + \frac{1}{2}v(j-1) + \frac{1}{2}v(j) \quad (8)$$

$$S'(j) = S'(j-1) + \frac{1}{2}S(j-1) + \frac{1}{2}S(j) \quad (9)$$

Because we do not have access to numerical values of w(x) except when x is integer, when j is odd we must do an approximation; therefore, between j and j + 1 we do a linear interpolation of w(j). Under these conditions, if we define j_0 as the sampling interval of interest, we may write:

$$2w(j/2) = 2w(j_0)\text{, if j is even and } j = 2j_0 \quad (10)$$

$$2w(j/2) = \frac{2S(j_0+0{,}5)}{(j_0+0{,}5)}\text{, if j is odd and } j = 2j_0{+}1 \quad (11)$$

Finally, if j is non-integer:

$$S(j_0+0{,}5) = S(j_0) + \frac{1}{4}v(j_0) + \frac{1}{4}v(j+0{,}5)\text{, for } j = 2j_0{+}1 \quad (12)$$

Now, the variance of the point selection error can be calculated.

The Variance of the Point Selection Error

Considering Q the number of increments taken for a given interval j, the variance of the point selection error can be defined as:

$$s^2(PSE) = \frac{W(j)}{Q} \quad (13)$$

Table 3 shows, for the block 4, the variances of the point selection error for each interval j. Multiplying j by the basic time interval (in this case 30'), we have the time interval considered for the variance calculation.

Table 3 - Calculation of the point selection error's variances (block 4).

j	v(j)	S(j)	w(j)	S'(j)	w'(j)	2w(j/2)	W(j)	Q	s²(PSE)
0	0.0071	0.0000	0.0071	0.0000	0.0071	0.0142	-	-	-
0.5	0.0084	0.0039	0.0077						
1	0.0097	0.0084	0.0084	0.0042	0.0084	0.0155	0.0071	19	0.00037
1.5	0.0107	0.0135	0.0090						
2	0.0117	0.0191	0.0095	0.0179	0.0090	0.0168	0.0078	18	0.00043
2.5	0.0131	0.0253	0.0101						
3	0.0145	0.0322	0.0107	0.0436	0.0097	0.0180	0.0083	17	0.00049
3.5	0.0168	0.0400	0.0114						
4	0.0192	0.0490	0.0123	0.0842	0.0105	0.0191	0.0086	18	0.00048
4.5	0.0157	0.0577	0.0128						
5	0.0122	0.0647	0.0129	0.1411	0.0113	0.0202	0.0089	15	0.00060
5.5	0.0131	0.0710	0.0129						
6	0.0140	0.0778	0.0130	0.2123	0.0118	0.0215	0.0097	15	0.00064
6.5	0.0125	0.0844	0.0130						
7	0.0111	0.0903	0.0129	0.2963	0.0121	0.0229	0.0108	14	0.00077
7.5	0.0112	0.0958	0.0128						
8	0.0113	0.1014	0.0127	0.3922	0.0123	0.0245	0.0123	14	0.00088
8.5	0.0128	0.1074	0.0126						
9	0.0143	0.1142	0.0127	0.5000	0.0123	0.0257	0.0133	12	0.00111
9.5	0.0147	0.1214	0.0128						
10	0.0151	0.1289	0.0129	0.6215	0.0124	0.0259	0.0135	12	0.00112
10.5	0.0157	0.1365	0.0130						
11	0.0162	0.1445	0.0131	0.7582	0.0125	0.0258	0.0133	11	0.00121

Analyzing the results presented and knowing that the measurement of the sampling precision is expressed by the variance of the sampling error, we conclude that the sampling precision decreases as: the time interval between increments collected increases and the number of increments decreases. This conclusion is well illustrated by the variances for j = 2 (Q = 18), j = 3 (Q = 17) and j = 4 (Q = 18), where we can see that the value of s^2(PSE) is not only a function of j or Q independently, but a combination of both.

For a 95% confidence interval ($CI_{95\%}$) and assuming errors present a normal distribution, sampling precision is given by ±2s(PSE). In the example presented in Table 3, for j = 1, the variance of the point selection error is 0,00037 and, so, s(PSE) is 0,0193. Knowing that the average grade for block 4 was 0,363 g/t Au, and assuming all other errors (fundamental, analytical, grouping and segregation) do not exist, the sampling precision for j = 1, or 30'- time interval, is: 0,363 ± 0,0387 g/t Au or 0,363 ± 10,7%.

Following the same procedure for blocks 1, 2 and 3, Table 4 presents the average grades, the basic time interval θ_{min}, the number of increments Q for this interval, the variance of the point selection error s^2(PSE) and the 95% confidence interval $CI_{95\%}$.

Table 4 - Estimates of point selection error for each block.

block	average grade Au (g/t)	θ_{min} (minutes)	Q	s^2(PSE)	$CI_{95\%}$	$CI_{95\%}$ (rel)
1	0.673	20	6	0.002717	0.1042	15.5%
2	0.524	60	7	0.002333	0.0966	18.4%
3	0.471	60	34	0.000474	0.0435	9.2%
4	0.363	30	19	0.000374	0.0387	10.7%

As previously described, all the errors generated by selection, preparation and analysis are to be found in the intercept v(0) of the variogram. Therefore, the sum of v(0) with s^2(PSE) shall represent the total variance of the sampling errors, as shown in Table 5. The last column shows the relative standard deviation and indicates the sampling precision considering all errors involved.

Table 5 - Variances and standard deviations considering all errors.

block	average grade Au (g/t)	v(0)	s^2(PSE)	$s^2_{v(0)+(PSE)}$	$s_{v(0)+(PSE)}$	$s_{v(0)+(PSE)}$ (rel)
1	0.673	0.0173	0.0027	0.0200	0.1414	21.0%
2	0.524	0.0163	0.0023	0.0186	0.1364	26.0%
3	0.471	0.0161	0.0005	0.0166	0.1288	27.4%
4	0.363	0.0063	0.0004	0.0067	0.0819	22.5%

The tables indicate that, as the number of increments increases and the time interval decreases, the power of the variogram in estimating sampling precision becomes greater. For an inadequate sampling protocol, the variogram can't indicate sample correlation.

CONCLUSIONS

Even considering all concepts involving the sampling theory, it is practically impossible to conduct in an industrial scale what is theoretically correct. Gold sampling has an odd behavior, particularly when one considers the segregation among particles as well as the difficulty in reducing sample mass without causing major change in grades. All gold sampling problems are overturned as much as its grade is reduced and as much as the deposit becomes marginal and the most irregular is the gold distribution in the rock.

Despite the characteristics of the deposit in study, it's possible to estimate the errors involved in every sampling stage. As has been showed, when sampling one-dimensional lots, two major causes decrease the sampling precision: the increase of the time interval j between samplings and the decrease of the number of increments collected. Moreover, it could be seen that the power of the variogram in estimating sampling precision becomes greater as the number of sampling increments increases.

REFERENCES

Grigorieff, A, 2002. Desenvolvimento de um novo protocolo de amostragem de carvão mineral. Dissertação de mestrado. PPGEM, UFRGS, Porto Alegre.

Grigorieff, A, Costa, J.F.C.L, and Koppe, J, 2005. Variogram of a continuous flow: a tool for designing sampling increments. In: Second World Conference on Sampling and Blending, Queensland, Australia. pp. 27-31.

Gy, P, 1998. Sampling for analytical purposes. 1st ed., translated by A.G. Royle, John Wiley & Sons, West Sussex, England.

Pitard, F F, 1993. Pierre Gy's sampling theory and sampling practice: heterogeneity, sampling correctness, and statistical process control. 2nd ed., CRC Press, Boca Raton, Florida.