A multivariate approach for process variograms

Q. Dehaine[a] and L. Filippov[b]

[a, b] **Université de Lorraine, CNRS, CREGU, GeoRessources lab., 2 rue du Doyen Marcel Roubault, F-54518, Vandœuvre-lès-Nancy, France.**
[a]**E-mail: quentin.dehaine@univ-lorraine.fr**
[b]**E-mail: lev.filippov@univ-lorraine.fr**

In the theory of sampling, variograms have proven to be a powerful tool to characterise the heterogeneity of 1-dimensional lots. Yet its definition and application in sampling for mineral processing have always been limited to one variable, typically ore grade. However this definition is not adapted to sampling for mineral processing where samples contain multiple properties of interest, i.e. variables, such as multiple element grades, grain size, etc. For such cases, the multivariable variogram, originally developed for spatial data analysis, can be used to summarise time variation of multiple variables (e.g. ore characteristics which are important for the process) and highlights the multivariate time auto-correlation of these variables. A case study of low-grade kaolin residue sampling for gravity processing shows that the multivariogram summarises the overall variability and highlights a periodic phenomenon when all variables are taken into account. This example illustrates the potential of the multivariable variogram compared to the classical approach.

Introduction

In every mining project, economic improvement goes through metallurgical assessment by means of series of metallurgical tests performed on the so-called process samples. Process samples are typically extracted from flowing stream, the so-called one-dimensional (1D) lots, at regular interval to obtain representative samples regarding the grade, mineralogical or physical characteristics. The metallurgical tests allow settling the best operating parameters which will allow reaching the desired recoveries and grades, and therefore improve the process. The effectiveness of these process improvements will depend directly on the representativeness of the samples initially collected for the tests.

The Theory Of Sampling (TOS) developed by Pierre Gy[1] gives a simple set of rules to eliminate sampling biases and minimises the sampling error (variance).TOS introduced the semi-variogram (referred as variogram in the text) adapted to sampling purpose as a way to characterise the autocorrelation between the units of a process and the heterogeneity of 1D lots. This tool provide critical information on[2,3]:

■ process variability over time and the magnitude of the different variability components,
■ the lot mean and the uncertainty of a single measurement with respect to the autocorrelation phenomenon,
■ the optimal design and scheme (*i.e.* random, stratified or systematic) for the sampling protocol.

In a typical variographic experiment, a set of N discrete units (*i.e.* increments) is collected from a one-dimensional flowing stream along a time period, representing the 1D lot. The relative heterogeneity associated with a property of interest, A, in a single unit of mass M_i, expressed in the proportion a_i, is defined as:

$$h_i = \frac{a_i - a_L}{a_L} \frac{M_i}{\bar{M}_i}, \ i = 1, \cdots, N \tag{1}$$

where $\bar{M}_i$ is the average mass increment and a_L the proportion of component A in the lot. This relative heterogeneity is dimensionless and hence the component A can describe any intensive property that characterise the material, *e.g.* grade, size distribution,

hardness or specific gravity. The variogram v_j is calculated for a sufficient number of units (up to a maximum of $N/2$) using the equation:

$$v_j = \frac{1}{2(N-j)} \sum_{i}^{N-j} \left(h_i - h_{i+j} \right)^2, \ j = 1, \cdots, N/2 \tag{2}$$

where j is a dimensionless lag-parameter, defining the distance between two increments. Thus, the variogram describes the variation, due to component A, between units as a function of the distance between them. An extensive description of the variographic technique and its practical application can be found in reference papers[2–4].

However, in the field of mineral processing, results from metallurgical tests often depends on several characteristics of the sample. Thus the samples need to be representative not only for one property (*i.e.* component) but for a certain range of properties. In these situations the practical approach is to identify the property with the most heterogeneous distribution and to take only this property into account. The main difficulty of this approach is that it doesn't account for the multivariate nature of heterogeneity, which can lead to underestimation of significant heterogeneity between close neighbours[5]. The importance of taking into account the multivariate character of the heterogeneity is well-known in geostatistics and particularly for spatial data analysis. The first solution to this problem was proposed by Oliver and Webster[6], who suggested to perform a Principal Component Analysis (PCA) on the data and to study the variogram of the first few principal components. While only a few studies have recently applied this approach to chemometrics, they show the usefulness of a variographic modelling based on PCA scores[7,8].

The purpose of this paper is to introduce the multivariate variogram, originally developed for spatial data analysis by Bourgault and Marcotte[9], which is defined in a way similar to that of the traditional variogram but in a multi-dimensional space. This new tool could be more adapted to process sampling of one-dimensional lots as it takes all properties of interest into account. To illustrate this, a variographic study is performed on a process stream from a kaolin mining plant which has been sampled for metallurgical testing.

Multivariate variogram applied to process sampling

We now assume that the heterogeneity contribution is a multivariate measure. If a material is characterised by a number p of parameters, the heterogeneity could therefore be represented as a vector of p individual heterogeneity contributions:

$$H_i = \left[h_1, \cdots, h_k, \cdots, h_p \right]_i \ , \ i = 1, \cdots, N \tag{3}$$

The univariate definition of the variogram is thus no longer adapted and need to be improved. G. Bourgault and D. Marcotte were the first to formalise the principle of a multivariate variogram[9] and it has been widely used for spatial data analysis and mapping since[5,10]. For every metric M it is possible to calculate the multivariate variogram V_j by analogy to the univariate case using the equation:

$$V_j = \frac{1}{2(N-j)} \sum_i^{N-j} \left(H_i - H_{i+j} \right) M \left(H_i - H_{i+j} \right)^t, \ j = 1, \cdots, N/2 \tag{4}$$

where ^{t} symbolise the transpose and M is positive definite $p \times p$ matrix which defines the metric in the calculation of the "distance" between the units. This metric defines the relation between the variables, such metrics are the identity matrix (Euclidian distance). The multivariate variogram is therefore simply the sum of the univariate variograms, or the inverse of the variance-covariance matrix (Mahalanobis distance)[9].

In contrast to variographic analysis of PCA scores, this approach captures all variables in a single variogram. Thus it is possible to calculate the auxiliary functions and consequently the error generating functions for each sampling scheme using classical point-by-point calculation[3], with the exception of the random selection scheme. Indeed the error generating function associated to this sampling scheme is equal to the constitutional heterogeneity of the lot (CH_L) which is defined as the variance of the (multivariate) heterogeneity contribution of all units making up the lot L:

$$CH_L = s^2 \left(H_i \right) = \frac{1}{N} \sum_i^N H_i M H_i^t \tag{5}$$

Material and methods
Material sampling

The samples used in this work were collected with the help of Imerys Ltd., UK. The primary objective of this sampling exercise was to design a protocol which allows collecting a representative sample of a shift (of approximately 2 h) for metallurgical testing. A total of 50 samples (of approximately 25 kg) were collected from the secondary hydrocyclones underflow stream of a kaolin dry mining plant operating at an approximate flow rate of 15 tons/hours corresponding to a micaceous residue which is studied as a potential source of metals[11]. Note that the increments are manually extracted every 2 minutes systematically using a by-pass which diverts the whole stream into the sample collector. This sample extraction protocol may lead to an Increment Extraction Error (IEE) which is difficult to assess.

All increments were weighted then dried directly without dewatering to avoid fine particles loss. Once dried the samples were weighted to estimate their initial pulp density and then riffled to obtain subsamples for particle size analyses. The remaining samples were then crushed and riffled alternatively in accordance with the theory of sampling to obtain representative subsamples for chemical analysis[12].

Analytical methods

The studied material has been sampled with the objective of metallurgical testing by gravity concentration. Thus the analytical methods chosen for the representativeness study must be adapted to this objective. In addition to classical chemical analysis, the critical characteristic of a material for gravity concentration is its size distribution[13].

Chemical Analysis. A set of 18 elements/ oxides were analysed, among which LREE (La, Ce, Nd), Nb and Sn. Representative 10 g aliquots were mixed with Cereox wax (Fluxana® GmbH & Co. KG) and pressed into pellets. Chemical analyses were carried out by Energy Dispersive X-Ray Fluorescence spectroscopy (ED-XRF) using a S2 Ranger (Bruker Corporation) at the GeoRessources laboratory (Vandoeuvre-lès-Nancy, France). The calibration of the XRF used results from Inductively Coupled Plasma Atom-Emission analysis (ICP-AES) for major elements and mass spectral analysis (ICP-MS) for the trace elements realised at the Service d'Analyses des Roches et des Minéraux (SARM-CNRS, Nancy, France).

Particle size analysis. A range of 4 parameters have been retained to describe the particle size distribution of the material: the D10, D50, D90 and Rosin R ammler (RR) slope which represent the particle sizes below which 10%, 50% and 90% of the particles are distributed respectively. And the slope of the size distribution using the Rosin-Rammler model[13]. Particle size analysis has been performed by laser light scattering using a Helium-Neon Laser Optical System MASTERSIZER 3000 (Malvern instruments Ltd.) coupled with a Hydro Extended Volume (EV) sample dispersion unit.

Case study
Experimental individual variograms

The analytical results of 7 selected variables (LREE, Nb, Sn, D10, D50, D90 and RR slope) for 50 micaceous residue samples are presented in Figure 1. The variation illustrates the stream material heterogeneity with time. The results show that, for variables (D10, Nb, Sn, and LREE), the variability expressed by the entire profile is equal to the global variation interval represented by the mean $\pm 2\sigma$ interval, whereas for variables (D50, D90, and RRslope), the variability seems associated with slight trends. However, there are no significant outliers in the profiles. Thus the analytical results can be used directly without any pre-treatment.

It is difficult to interpret from these different scales profiles which variable contributes most to the heterogeneity of the lot. One can thus compare the individual heterogeneity contribution, calculated using equation (1) for each variable (Figure 2). It is observed that the LREE content has the largest overall variability.

From these individual heterogeneity contributions the individual variograms are calculated using equation (2) and the nugget effects V_0 are estimated by backward extrapolation (Figure 3A). The auxiliary functions noted w_j and w'_j are shown in Figure 3C and D. The individual variograms distinguish two main groups, a high-sill variables group (LREE, D90 and Sn) and a low-sill variables group (D10, D50, RR Slope and Nb). The overall range is difficult to estimate using directly the variograms, but the auxiliary functions suggest an overall range around 5-7. The variograms of the low-sill groups
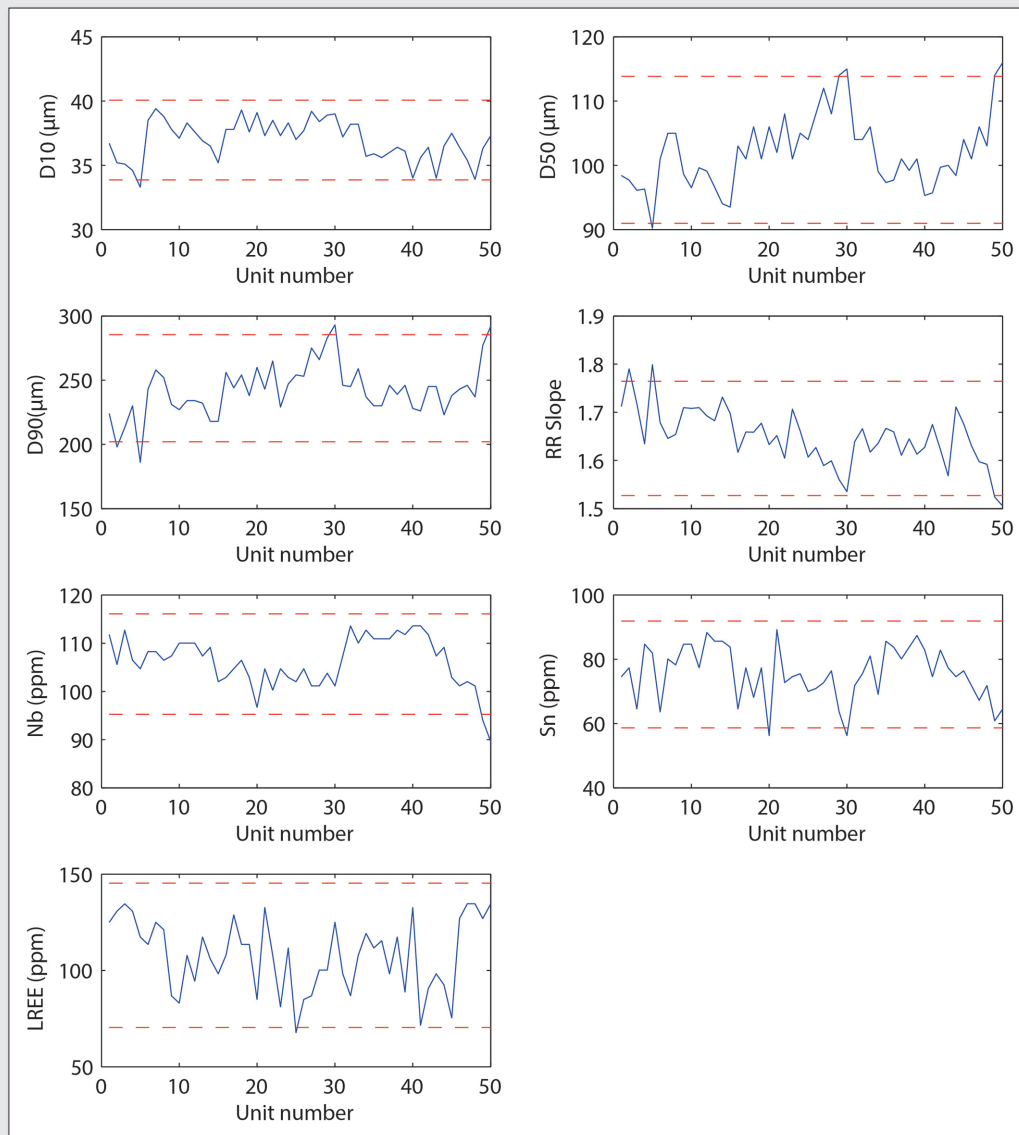
Figure 1. Analytical results characterising the variations of the geochemical compositions (Nb, Sn, and LREE) and size distributions (D10, D50, D90, and RR slope) during time, each unit being extracted at 2 min intervals. The dashed lines represent the mean ±2σ of the analytical results. It can be seen that there is no significant outliers.

appear to 'flat' as well as the variogram of Sn. A minimum can be observed in the variogram of LREE at j = 15 indicating the existence of a possible cyclic fluctuation with a too long period of (*i.e.* j = 15 = 30 min) to see another minimum in the variogram. A similar observation is observed for the D90 variogram (Figure 3B). Indeed, a local minimum is observed at j = 7-9 and a tentative repetition at j = 20 (this can also observed for w_j but not for w'_j since the curve is too smooth). This suggests the existence of some periodic phenomenon for the D90 with a rather short period of approximately 9 lags (*i.e.* j = 9 = 18 min).

The classical conclusion at this point will be to focus the sampling protocol on the LREE content taking care of the periodic phenomena.

Figure 4 show the error generation functions for LREE according to the sampling scheme which is used to choose a protocol with the lowest sampling variance. It can be seen that the 3 sampling schemes are quite close but the systematic sampling stay the sampling scheme with the lowest variance. The recommended

sampling protocol is thus hard to define it could be recommended to use a stratified random sampling or systematic sampling with at least 5 or 10 increments with a sampling frequency higher than two per period of 18 min and 30 min. Since the average shift duration is around 3h this would imply to sample not 5 or 10 but at least 20 increments.

Experimental multivariograms

A multivariate analysis highlights the relationships between the variables which are not taken into account in classic variographic studies. Table 1 presents the correlation matrix for the selected 7 variables. As predicted all the variables referring to the size distribution are strongly to moderately correlated with the exception of D10 which only display moderate correlations (with D50 and D90) or no correlation at all (with RRslope). Sn and Nb are both moderately correlated with the D50 whereas LREE display a clear independency.
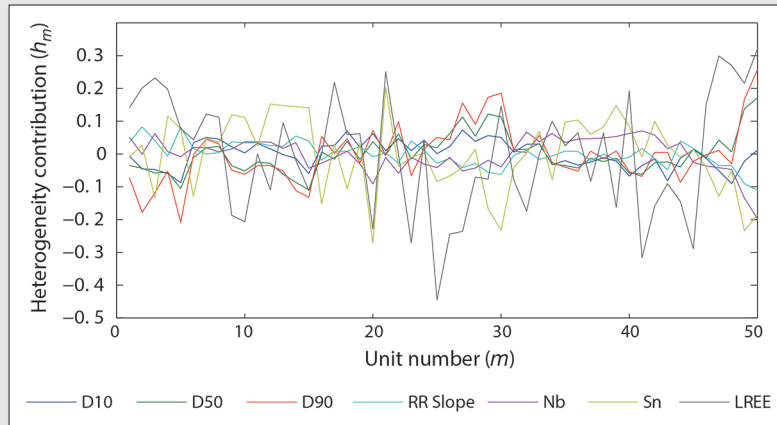
Figure 2. Individual heterogeneity contributions h_m of the 7 variables of interest for the 50 units.
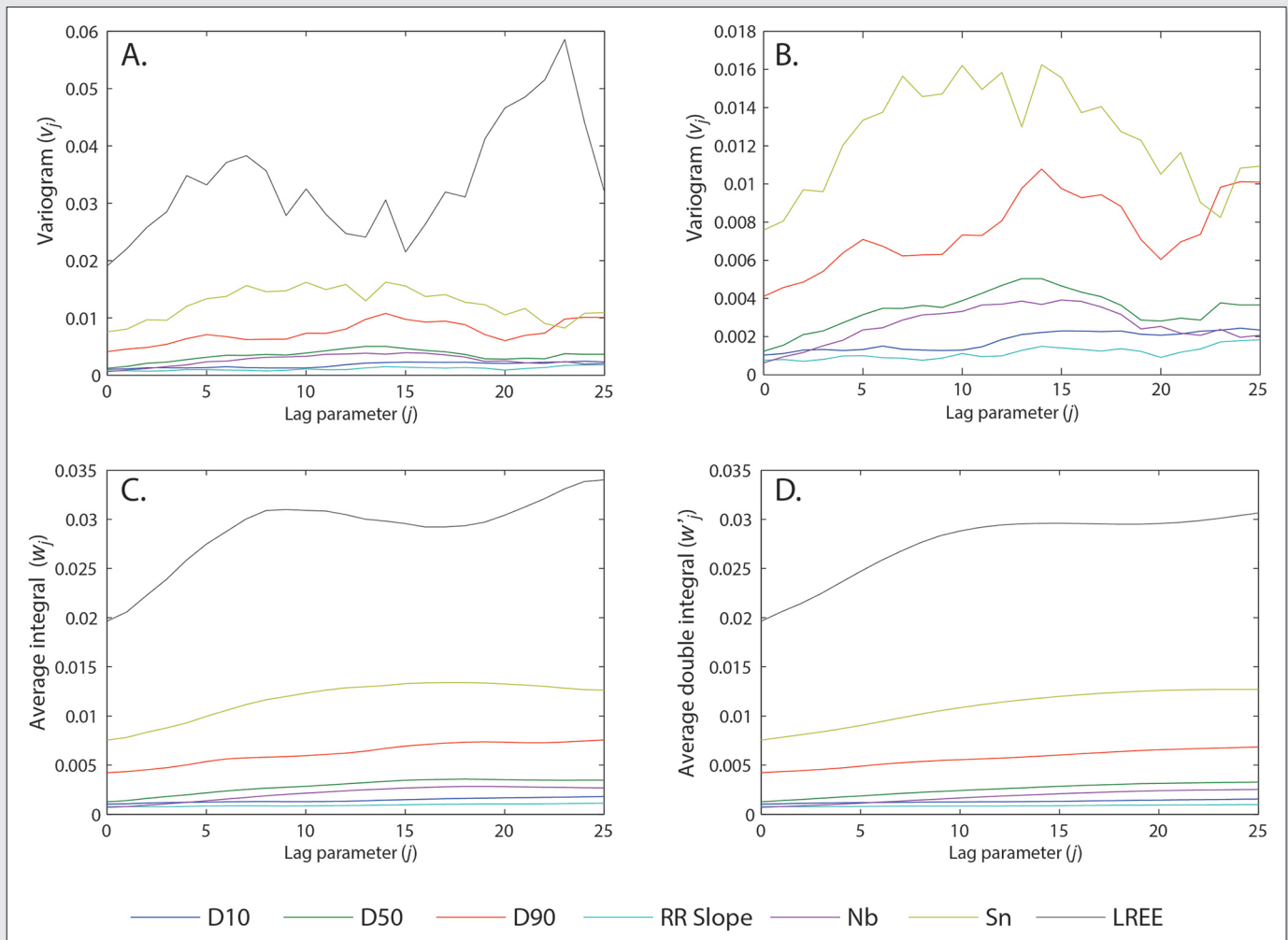

Figure 3. (A) Experimental variograms V_j of the 7 variables of interest. (B) Experimental variograms V_j of 6 of the variables of interest without LREE. (C) Average first order integral w_j and (D) Average second order integral w'_j. A common range of approximately 5-7 lags (10-14 min) is observed.

The multivariogram is computed using formula (4) with Mahalanobis metrics (Figure 5). The general shape of the multivariogram is approximated by a smoothed curve with a spherical model[14]:

$$V_j = \begin{cases} s\left[\dfrac{3j}{2r} - \dfrac{1}{2}\left(\dfrac{j}{r}\right)^3\right], & j \leq r \\ \\ s & , \; j > r \end{cases} \qquad (6)$$
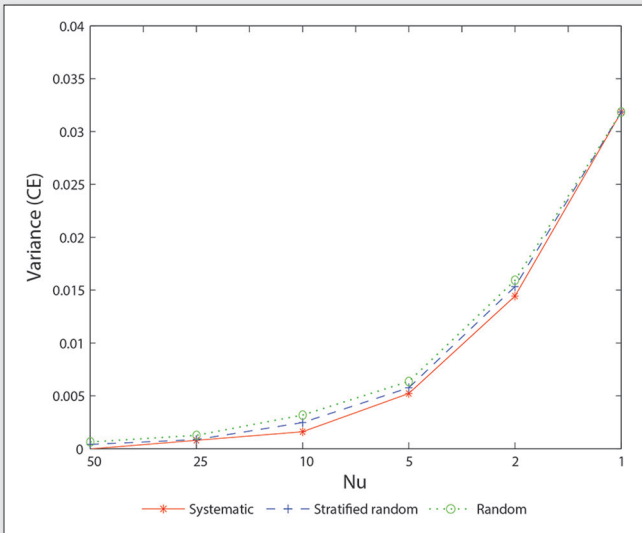
Figure 4. Plot of the error generating functions associated to the LREE content for the 3 sampling schemes as a function of the number of units/increments collected to make the final sample (Nu).

Two minimums at j = 9 and j = 20 on the multivariogram curve suggest a periodic phenomenon with a period of approximately 10 lags (j = 10 = 20 min) which is a results of the periodic phenomena observed in the individual variograms. The multivariogram also displays a high sill of approximately 7, which is due to the metric used in the computation of the multivariogram. Hence the sampling variance is much more important too, and with 5 increments to make the final sample, the sampling variance is still about 0.67 and 0.25 if 10 increments are collected (Figure 6). Based on this multivariogram a sampling protocol could be to take at least 10 increments with a sampling frequency higher than two per period of 20 min.

The multivariogram has allowed proposing a more adapted sampling protocol which takes into account a periodic phenomenon. However the estimated global variance with this approach is very high and implies a large number of increments should be sampled to achieve a reasonably lower sampling variation. This is a direct consequence of the choice of the variables of interest for the variographic study which all contribute at various degrees to the heterogeneity. Note that their importance for the process tested could be completely different from one variable to the other. Thus the sampler must pay attention to the choice of the variables of interest to avoid overestimation of the sampling variance. Another way to have a sampling variance more adapted to the tests for which the samples are collected would be to weight the variables by the mean of an adapted metric.

where s represent the sill of the variogram and r the range. The range suggested by the spherical model is around 11, which is twice the general range observed for the individual variograms.

Table 1. Correlation matrix for all the variables of interest. The high correlation coefficients (>|0.75|) are noted in italic.

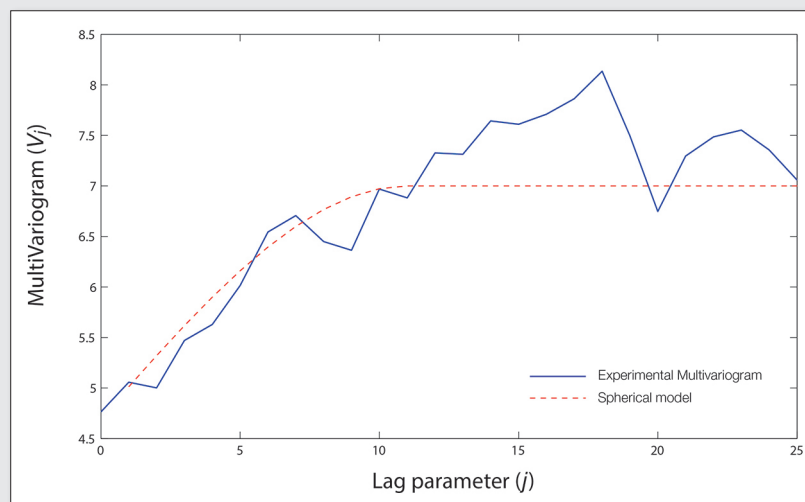| Variables	D10	D50	D90	RR Slope	Nb	Sn	LREE
D10	1.00	0.61	0.62	-0.20	-0.24	-0.30	-0.25
D50	0.61	1.00	0.93	-0.78	-0.64	-0.64	0.02
D90	0.62	0.93	1.00	-0.88	-0.54	-0.54	-0.02
RR Slope	-0.20	-0.78	-0.88	1.00	0.48	0.45	-0.09
Nb	-0.24	-0.64	-0.54	0.48	1.00	0.52	-0.14
Sn	-0.30	-0.64	-0.54	0.45	0.52	1.00	-0.09
LREE	-0.25	0.02	-0.02	-0.09	-0.14	-0.09	1.00

Figure 5. Multivariogram for the 7 variables of interest and fitted spherical model. The range given by the spherical model is approximately 11 with a sill around 7. However two minimums at j = 9 and j = 20 suggest a periodic phenomenon with a period of approximately 20 min.
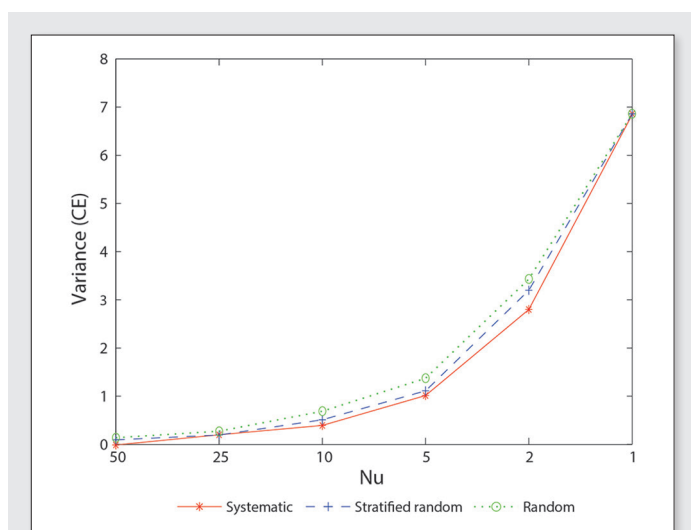
Figure 6. Plot of the error generating functions of the multivariogram for the 3 sampling schemes as a function of the number of units/increments collected to make the final sample (Nu).

Conclusions

The multivariate approach of process variograms described in this work has allowed a better description of the heterogeneity of a material taking into account all the variables of interest simultaneously. This approach give to the sampler the opportunity to choose all the parameters that characterise the material for a given objectives and to use the multivariate variogram as a summarising tool to describe the variability of this material and to design an adapted sampling protocol.

Acknowledgments

References

1. P. Gy, *Sampling for Analytical Purposes*. John Wiley & Sons, Chichester (1998).
2. P. Gy, "Sampling of discrete materials III. Quantitative approach—sampling of one-dimensional objects". *Chemometrics and Intelligent Laboratory Systems*. **74**, 39–47 (2004). doi: http://dx.doi.org/10.1016/j.chemolab.2004.05.015
3. L. Petersen, K.H. Esbensen, "Representative process sampling for reliable data analysis—a tutorial". *Journal of Chemometrics*. **19**, 625–647 (2005). doi: http://dx.doi.org/10.1002/cem.968
4. P. Minkkinen, "Practical applications of sampling theory". *Chemometrics and Intelligent Laboratory Systems*. **74**, 85–94 (2004). doi: http://dx.doi.org/10.1016/j.chemolab.2004.03.013
5. G. Bourgault, D. Marcotte, P. Legendre, "The multivariate (co)variogram as a spatial weighting function in classification methods". *Mathematical Geology*. **24**, 463–478 (1992). doi: http://dx.doi.org/10.1007/bf00890530
6. M.A. Oliver, R. Webster, "A geostatistical basis for spatial weighting in multivariate classification". *Mathematical Geology*. **21**, 15–35 (1989). doi: http://dx.doi.org/10.1007/bf00897238
7. P. Minkkinen, K.H. Esbensen, "Multivariate variographic versus bilinear data modeling". *Journal of Chemometrics*. **28**, 395–410 (2014). doi: http://dx.doi.org/10.1002/cem.2514
8. Z. Kardanpour, O.S. Jacobsen, K.H. Esbensen, "Soil heterogeneity characterization using PCA (Xvariogram) - Multivariate analysis of spatial signatures for optimal sampling purposes". *Chemometrics and Intelligent Laboratory Systems*. **136**, 24–35 (2014). doi: http://dx.doi.org/10.1016/j.chemolab.2014.04.020
9. G. Bourgault, D. Marcotte, "Multivariable variogram and its application to the linear model of coregionalization". *Mathematical Geology*. **23**, 899–928 (1991). doi: http://dx.doi.org/10.1007/bf02066732
10. R. Kerry, M.A. Oliver, "Variograms of Ancillary Data to Aid Sampling for Soil Surveys". *Precision Agriculture*. **4**, 261–278 (2003). doi: http://dx.doi.org/10.1023/a:1024952406744
11. Q. Dehaine, L.O. Filippov, "Rare earth (La, Ce, Nd) and rare metals (Sn, Nb, W) as by-product of kaolin production, Cornwall: Part1: Selection and characterisation of the valuable stream". *Minerals Engineering*. (2014). doi: http://dx.doi.org/10.1016/j.mineng.2014.10.006
12. L. Petersen, C.K. Dahl, K.H. Esbensen, "Representative mass reduction in sampling—a critical survey of techniques and hardware". *Chemometrics and Intelligent Laboratory Systems*. **74**, 95–114 (2004). doi: http://dx.doi.org/10.1016/j.chemolab.2004.03.020
13. B.A. Wills, T. Napier-Munn, *Wills' Mineral Processing Technology*. Elsevier (2005).
14. R. Webster, M.A. Oliver, *Geostatistics for Environmental Scientists*. John Wiley & Sons, Ltd, Chichester, UK (2007).