

Article

Sampling Theory for Mineral Process Flows

Geoffrey Lyman

Materials Sampling & Consulting, Mapleton, QLD 4560, Australia; geoff.lyman@gmail.com

Abstract: Current literature dealing with the theory of sampling process flows is limited in scope, a situation addressed by papers presented at conferences by the author. This paper provides a summary of methods to quantify process sampling variance in a sound statistical manner. Two methods of calculation of the process sampling variance are presented here with models for robust estimation of the process covariance/variogram. The strategies for process sampling to acquire the data for process characterisation are described. The method for process characterisation can be based on special sampling campaigns, on the analysis of on-line measurement data, or on the analysis of historical data, which may be based on composite samples, such as shift samples. These scenarios are all dealt with as well as methods based on analysis of process dynamics. The latter analysis demonstrates that use of covariance functions drawn from geostatistical practice can be inappropriate for process analysis. The work is supported with plant data examples and quantitative analyses of process conditions.

Keywords: sampling theory; process dynamics; variograms

1. Introduction

There are three components of uncertainty or variance due to sampling a mineral process flow. The first is ‘distributional variance’ due to the variation in the mineral grades with time (or tonnage) at the primary sampling point. The second uncertainty is ‘sampling variance’ due to the grade and size variations of the particles in the flow, which has been fully described in a previous paper [1]. The last uncertainty is due to the ‘analytical variance’ associated with the chemical or physical analysis of the final sub-sampled aliquot; it is specific to the method of analysis and beyond the scope of this paper. The first two uncertainties express themselves in the primary sample and subsamples thereof. An optimal sampling, sample preparation, and analysis protocol is achieved by balancing these sources of uncertainty with the analytical variance.

The objective of this paper is to clarify the first component, distributional variance. Because this component may often be large compared to the other two, it is often referred to generically as ‘the sampling variance’, a practice also used in the paper below, but not to be confused with the second component, ‘sampling variance’, which is due only to the particulate nature of the flow. Distributional variance is often the largest component of the three because of the uncontrollable in situ variability of the resource, materials handling arrangements, the opportunities for sampling the front end of the process, and the physical and economic constraints on sampling equipment during the processing to final output streams.

As is the case with the statistical theory of particulate sampling [1], it is necessary to have a statistical framework or model within which to describe variations in process flows. The simplest form of model for the variations is a stationary random function. This model is, in fact, the basis for geostatistics [2–4], which is used to describe the spatial variation of mineral grades within mineral deposits. Geostatisticians have developed variants of the simplest form of random functions to enhance its descriptive power and to deal simultaneously with several mineral types [3]. In this paper, the random function is constrained to the simplest type, a stationary Gaussian random function.



Citation: Lyman, G. Sampling Theory for Mineral Process Flows. *Minerals* **2023**, *13*, 922. <https://doi.org/10.3390/min13070922>

Academic Editor: William Skinner

Received: 22 April 2023

Revised: 1 July 2023

Accepted: 4 July 2023

Published: 9 July 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

The advantage of a Gaussian random function is that it requires only one descriptor, which is its covariance function. Mild departures from the Gaussian form can be dealt with by suitable approximations.

The process is sampled either by extraction of physical increments of ore from the material flow or by use of an on-line analyser to interrogate the elemental composition of the material in situ. The sampling is usually carried out at convenient locations on conveyors, slurry pipes, or launders.

Section 2 provides the necessary mathematical framework of Gaussian random functions and derives the sampling variance from the properties of the random function. Section 3 provides the tools required to estimate the covariance function of the process flow. Section 4 discusses the acquisition of process data and how this influences the information gained from a variogram. Section 5 applies this methodology to both simulated and actual process data to show its usefulness. Section 6 combines the process distributional variance with sampling variance and analytical variance from subsequent sample preparation and analysis procedures to arrive at a total sampling variance.

2. Derivation

A second-order stationary Gaussian random function is a collection of Gaussian (normally distributed) random variables that are connected through a covariance function. For the stationary random function, the covariance between a value attached to a time t and one at $t + h$ is given by $C_X(h)$, where, if the random function is denoted by $X(t)$,

$$C_X(h) = \text{cov}\{X(t), X(t+h)\} \quad (1)$$

The restriction to second-order stationary random functions requires both the means and variances of the random variables to be constant; they are not a function of time t . This assumption of stationarity is enormously important as lack of stationarity will distort the estimates made of the covariance function and lead to substantially incorrect estimates of sampling variance. This issue of stationarity is ignored in the current sampling literature [5,6]. The mean of the random function is denoted as μ_X and its variance as σ_X^2 . The Gaussian random function is the simplest as a Gaussian random variable is fully described by the mean and variance. The probability density function (PDF) is a function only of the mean and variance and all central moments higher than the second (the variance is the second central moment) are a function only of the variance.

The covariance of a random variable is defined as

$$C_X\{h\} = E\{(X(t) - \mu_X) \cdot (X(t+h) - \mu_X)\} \quad (2)$$

where $E\{\cdot\}$ denotes an expected value. Recall that the correlation coefficient between two random variables is simply the covariance divided by the product of their standard deviations,

$$\text{corr}\{X, Y\} = \frac{\text{cov}\{X, Y\}}{\sigma_X \sigma_Y} \quad (3)$$

A correlation coefficient of unity (1) means that one random variable tracks the other perfectly and one of zero means that one does not follow the other at all. The behaviour of a covariance function is restricted in some important ways [2–4]:

- The covariance is, at most, a periodic function or a decreasing one; it can resemble a decaying cosine function.
- It must be a so-called positive semi-definite function, meaning that for a discretely defined function, it forms a matrix that has eigenvalues that are positive or zero.
- For a real-valued continuous random function, the covariance function must be the cosine transform of a function that has the properties of a PDF (zero or positive everywhere).
- The covariance function is an even function so that it is reflected through the ordinate axis, $C(-h) = C(h)$.

These properties of the covariance function mean that not all functional forms can be used as a covariance function. This fact is also ignored in current sampling theory.

As it is in geostatistics, it is common practice to employ the ‘variogram’ function instead of the covariance function; this function was used by Matheron [3] in the development of geostatistics because the conditions on the associated random functions are less restrictive. However, in almost all practical simulations of random functions, the covariance function is used even if the data were analysed using a variogram calculation.

The variogram (strictly called the ‘semi-variogram’) $\gamma(t)$ is related to the covariance function by

$$\gamma(h) = \sigma_X^2 - C_X\{X(h)\} \quad (4)$$

and since

$$\text{cov}\{X(t), X(t)\} = C_X(0) = \sigma_X^2 \quad (5)$$

the variogram must have a value of zero at the origin $\gamma(0) = 0$. It is therefore, at most, a periodic function that rises from zero to oscillate about a value of σ_X^2 , which is called the ‘sill’ in geostatistical parlance. The variogram is everywhere positive. Figure 1 illustrates several forms of the covariance function and the corresponding variograms.

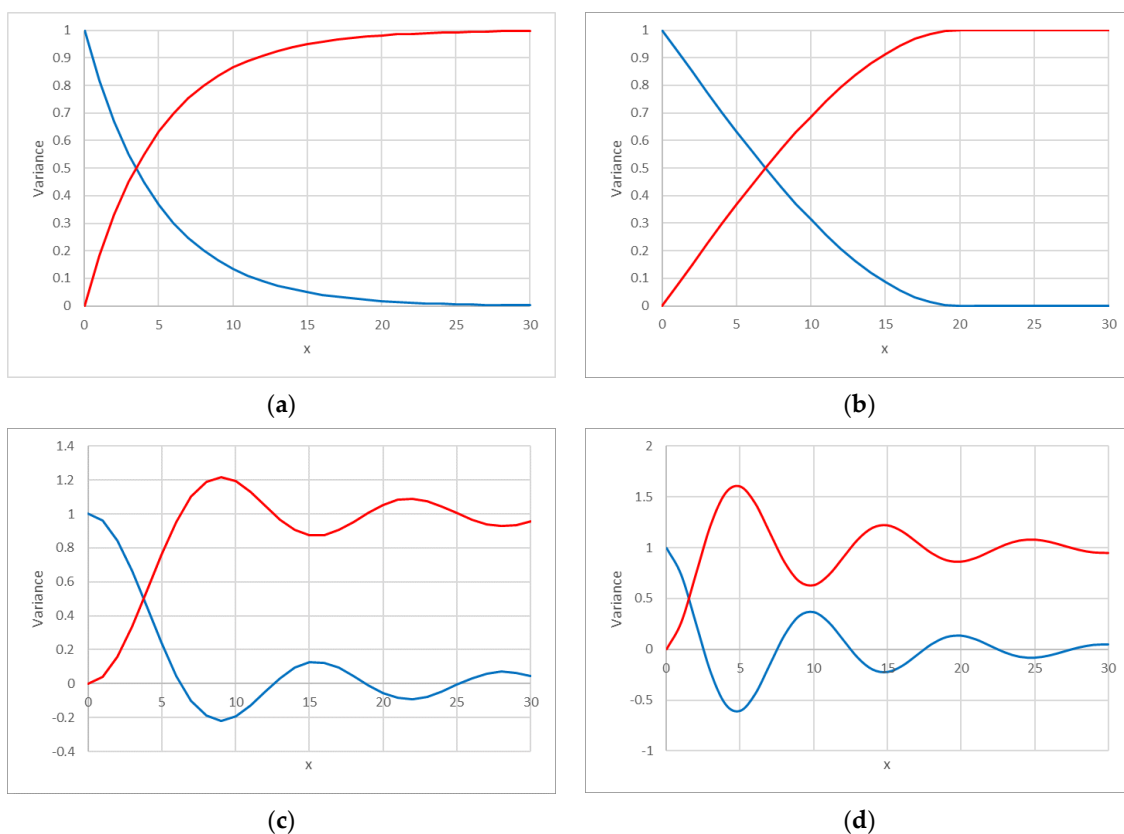


Figure 1. Examples of continuous covariance (blue) and variogram (red) functions. (a): exponential. (b): spherical. (c): hole effect. (d): damped cosine.

The practical applications of sampling theory deal almost exclusively with discrete random functions, which are defined at a series of points in increasing time, $\{t_1, t_2, \dots, t_j, \dots\}$. This happens because measurements of the process grades are taken or are available at discrete times rather than strictly continuously. The variogram and covariance functions can therefore be estimated at a series of time increments $\Delta t_{ij} = t_j - t_i$; $t_j > t_i$. While equally spaced times greatly aid in calculating and graphing the variogram estimate, equal spacing is not a requirement for detailed analysis using either variogram or covariance techniques; indeed, equal spacing restricts the information that can be extracted in a serious manner.

There are important distinctions to be made in the analysis method between collecting single physical increments of material, collecting multiple increments for incorporation into a composite sample, or making continuous on-line measurement of elemental grade, which is integrated/averaged over a fixed time interval and then recorded. These issues of sample timing and aggregation are discussed further in Section 4.

The sampling variance is the variance of the difference between the true average grade of the material flowing past the sampling point and the grade of a sample collected by extracting increments from the flow or making measurements over specific time intervals with an on-line analyser. The usual derivation of this variance takes the true variation in grade (analyte content) to be a continuous stationary random function $X(t)$ and the sample to have the average grade of finite increments collected at times t_i ; $i = 1, 2, \dots, n$.

The impact of random variation in the mass flow of the solids in the process stream is usually ignored and is omitted in the analysis below. Such mass flow variations can also be modelled with a random function, as shown in the author's publications [7,8], which conclude that the mass flow variations must be large to impact the sampling variance when the mass flow and grade variations are uncorrelated.

Note, however, that instead of using time as the independent variable of the random function, the totalized mass of material passing the sampling point can be used. In such a case, the increments extracted from the process stream should usually be of equal mass if the increments are intended to represent the average grade of material passing the sampling point.

The true average grade in the process stream over a period T is,

$$\bar{X} = \frac{1}{T} \int_0^T X(t) dt \quad (6)$$

and the grade of each sample is,

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X(t_i) \quad (7)$$

given that the increments are all of the same mass and the t_i are located within the interval T . The quantity to be calculated is the variance of the difference in these two grade values,

$$Y = \bar{X} - \bar{X}_n \quad (8)$$

The expected value of the difference is obviously zero. The variance is,

$$\begin{aligned} \text{var}\{Y\} &= E\{(Y - E\{Y\})^2\} = E\{Y^2\} \\ &= E\{[\bar{X} - \mu_X - (\bar{X}_n - \mu_X)]^2\} \\ &= E\{(\bar{X} - \mu_X)^2 - 2(\bar{X} - \mu_X)(\bar{X}_n - \mu_X) + (\bar{X}_n - \mu_X)^2\} \\ &= \text{var}\{\bar{X}\} - 2\text{cov}\{\bar{X}, \bar{X}_n\} + \text{var}\{\bar{X}_n\} \end{aligned} \quad (9)$$

which can be written in terms of the variogram, $\gamma(t)$, as

$$\begin{aligned} \text{var}\{Y\} &= -\frac{1}{T^2} \int_0^T \int_0^T \gamma(|u-v|) du dv + \frac{2}{nT} \sum_{i=1}^n \int_0^T \gamma(|t-t_i|) dt - \\ &\quad - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma(|t_i-t_j|) \end{aligned} \quad (10)$$

Geostatisticians have the same problem when estimating the grade of a deposit from drill hole samples taken at discrete locations, and this variance is called the 'estimation

variance'. The geostatistical analysis of this problem [2] involves two so-called auxiliary functions. The first is $\chi(u)$, which is the average value of the variogram over an interval u from the origin.

$$\chi(u) = \frac{1}{u} \int_0^u \gamma(t) dt \quad (11)$$

The second is a time-weighted average of $\chi(u)$ over the same interval.

$$F(u) = \frac{2}{u^2} \int_0^u t \chi(t) dt \quad (12)$$

By making a clever change of variable, the second function can be shown to be,

$$F(u) = \frac{2}{u^2} \int_0^u (u - \tau) \gamma(\tau) d\tau \quad (13)$$

In general, the timing of increment extraction can follow three different schemes. The sample times within a sampling period T may either be 'random' by uniform random placement within T , 'random stratified' by random placement within n equal intervals in T , or 'systematic' by placement at strictly equal intervals $\Delta t = T/n$.

Equation (10) can now be simplified for each of these cases, neglecting the first case of uniform random sampling because it is impractical to implement. The simplification for random stratified sampling is complex and will not be shown here. The evaluation of the second and third terms in the variance requires the expected values of the expressions over the random position of the sample within each stratum. However, the result of this analysis is straightforward, giving a sampling variance of

$$\text{var}_{RS}\{Y\} = \frac{F(\Delta t)}{n} \quad (14)$$

Note that this expression does not require knowledge of the variogram over its full range. Only the part of the variogram up to the time Δt is required. Any expression of periodicity in the later parts of the variogram is irrelevant as the randomising of the location of the sample within the stratum effectively destroys any influence of periodicity in the limit of a sufficient number of samples being taken.

For the special case of the exponential variogram with unit sill,

$$\gamma(t) = 1 - e^{-t/\alpha} \quad (15)$$

which is often a good choice for the form of the variogram, it is possible to determine the formula for the sampling variance analytically. For random stratified sampling with the exponential variogram of unit variance,

$$\text{var}_{RS}\{Y\} = \frac{1}{n} \left[1 - 2 \frac{\alpha}{\Delta t} + 2 \left(\frac{\alpha}{\Delta t} \right)^2 \left(1 - e^{-\Delta t/\alpha} \right) \right] \quad (16)$$

When dealing with systematic sampling, the level of simplification that can be achieved is not so great. The exact result for unit variance of the variogram is

$$\text{var}_S\{Y\} = F(T) - \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n \gamma(|j-i|\Delta t) \quad (17)$$

noting that if a matrix is constructed such that its ij th element is $\gamma(|j-i|\Delta t)$, the last term above is the average value of all the elements.

For the special case of an exponential variogram and systematic sampling only, the sampling variance for unit variance can be written as,

$$\text{var}_S\{Y\} = \frac{1}{n} + 2r \left[\frac{1}{n(1-r)} - \frac{1-r^n}{(n(1-r))^2} \right] - 2\frac{\alpha}{T} + 2\left(\frac{\alpha}{T}\right)^2 (1 - e^{-T/\alpha}) \quad (18)$$

where

$$r = e^{-\Delta t/\alpha} \quad (19)$$

and

$$T = n\Delta t \quad (20)$$

For any other functional model of the variogram, the summation term probably cannot be simplified, but the sampling variance expressions can still be readily evaluated. In Figures 2 and 3, T is a specified multiple of the parameter α . There is not much difference in the results, but systematic sampling provides a slightly lower variance than random stratified sampling in the case of an exponential variogram.

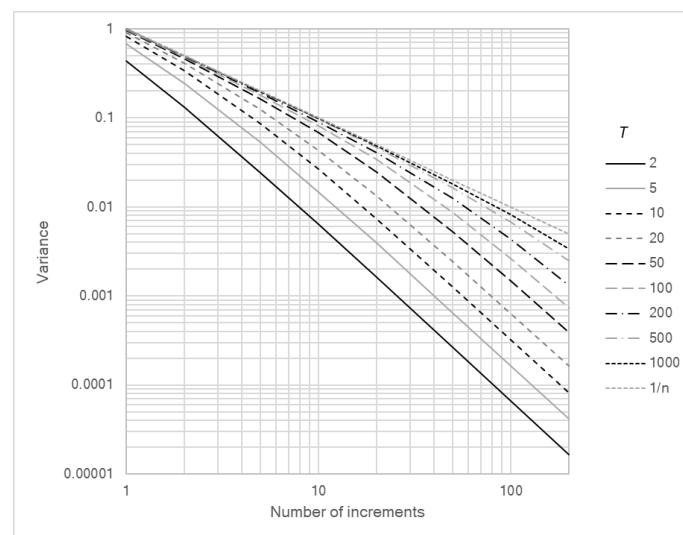


Figure 2. Sampling variance for a unit variance process as a function of numbers of increments and the period as a multiple of the parameter α Exponential covariance and random stratified sampling. Exact results.

In the absence of correlation between grades with time (the covariance function is zero at a time greater than zero and the variogram is equal to the sill beyond the origin—a ‘flat’ variogram), the sampling variance is given by the line for a value of $1/n$. The values in the figures are exact for the exponential variogram.

The effect of the strength of the correlation on the sampling variance is very substantial. A low value of T corresponds to strong correlation over the sampling period, while a large one indicates weak correlation with time within the sampling period.

It is often very useful to be able to compute examples, or ‘realisations’, of a random function with a known covariance function. The realisations are always calculated as discrete versions of the random function, calculated at chosen points on the time axis. There is a method that provides realisations that correctly follow the desired covariance function. The method is best based on the calculation of the covariance matrix for the random function at the chosen time values and the subsequent calculation of the eigenvalue decomposition of the covariance matrix. The times need not be equally spaced. Because the covariance matrix must be positive semi-definite, all the eigenvalues must be positive or zero. The calculation for a large matrix is a relatively large task, but can be carried out

with exact results. Once the decomposition has been calculated, it can be used to rapidly generate any number of realisations.

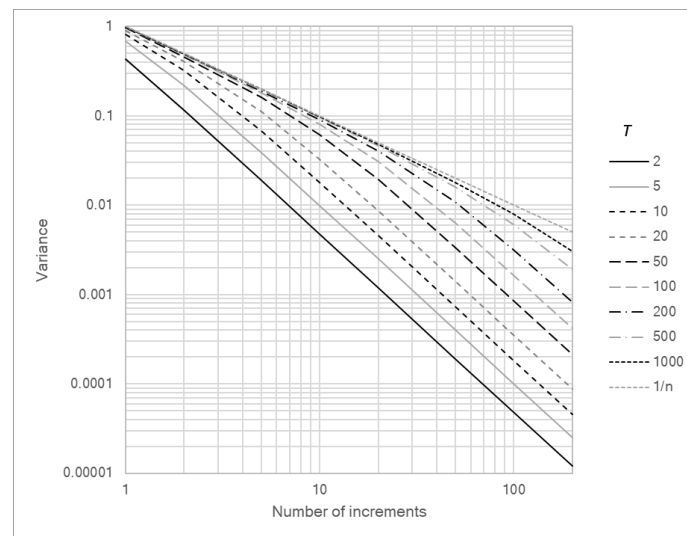


Figure 3. Variance for a unit variance process as a function of numbers of increments, n , and the period, T , as a multiple of the parameter α Exponential covariance and systematic sampling. Exact results.

This method of decomposition of the covariance matrix for discrete Gaussian random functions can be exploited for the calculation of the sampling variance in a very powerful way that is computationally useful. The covariance matrix for a realisation of the random function over the sampling period is denoted by Σ and it has a decomposition of

$$\Sigma = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^T \quad (21)$$

where $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues and $\mathbf{W}$ is an orthonormal matrix of eigenvectors. The points in time at which increments are extracted or which correspond to measurements are included in the set of times used to construct the covariance matrix. The set of points will, in general, be large so that the random function is well-defined.

To generate a discrete realisation $\mathbf{x}$ of the random function, one has

$$\mathbf{x} = \mathbf{W}\mathbf{\Lambda}^{1/2}\boldsymbol{\varepsilon} \quad (22)$$

where $\mathbf{\Lambda}^{1/2}$ is a diagonal matrix containing the square root of the eigenvalues in $\mathbf{\Lambda}$ and $\boldsymbol{\varepsilon}$ is a vector of independent random numbers of a chosen distribution, usually normal with unit variance.

The true average assay over the sampling period is

$$\bar{X} = \frac{1}{M} \sum_{i=1}^M X_i \quad (23)$$

where there are M points in the realisation. The points at which increments are extracted can be denoted in general terms as an indicator function $S(i)$ defined on the set $\mathcal{S}$ of M points such that

$$\begin{aligned} S(i) &= 1; i \in \text{increment or measurement} \\ &= 0; \text{otherwise} \end{aligned} \quad (24)$$

The sample is then defined as

$$X_S = \frac{\sum_{i=1}^M S(i)X_i}{\sum_{i=1}^M S(i)} \quad (25)$$

The sum in the denominator can be denoted as n , since it is assumed that there are n increments taken. The difference between a sample and the true average is

$$\begin{aligned} y &= \bar{x} - x_S \\ &= \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^M w_{ij} \lambda_j^{1/2} \varepsilon_j - \frac{1}{n} \sum_{i=1}^M \sum_{j=1}^M S(i) w_{ij} \lambda_j^{1/2} \varepsilon_j \\ &= \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^M \lambda_j^{1/2} \varepsilon_j w_{ij} \left[1 - \frac{M}{n} S(i) \right] \end{aligned} \quad (26)$$

Defining

$$t_j = \frac{1}{M} \sum_{i=1}^M w_{ij} \left[1 - \frac{M}{n} S(i) \right] \quad (27)$$

the difference is then

$$y = \sum_{j=1}^M t_j \lambda_j^{1/2} \varepsilon_j \quad (28)$$

which is a simple linear function of the random variables ε_j , which are independent and of zero mean and unit variance.

The mean of y is clearly zero and its variance must be

$$\sigma_y^2 = \sum_{j=1}^M t_j^2 \lambda_j \quad (29)$$

and this is the sampling variance. The values of t_j^2 can be calculated from (27). As long as the sampling times are included in the covariance matrix together with a sufficient number of other points to provide a good estimate of the true average value of the random function over the sampling period, this method can accommodate any sampling scheme and any covariance function. The result from this method will differ slightly from the estimators presented since the true value of the average grade of the lot being sampled is based on a discrete function rather than a continuous one. It also depends on the nominated set of sampling times for the chosen sampling scheme. The sampling variance for systematic sampling will depend weakly on the placement of the sample in the first stratum. Similarly, for random stratified sampling, the result will depend on the choice of sampling times; an average value requires averaging values of t_j^2 over all possible placements in the sample increments.

3. Estimation of the Variogram and Covariance Function

The current sampling literature and even the current tools for geostatistical analysis do not exploit the statistical properties of the variogram or covariance function. Many ‘packages’ for geostatistical analysis allow the user to fit a variogram ‘by eye’, which completely ignores the statistical properties of the estimated variogram. This rather simplistic approach is not mathematically sound and will adversely influence the outcomes when a variogram is used for estimating the grade variations in a resource. In the context of process analysis, use of a rather arbitrary variogram will cause the estimate of the sampling variance to be incorrect.

There are a number of issues with current variogram/covariance function estimation in the sampling literature. Attention will be confined here to the sampling context and the geostatistical context ignored, although a number of the issues in sampling are also relevant to geostatistics.

The first issue is the need to have the covariance function estimate admissible as a covariance function; it must be positive definite (the variogram negative definite). The standard estimator for the variogram is a method of moments one proposed by Matheron [3]:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N_{\mathbf{h}}} \sum_{i=1}^{N_{\mathbf{h}}} [z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})]^2 \quad (30)$$

where $N_{\mathbf{h}}$ is the number of pairs of data used in the calculation at distance $\mathbf{h}$. The distance is denoted in general as a vector (bold type), as in the geostatistical context, the displacements are defined in three dimensions. The corresponding covariance function estimate is calculated as

$$\hat{C}(\mathbf{h}) = \hat{\sigma}^2 - \hat{\gamma}(\mathbf{h}) \quad (31)$$

The covariance determined this way is not positive definite [3], and therefore working with this variogram estimator to determine sampling variances is not correct; an admissible positive semi-definite function must be fitted to the covariance before variance calculations can be made. The current sampling literature suggests that using the raw variogram for variance calculations is acceptable ([6], Sections 7.9.4 and 7.12).

The sampling literature focusses on the parabolic functional model for the variogram introduced by Gy [5]. He states that the variogram can be concave upwards or downwards, which is not true, as parabolic variograms that do not have a constant slope (a linear variogram) or a slope that decreases with distance from the origin are not admissible variograms [7]. Such a model is inflexible. An admissible covariance function has a cosine transform that is everywhere positive or zero [9].

There are many possible covariance functions; many have been studied by geostatisticians [2–4]. Most functional forms are based on geometric concepts (the ‘spherical’ model is related to the area of overlap of two equal spheres as a function of their separation) and are not suitable in the context of process variation.

For process applications, covariance functions should be based on process dynamical considerations. For example, the exponential covariance is the covariance function for the output from a perfect mixer which is fed with a random input.

The next issue, which is a very practical one and is ignored in the current and geostatistical literature, is the calculation of measures of uncertainty in the estimation of the variogram using (30). The variogram itself is a random function, [3] and, as such, it will usually be just one realization of many; each variogram value is itself also a statistic with a mean, variance, and PDF.

Since there are $N_{\mathbf{h}}$ pairs of data available at a distance $\mathbf{h}$ and the estimate of the variogram at that point is the mean of the available squared differences, it is possible to use the set of squared differences to calculate a standard deviation of the mean value as a measure of its precision. Geostatisticians sometimes plot what they call the ‘variogram cloud’, which is a plot of all mean squared differences, and this can provide a measure of the precision of definition of the variogram. A simple standard deviation of the variogram mean, used to put error bars on each variogram point estimated, gives an excellent guide to the nature of the variogram. The author has seen variograms interpreted as periodic when the addition of error bars clearly indicates that the apparent (weak) periodicity does not, in fact, exist and is a statistical artifact. Such misinterpretation can cause a useless ‘witch-hunt’ in the process context, where a periodic component may be an indicator of poor process performance or control. The variogram itself is a random function [3] and the neighbouring points are strongly correlated [7,9], which is certain to give rise to a ‘wavy’ behaviour of the variogram point estimates.

Another practical issue arises; using the variogram as a statistical tool for process analysis assumes a stationary process where the mean and covariance of the data are supposed to be constant. The interpretation of the variogram based on data that do not appear to be stationary over the analysis period will be incorrect. It is therefore mandatory to ‘detrend’ the data used so that the data appear to have a constant mean within the data set and that the magnitude of the variation with time seems constant.

Similarly, mixing two sequences of data where the variances of each are dissimilar will likewise produce an outcome of little value. The available examples of the use of the variogram by the sampling community rarely, if ever, involve detrending the data prior to variogram calculation. This indicates a significant misunderstanding of the variogram as a statistical analysis tool.

The author uses a very effective and flexible tool to detrend data sets, which is known as locally weighted regression [10]. This method employs a variable size of window on the time axis within which a weighted polynomial regression is carried out at a nominated point on the time axis (there need not be a data point at the chosen point). Data closest to the time point are weighted more heavily, with the weighting going to zero at the ends of the window, which is centred on the time point once the window ‘fits’ entirely within the data set. Weighting is also applied so that ordinate values distant from the regression do not pull the regression away from the majority of points. Correct application of this latter weighting requires a second pass through the calculation, the first pass being used to estimate the uncertainties in the data relative to the regression model. Altering the size of the window controls the amount of smoothing of the estimated data set mean. A wide window removes only low-frequency components of the random function, while narrowing the window removes higher and higher frequency components. The detrending and variogram estimation methods can be packaged (Excel) into a very simple pre-processing procedure.

A final point involves the handling of unequally spaced measurements on the time axis or spatially. Geostatisticians deal with this issue by creating an optimized regular grid and then moving points to the nearest grid points. In reality, this is a very poor process as it immediately removes higher-frequency information from the variogram, limiting the estimation of the behaviour of the variogram near the origin. Unequal spacing can be correctly dealt with using more complex estimation methods that are now computationally practical, while they were not in the early days of geostatistics.

Having usefully detrended the data, a variogram or covariance function can be found via the maximum likelihood method (see Section 3.1) after choosing a functional form for the covariance function. The Matheron variogram calculation can also be run to provide some guidance in the choice of covariance functional form. It may be useful to fit an admissible variogram/covariance function to the raw variogram. This latter step is commonly, but incorrectly, performed by ordinary least squares or weighted least squares using the standard deviations of the variogram estimates; this method seems attractive and simple but is statistically unsound. However, with a sufficiently large data set, it can provide a useful guide to the form of the admissible function.

Correctly fitting a functional form to variogram values is hampered by the complex covariance structure for the variogram estimator. There is a high covariance between adjacent points in the variogram [7,9]. While a correct method for the calculation of the covariance matrix of the variogram points was outlined by Cressie ([4], Section 2.4.2), who recognised that the variogram estimator can be written as a quadratic in matrix form, he did not pursue the use of that information in variogram estimation.

The author followed on from Cressie’s work and showed how the full covariance for the variogram estimates can be calculated, given a proposed form of the covariance matrix. It is also shown that the variogram estimator is a weighted sum of chi-squared random variables with one degree of freedom, which shows that the estimator has a skewed distribution with a substantial upward tail [7].

The complexity of the variogram estimator can be avoided using a method of covariance function estimation based on the maximum likelihood method. The likelihood of the

data can be formulated when it is assumed that the measurements are normally distributed, given a functional form for the covariance. Minimum variance estimates of the covariance function parameters can then be made and, for large enough data sets, the covariance matrix of the parameter estimates can be estimated. Further, reasonable tests of goodness of fit of the covariance function to the data can be made to find a function that best describes the data using likelihood ratio tests. A very important aspect of this estimation method is that the exact location of data on the time axis can be taken into account, eliminating the loss of information that comes from moving data points to grid locations. In fact, optimal spacings of samples can be determined which permit the most precise estimation of a covariance function for a fixed number of measurements [7,11] (see below).

3.1. The Maximum Likelihood Method for Covariance Function Estimation

Consider a data set $\{a_i; i = 1, 2, \dots, N\}$ representing assays at a set of points, $S : \{\mathbf{x}_i; i = 1, 2, \dots, N\}$. If the underlying model for these data is a stationary Gaussian random function of mean μ and spatial covariance $C(\mathbf{h})$ corrupted by a Gaussian measurement variance σ_{a0}^2 where $\mathbf{h}$ is the (vector) distance between observations, then the probability of observing the data is

$$\Pr\{\mathbf{a}|\mathbf{p}, S\} = \frac{1}{(2\pi)^{N/2} \|\mathbf{R}(\mathbf{p}, S)\|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{a} - \mu)^T \mathbf{R}^{-1}(\mathbf{p}, S)(\mathbf{a} - \mu)\right] \quad (32)$$

where $\|\cdot\|$ denotes the determinant of the matrix and $\mathbf{R}$ is the covariance matrix of the data, which is taken to be a function of the parameters $\mathbf{p}$ of the covariance model for the data and the structure of the lattice of the data points, S . The measurement variance, or nugget, is included in the covariance matrix. The mean of the data, μ , must also be taken into the model parameter set to be estimated.

It is useful to separate the total variance of the mean, denoted by σ^2 , from the covariance by using the autocorrelation matrix $\mathbf{A}$:

$$\mathbf{A} = \frac{\mathbf{R}}{\sigma^2} \quad (33)$$

The probability is then

$$\Pr\{\mathbf{a}|\mathbf{p}, S\} = \frac{1}{(2\pi)^{N/2} \sigma^N \|\mathbf{A}(\mathbf{p}, S)\|^{1/2}} \exp\left[-\frac{(\mathbf{a} - \mu)^T \mathbf{A}^{-1}(\mathbf{p}, S)(\mathbf{a} - \mu)}{2\sigma^2}\right] \quad (34)$$

The standard approach to maximum likelihood (ML) estimation then permits the statement that the log-likelihood of the model parameters is

$$L(\mathbf{p}|S, \mathbf{a}) = -\frac{N}{2} \ln(2\pi\sigma^2) - \frac{(\mathbf{a} - \mu)^T \mathbf{A}^{-1}(\mathbf{p}, S)(\mathbf{a} - \mu)}{2\sigma^2} - \frac{1}{2} \ln \|\mathbf{A}(\mathbf{p}, S)\| \quad (35)$$

The influence of the parameters that describe the shape of the variogram and the spatial location of the points at which the data are determined is confined to the autocorrelation matrix $\mathbf{A}$.

Computationally, it is necessary to maximise the log-likelihood or minimise the negative of the log-likelihood. A generalised minimisation procedure can be used. In the computational scheme, if the current proposal for the autocorrelation function is $\tilde{\mathbf{A}}(\mathbf{p}, S)$ based on the proposed parameters shaping the autocorrelation function or normalised variogram, then the ML estimate of the data mean, conditional on the choice of the autocorrelation function, is

$$\hat{\mu}_{ML} = \frac{\sum_{j=1}^N \sum_{i=1}^N \alpha_{ij} (a_i + a_j)}{2 \sum_{j=1}^N \sum_{i=1}^N \alpha_{ij}} \quad (36)$$

where α_{ij} is the ij th element of the inverse of the matrix $\tilde{\mathbf{A}}$. Note that the ML estimate of the mean is asymptotically unbiased.

The variance of the mean (the sill) of the random function can also be estimated analytically as

$$\hat{\sigma}_{ML}^2 = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N (a_i - \hat{\mu}_{ML})(a_j - \hat{\mu}_{ML}) \alpha_{ij} \quad (37)$$

Having extracted these estimates of the mean and sill from the data, it is possible to continue with the estimation problem, which, after substituting the ML estimates for the mean and sill, becomes

$$-L(\mathbf{p}|\mathcal{S}, \mathbf{a}) = \frac{N}{2} \left[1 + \ln(2\pi) + \ln(\hat{\sigma}_{ML}^2) \right] + \frac{1}{2} \ln \|\mathbf{A}(\mathbf{p}, \mathcal{S})\| \quad (38)$$

This problem is reasonably well-posed but is likely to depend for its solution on a reasonable number of data values, well distributed over a region. So $\mathcal{S}$ and the number of observations can be expected to have an impact on the extent to which the solution is well-defined.

Note also that if the data follow a distribution other than the Gaussian, that density can be used in (32) instead, but the formulation may be substantially more complex.

From the computational point of view, it is clear that minimisation of the function is not a simple task. The calculation involves the determinant and the inverse of the covariance matrix. Since the inverse as well as the determinant are needed, and $\tilde{\mathbf{A}}$ is positive definite, the most efficient way of dealing with the problem is to make a Cholesky decomposition [12] of $\tilde{\mathbf{A}}$ as

$$\tilde{\mathbf{A}} = \mathbf{L}\mathbf{L}^T \quad (39)$$

where $\mathbf{L}$ is lower triangular. The determinant of $\tilde{\mathbf{A}}$ is then simply the square of the product of the diagonal elements of $\mathbf{L}$, and the inverse of $\tilde{\mathbf{A}}$ can be found from an extension of the usual Cholesky calculations. The inverse of $\tilde{\mathbf{A}}$ can be found directly from $\mathbf{L}^{-1}$. Even though the Cholesky method is more efficient than other methods, the matrix is $N \times N$ and the Cholesky inversion is an $O\{N^3\}$ process. This assessment of the computational work needed to find the ML estimate of the variogram makes it clear why the usual method of estimation of the variogram involves the Matheron estimator and least squares fitting instead of the ML estimate. At the time of the development of geostatistics, the computational effort required by the ML method would have put the method out of reach of geostatisticians. Now, unless the data set is bigger than $N = 1000$, the ML method using Cholesky inversion is practical. For the usual amount of data available in a process sampling context, the method has no disadvantage at all. The use of large data sets from on-line measurement may preclude the use of the ML method, unless the data are equally spaced and the set is augmented to make its covariance matrix circulant. In such a case, very large data sets can be handled and the circulant calculations are $O\{N \ln N\}$.

The effectiveness of the ML method can be demonstrated by a simple simulation example. A total of 200 random function realisations of 100 points with an exponential variogram with a nugget variance of 0.1 times the sill and decay constant of 20 were created and the Matheron estimator used to calculate variograms. The nugget variance, decay constant, and sill were then estimated by ML for the same random function realisations and the variogram calculated using these values. Then the averages of the variograms were

calculated for the two methods along with the standard deviations of the variogram values at each lag. The mean variograms with ± 1 SD error bars are shown in Figure 4. When the variogram form is fitted to the average Matheron variograms, the estimates of the nugget and decay constant were found to be 0.08578 and 17.91 and 0.09999 and 19.998 for the ML estimator. The ML estimates are essentially exact while the Matheron values are biased. In addition, the variation of the ML variograms with respect to their mean is smaller at the higher lags. While it can be shown [7,9] that the Matheron estimator is unbiased, as an estimator for a finite data set, it is less effective in that it converges to the true value more slowly than the ML estimator and has a higher variance. The unbiasedness and lower variance of the ML estimator is expected due to the properties of ML estimators [9].

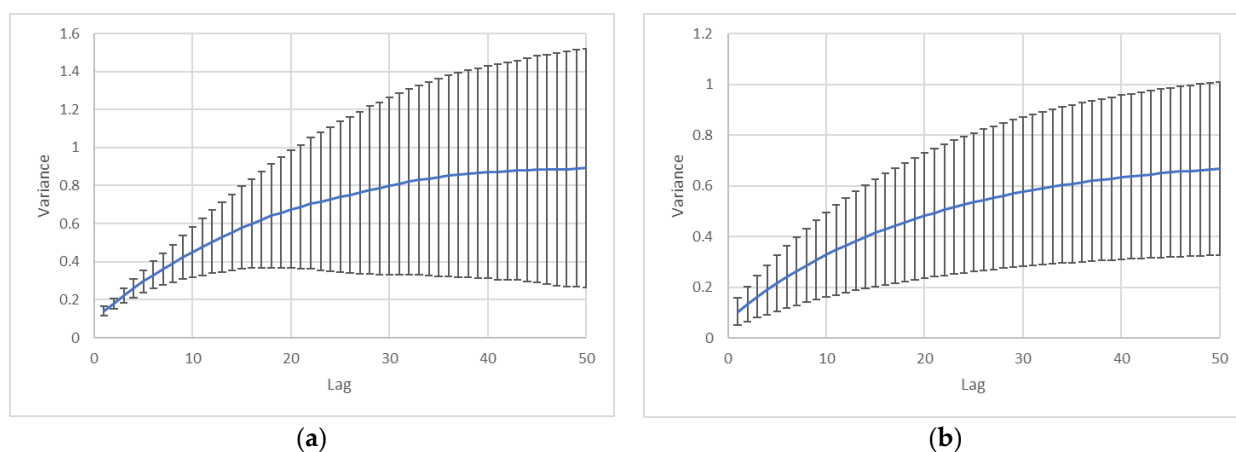


Figure 4. Average variograms with ± 1 SD error bars from 200 realisations of a random function with a nugget of 0.1 and decay constant of 20. (a): Matheron estimator. (b): ML estimation.

In the ML method, a matrix known as the Fisher information matrix is defined as below. In a ML parameter estimation problem, where the model for the data involves a set of parameters $\mathbf{p}$, the information matrix is given by

$$I_{ij} = -E \left\{ \frac{\partial^2 L}{\partial p_i \partial p_j} \right\} \quad (40)$$

where L is the logarithm of the likelihood function. The expectation operator $E\{\cdot\}$ is taken over many realisations of the data that follow the model that is being sought. The covariance matrix of the model parameters is the inverse of the information matrix. Having found the maximum of the likelihood function, the information matrix can be calculated numerically and inverted. Further, it can be shown [13] that the distribution of the parameter estimates is asymptotically jointly normal as the size of the data set increases with the variance derived from the information matrix.

The ML method has many virtues compared to least squares estimation applied to data whose distribution is not correctly recognised; the extra computational effort is worthwhile. The ML method avoids the bias involved in the estimated variogram, recognises the exact location of data that are unequally spaced, applies a mathematically sound parameter estimation procedure, and provides an estimate of the covariance matrix of the parameter estimates. It is far superior to making a fit by eye to data points whose uncertainty is not known, as is frequently current practice in geostatistical settings.

4. Data Acquisition for Process Analysis

In the sampling literature, there is not a great deal of data from processes and perhaps even less from conventional sampling plants. This suggests that industry is not making the effort to collect process data with the objective of determining their actual process sampling variance, or they do not want to publish data.

The variogram has great utility for process analysis as it reveals the dynamic behaviour of the process, and, if the data acquisition for its estimation is performed carefully, it can also provide an estimate of the total sampling variance that is achieved with a given sampling scheme and sample preparation protocol. The last fact derives from the fact that the component of uncertainty in a measurement, which is not correlated in time, tonnage, or space and characterised by the variogram, shows up on the variogram as the intercept of the variogram at zero separation of time difference (lag). The intercept is called the ‘nugget variance’ by geostatisticians. By definition, this intercept is the ‘variance of a measurement with itself’. This variance includes all the sample preparation and analysis variances as well as any variance associated with the primary increment due to particulate heterogeneity. Using the variogram with the nugget variance removed—the variance due to time-wise variation and the particular spacing between measurements—the sampling variance due to the time variation can be calculated.

Combining this with variance due to sample preparation and analysis, the total sampling variance can be determined as the sum of the three underlying uncertainties in sampling (distributional variance, particulate sampling variance, and analytical variance)

4.1. Sources of Plant Data

The collection of data for sampling analysis can come from three basic sources. First, a stream (such as a mineral processing plant product) can use a physical sampler installed on the stream to produce a set of increments taken at particular times during a day, a shift, or part of a shift and composited prior to analysis. Second, the data set may consist of measurement of each individual increment taken by a physical sampler. Lastly, the data may come from an on-line analyser generating new readings at defined intervals.

In the first case, with increments composited over a significant timespan, the variogram calculated from the measurements will not be what can be called the ‘point variogram’, which gives the full picture of process variation; the point variogram is the ideal source of information on process variation. A variogram based on composite samples results in an averaging or integration of the process variations. It can be calculated from the point variogram, but the process can only be approximately reversed, finding the point variogram from the variogram based on composite samples. The latter variogram will show a lower total variation (variogram sill) and will be what is called a ‘regularisation’ of the point variogram. Consequently, if process dynamics indicate that the point variogram should have a particular form, the variogram from the composite samples will have a different form. However, the intercept of the variogram for the composite samples will still provide a good estimate of the sampling and preparation variance, but that variogram cannot be used to determine total sampling variance over the period of the compositing as the point variogram is required for this purpose. It is nonetheless a very useful starting point for process analysis and can be useful in designing a special sampling program for the estimation of the point variogram.

Physical sampling to establish the point variogram involves extracting correct samples from a process stream at a chosen set of times. The recommendations in the current literature indicate that increments should be extracted at a set of times that fall on a regular grid. With some increment spacings that are close together and others that are multiples of the closer spacing, an estimate of the variogram can then be made to cover a relatively wider range of intervals. However, this requirement to use a grid is outmoded by the fact that correct variogram estimation can be made without equally spaced points. More information can be obtained for less cost by using a set of sampling intervals following a particular distribution. Such a strategy can provide a much better estimate of the variogram nugget variance if a mixture of closely and wider-spaced sampling times are used. The advantage of the statistically spaced points is that information on a full range of frequencies is collected. More information close to the origin of the variogram is collected, better defining the intercept of the variogram.

The author carried out a study [11] to estimate the distribution of sample spacings that would provide the most information about the variogram with a fixed number of samples. The study was aimed at sampling for diamonds using a large core, where the cost per sample is very large, but the result is clearly useful in any circumstance. The method was based on maximising the Fisher information for the maximum likelihood estimation of the covariance function parameters. The covariance matrix of the parameters is the inverse of the Fisher information matrix, so the maximisation minimised the determinant of the covariance matrix and, thus, the uncertainty in parameter estimates. The computations were based on multiple 400-point realisations of a random function on a regular spacing, and the data used in a given parameter estimation were selected by sampling spacings of the samples from a spacing distribution function. The number of points used for estimation was fixed to a fraction of the 400 equally spaced points where the random function had been calculated. For each distribution of spacings, the average value of the Fisher information matrix over the fits to the 30 realisations was calculated and optimization of the spacing distribution was achieved using the method of simulated annealing [12]. The variogram function studied was an exponential variogram with a non-zero nugget variance.

The evolution of the spacing distribution, which started with a uniform random distribution to a final distribution, is shown in Figure 5. A total of 50 sample points were used. The use of an optimized unequal sample spacing can reduce the cost of the variogram acquisition and maximise the information on the variogram.

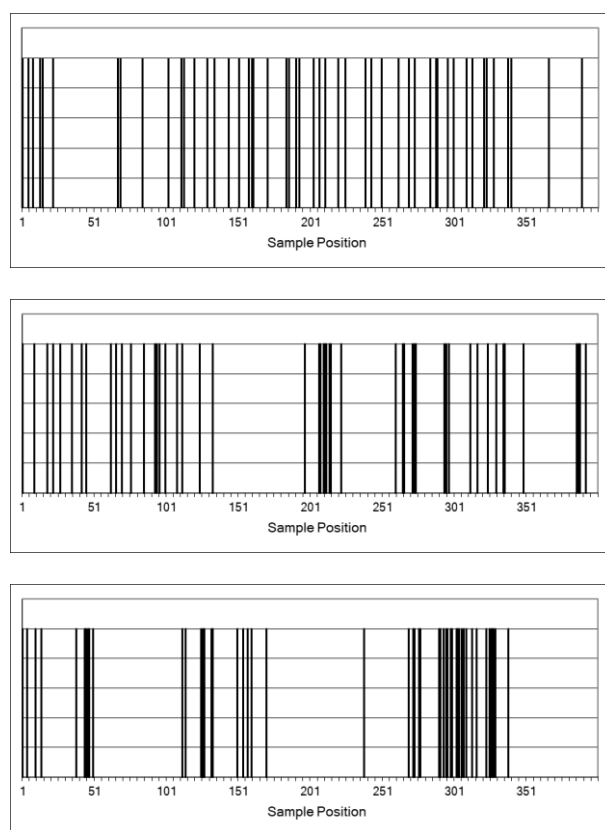


Figure 5. Evolution of sample spacing distributions towards an optimal spacing.

The presence of an on-line analysis unit such as a prompt gamma neutron activation gauge (PGNA) on the process stream opens up the possibility of constant monitoring of the process stream characteristics. The gauge typically produces a reading every few minutes. It is important to know how the measurements have been calculated. For some gauge models, the value produced every minute is, in fact, a moving average of some set of past raw measurements. For the AllScan produced by RealTime Instruments (RTI), the raw

readings can also be subjected to a statistical procedure that updates the gauge output only when there has been a significant change in the material composition. Such a procedure does not distort the information and is effective in removing noise from the gauge output.

The ability to add real-time estimation of the variogram or covariance function for a process reading is something that is not exploited in industry even though there is no computational barrier to doing so. The estimation of the variogram requires a set of historical data—at least 30 immediate past readings—and therefore its response to process changes will not be extremely rapid. Signalling of a change in a process variable may be better indicated using a cumulative sum statistic. However, a simple change in a process variable does not have the ability, as a variogram does, to detect changes in process dynamics. The variogram can also indicate a change in the precision of measurement of the on-line gauge from the change in the nugget variance of the variogram. Using maximum likelihood estimation of the variogram, the precision of the estimates of the variogram parameters is also accessible.

An on-line analyser providing measurements on a minute-by-minute basis captures all process dynamics in a typical mineral processing plant. The same is true when a special sampling campaign has been carried out with the objective of capturing the full variogram for the process stream. When historical data are the only source available and they have come from a physical sampling system, the sample grades are the result of extraction of increments on a regular time basis or a random stratified basis, and the sample is a composite of the increments taken over the sampling period. As a result of the compositing, there will be a loss of information. The current literature does not address this issue in any form, but it is possible to analyse this case of data collection and understand how the variogram calculated from analysis of individual increments is related to the variogram based on composite samples.

4.2. The Relationship between Point Variograms and Variograms of Composite Samples

In this analysis, the nugget variance must be removed from the variogram as this component of variance is independent from the covariance with time and relates only to the variance introduced by sample preparation and analysis.

An equation is now derived relating the point covariance/variogram function to those of the composite samples.

Let the point analyte variation be given by the discrete random function $X(t_i)$; $i = 1, 2, \dots$ and assume that increments are collected into the composite sample every Δt time units so they have analyte contents $X(i\Delta t)$; $i = 1, 2, \dots, N$ with N increments in each composite sample. The analyte contents of the composite samples are

$$Y(j) = \frac{1}{N} \sum_{i=1}^N X[i\Delta t + (j-1)N\Delta t] ; j = 1, 2, \dots \quad (41)$$

so the covariance function for the composites must be

$$C_Y(h) = E\{(Y(j) - Y_0)(Y(j+h) - Y_0)\} \quad (42)$$

Writing this in terms of the function X ,

$$\begin{aligned} C_Y(h) &= E\left\{ \frac{1}{N^2} \left[\sum_{i=1}^N X[i\Delta t + (j-1)N\Delta t] - X_0 \right] \left[\sum_{k=1}^N X[k\Delta t + (j+h-1)N\Delta t] - X_0 \right] \right\} \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{k=1}^N C_X[|i\Delta t + (j-1)N\Delta t - k\Delta t - (j+h-1)N\Delta t|] \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{k=1}^N C_X[|hN\Delta t + (k-i)\Delta t|] \end{aligned} \quad (43)$$

This summation is best understood by explicitly evaluating some of the terms. Consider the case of $N = 4$, then for $h = 0$

$$C_Y(0) = \frac{1}{N^2} \sum_{i=1}^N \sum_{k=1}^N C_X[|(k-i)\Delta t|] \quad (44)$$

and there are N^2 terms involved which can be structured as the average of the entries in a matrix (note that the absolute value of the difference is used since the covariance is an even function):

$$\begin{bmatrix} C_X(0) & C_X(\Delta t) & C_X(2\Delta t) & C_X(3\Delta t) \\ C_X(\Delta t) & C_X(0) & C_X(\Delta t) & C_X(2\Delta t) \\ C_X(2\Delta t) & C_X(\Delta t) & C_X(0) & C_X(\Delta t) \\ C_X(3\Delta t) & C_X(2\Delta t) & C_X(\Delta t) & C_X(0) \end{bmatrix} \quad (45)$$

This term is, in fact, the sill of the variogram for the composite sample variogram, and since the covariance function is a decreasing function, this sill must be lower than the sill for the variogram of $X(t)$. For $h = 1$, the term is the average of the elements of the matrix

$$\begin{bmatrix} C_X(4\Delta t) & C_X(5\Delta t) & C_X(6\Delta t) & C_X(7\Delta t) \\ C_X(3\Delta t) & C_X(4\Delta t) & C_X(5\Delta t) & C_X(6\Delta t) \\ C_X(2\Delta t) & C_X(3\Delta t) & C_X(4\Delta t) & C_X(5\Delta t) \\ C_X(\Delta t) & C_X(2\Delta t) & C_X(3\Delta t) & C_X(4\Delta t) \end{bmatrix} \quad (46)$$

and so on for higher values of h .

The averaging process involved in the matrix of Equation (44) deserves further consideration. The average of the values in the matrix can be written as (omitting the Δt for simplicity)

$$C_Y(h) = \frac{1}{N^2} \left[NC_X(h) + (N-1)(C_X(h+1) + C_X(h-1)) + (N-2)(C_X(h+2) + C_X(h-2)) + \dots + C_X(h+N-1) + C_X(h-N+1) \right] \quad (47)$$

If the point covariance is almost linear around the value of $C_X(h)$, then

$$C_X(h+j) + C_X(h-j) \approx 2C_X(h) \quad (48)$$

and

$$\begin{aligned} C_Y(h) &= \frac{1}{N^2} [NC_X(h) + 2(N-1)C_X(h) + 2(N-2)C_X(h) + \dots + 2C_X(h)] \\ &= \frac{C_X(h)}{N^2} N^2 = C_X(h) \end{aligned} \quad (49)$$

It is not until the slope of the covariance function changes rapidly as the range of the covariance function is reached that the above approximation fails. The conclusion is that, while the first point on the composite covariance function differs from that of the point covariance function in a known manner, the composite covariance function tracks the point covariance function closely until the covariance approaches zero and remains there. This approximation will be valid until the increments taken into the composite cover a significant portion of the range of the point variogram. In that case, the dynamics of the process flow will be significantly obscured in the variogram from the composite samples but may be recoverable, but perhaps not uniquely, using the relationships derived here.

Figure 6 shows a point covariance function and the corresponding composite covariance function, where 10 increments are taken into the composite as well as the resulting variograms. The critical point is that while the sill of the variogram changes, the shape of the variogram is preserved so that the dynamics of the process are not obscured. This aspect of variogram/covariance function behaviour has not been addressed by the current sampling literature, even though it is critical to the correct determination of sampling variance in sampling system design when experimental data are available.

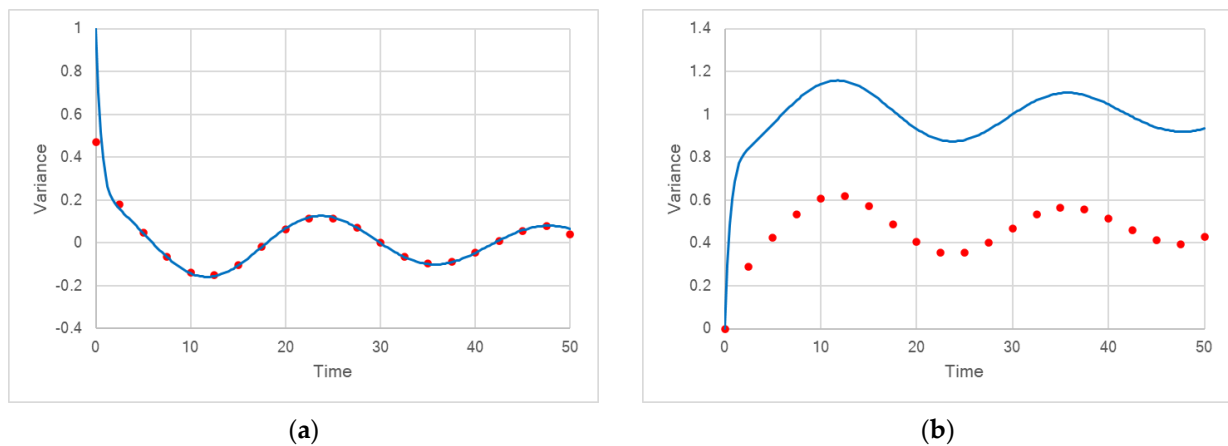


Figure 6. (a): point covariance function and composite covariance function for a periodic variogram. (b): corresponding variograms. Solid line: point function; dots: composite function.

5. Examples of Variogram Analysis

An interesting example of the use of process dynamic knowledge to determine a variogram or covariance function for a part of a process is the modelling of the operation of a flotation cell as a combination of a stirred tank and a plug flow device. The arrangement is shown in Figure 7.

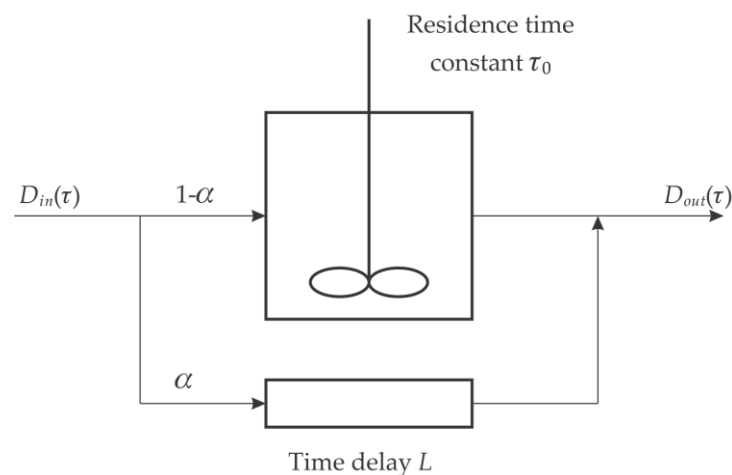


Figure 7. Combination of a stirred tank and plug flow device.

The output of a system with a stochastic input can be determined using the residence time distribution for the system. For the tank and delay model, the residence time distribution is known to be

$$RTD(\tau) = \frac{1-\alpha}{\tau_0} e^{-\tau/\tau_0} + \alpha \delta(\tau - L) \quad (50)$$

where $\delta(\cdot)$ is a Dirac or impulse distribution which takes on the value of unity when the argument is zero and zero otherwise and α is the fraction of the total flow through the plug flow device. L is the time lag across the plug flow device and τ_0 is the time constant for the stirred tank, so the output of the system is

$$\begin{aligned}
D_{out}(t) - d_0 &= \int_{\tau=0}^{\infty} RTD(\tau) [D_{in}(t - \tau) - d_0] d\tau \\
&= \int_{\tau=0}^{\infty} [D_{in}(t - \tau) - d_0] \left[\frac{1-\alpha}{\tau_0} e^{-\tau/\tau_0} + \alpha \delta(\tau - L) \right] d\tau \\
&= \frac{1-\alpha}{\tau_0} \int_{\tau=0}^{\infty} [D_{in}(t - \tau) - d_0] e^{-\tau/\tau_0} d\tau + \alpha [D_{in}(t - L) - d_0]
\end{aligned} \quad (51)$$

The output covariance function is then defined as

$$C_{out}(x) = E\{[D_{out}(t) - d_0][D_{out}(t + x) - d_0]\} \quad (52)$$

where d_0 is the mean concentration and $E\{\cdot\}$ denotes an expectation. When the deviation from the mean input concentration is white noise, $Z(t)$, the covariance becomes

$$C_{out}(x) = E\left\{ \left[(1-\alpha) \int_{\tau=0}^{\infty} e^{-\tau/\tau_0} Z(t-\tau) \frac{d\tau}{\tau_0} + \alpha Z(t-L) \right] \left[(1-\alpha) \int_{s=0}^{\infty} e^{-s/\tau_0} Z(t-s+x) \frac{ds}{\tau_0} + \alpha Z(t+x-L) \right] \right\} \quad (53)$$

Multiplying out the terms and taking the expectation over the white noise function,

$$C_{out}(x) = (1-\alpha)^2 \frac{e^{-x/\tau_0}}{2\tau_0} + \alpha(1-\alpha) \frac{e^{-(L-x)/\tau_0}}{\tau_0} + \alpha(1-\alpha) \frac{e^{-(L+x)/\tau_0}}{\tau_0} + \alpha^2 \delta(-x) \quad (54)$$

where the second term is as written for $x \leq L$ and zero otherwise. Noting that

$$C_{out}(0) = \frac{(1-\alpha)^2}{2\tau_0} + 2\frac{\alpha(1-\alpha)}{\tau_0} e^{-L/\tau_0} + \alpha^2 \quad (55)$$

the variogram can be written as

$$\begin{aligned}
\gamma(x) &= C_{out}(0) - C_{out}(x) \\
&= \frac{(1-\alpha)^2}{2\tau_0} (1 - e^{-x/\tau_0}) + \frac{\alpha(1-\alpha)}{\tau_0} (e^{-(L-x)/\tau_0} - e^{-L/\tau_0}) + \frac{\alpha(1-\alpha)}{\tau_0} (e^{-(L+x)/\tau_0} - e^{-L/\tau_0}) + \alpha^2 (\delta(-x) - 1)
\end{aligned} \quad (56)$$

with the condition noted above on the second term.

Figure 8 shows this variogram and a corresponding variogram from a discrete simulation of the system using a white noise input and no measurement error. The scales do not match as the discrete simulation cannot exactly match the (continuous) theoretical function. This example demonstrates that ‘blind’ fitting of variogram functions to process data is a poor practice and that process knowledge should be used whenever possible. Process covariance/variogram functions can differ significantly from those used by geostatisticians.

The following analysis of a data set from an intermediate stage of recovery of a precious metal in a base metal smelter provides an excellent insight into the utility and importance of data detrending prior to calculation of variograms. Figure 9 presents data detrending (red line) and the corresponding variogram (Matheron estimator) for a series of detrending levels. The window width varies from 100% of the data set (more than 900 values from historical assays) down to 10% of the data set. As the detrending level increases (smaller windows), the lower-frequency components of the data variation are progressively removed, and higher-frequency components become more apparent.

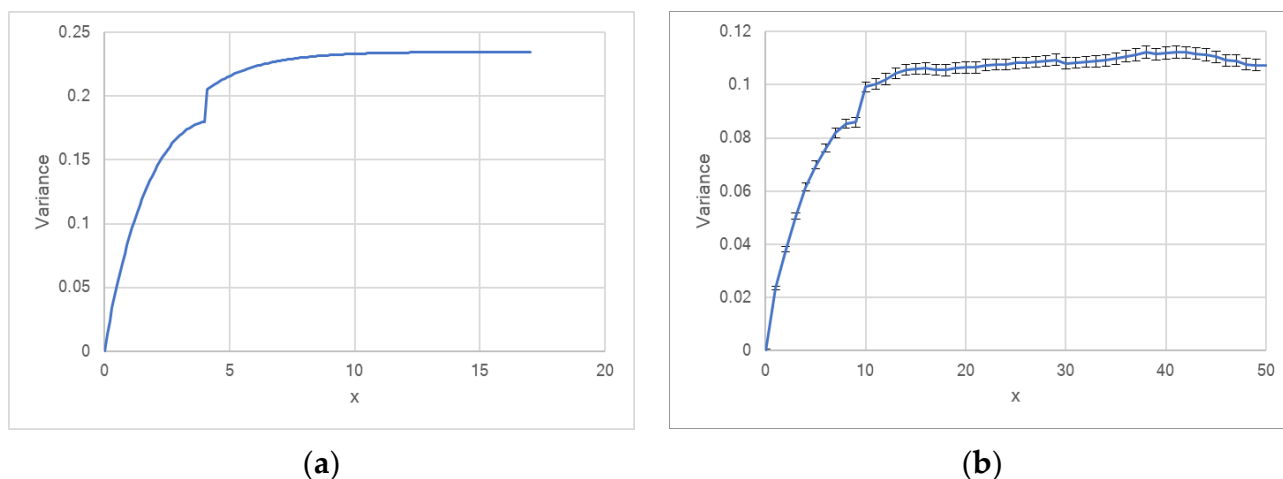


Figure 8. Comparison of theoretical variogram (a) and simulated variogram (b) for a stirred tank/plug flow device system. The split to the plug flow device is $\alpha = 0.05$ and the time delay is twice the residence time in the tank.

Using a 100% window for the detrending minimises the extent to which the trend follows the data. The corresponding variogram picks up the large periodic variation in the data with a period of about 500 lags. This low-frequency variation dominates the variogram. The small error bars (± 1 SD) indicate that the variogram is very well-defined. With a 70% window, the low-frequency periodicity is reduced and a significant higher-frequency variation with a period just under 100 lags becomes apparent. That this variation is significant is indicated by the fact that the variation does not fall inside the error bars. At 40% and 20% windows, this higher-frequency periodicity is even better defined. However, with the 10% window, the periodicity at 100 lags has been almost completely filtered out. It might be tempting to identify an even higher-frequency variation at this level of filtering, but the magnitude of the error bars suggests that this variation is not statistically significant. If the error bars had been drawn at ± 2 SD, the variations in the variogram would be fully contained within the error bars and the conclusion would be that the remaining variation is just noise.

A second example of a process analysis using variograms comes from a flotation plant with data from an on-line analysis system producing readings every 15 min for a rougher concentrate. Figure 10 presents the raw data with the detrended data and the raw and fitted variogram. The variogram has been normalised to unit sill. The raw sill is 2.34 from the detrended data and the average grade is 6.13. The variogram shows a 24 h periodic signal. The fitted function is a combination of an exponential function (mixing) and a damped periodic function. The equation is

$$\gamma(h) = f_n + (1 - f_n) \left[\alpha \left(1 - e^{-h/\delta_1} \right) + (1 - \alpha) \left(1 - e^{-h/\delta_2} \cos \left(\frac{2\pi h}{P} \right) \right) \right] \quad (57)$$

The following parameter values were found by weighted least squares fitting of the variogram to the normalised variogram: $f_n = 0.0838$, $\alpha = 0.805$, $\delta_1 = 2.21$, $\delta_2 = 214$ when the time variable is in units of 15 min. The period was fixed at 24 h. Using the average grade and the nugget variance of the normalised variogram and the sill variance, the relative standard deviation of a single measurement is 7.2%. Since the time variable h is in units of 15 min, the mixing term suggests a residence time in the bank of 2.21 15 min steps, or 33 min, which is quite reasonable for a rougher bank. The strong diurnal correlation suggests that some daily procedure in operating practice is affecting the plant feed.

The data set is far too large to permit maximum likelihood fitting of the model to the variogram without the circulant embedding method, so it was carried out by weighted least squares.

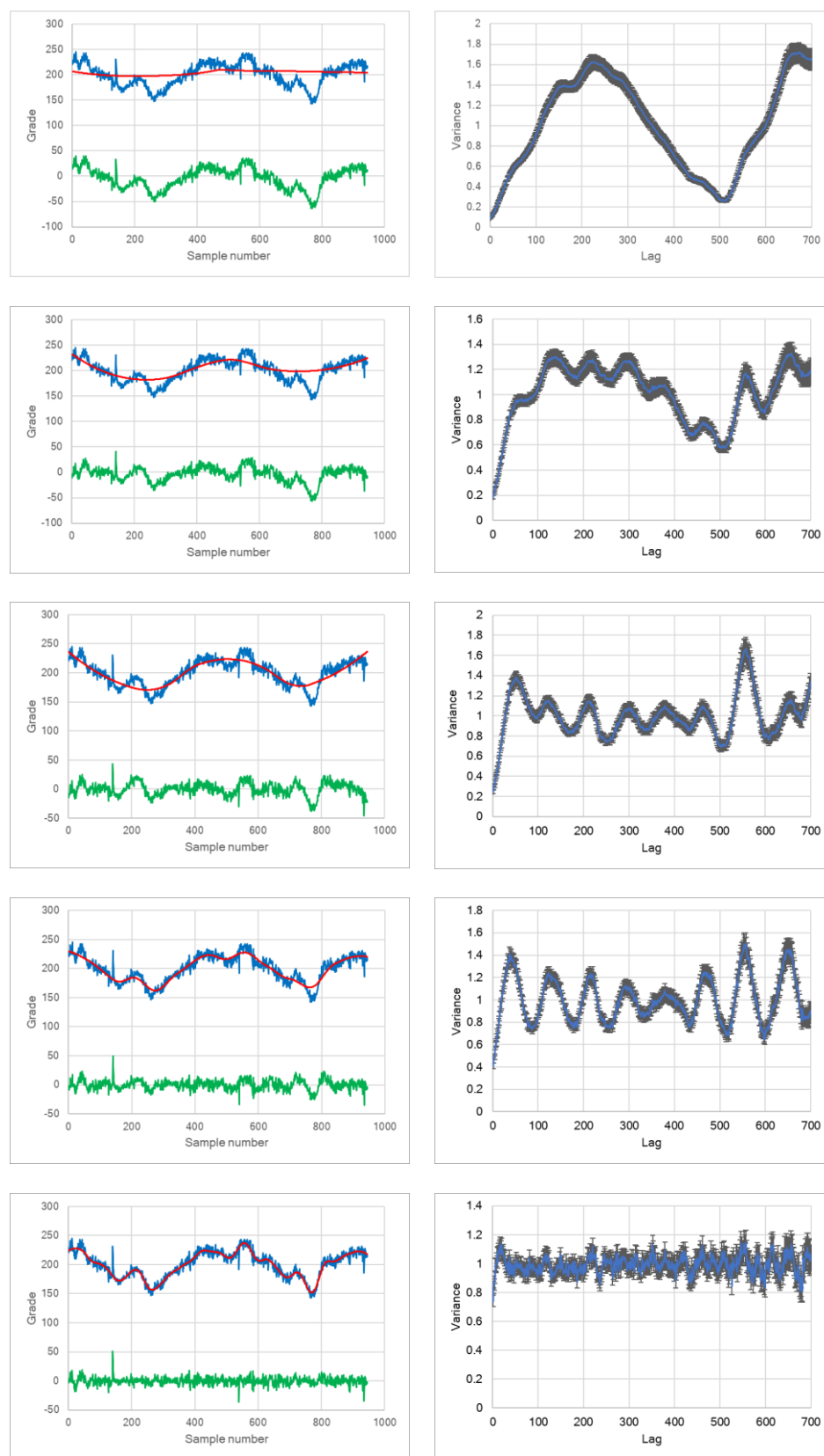


Figure 9. Precious metal grade variation with various degrees of detrending of the data, with corresponding detrended data (green) and variograms. Window widths for detrending, top to bottom, 100%, 70%, 40%, 20%, 10%.

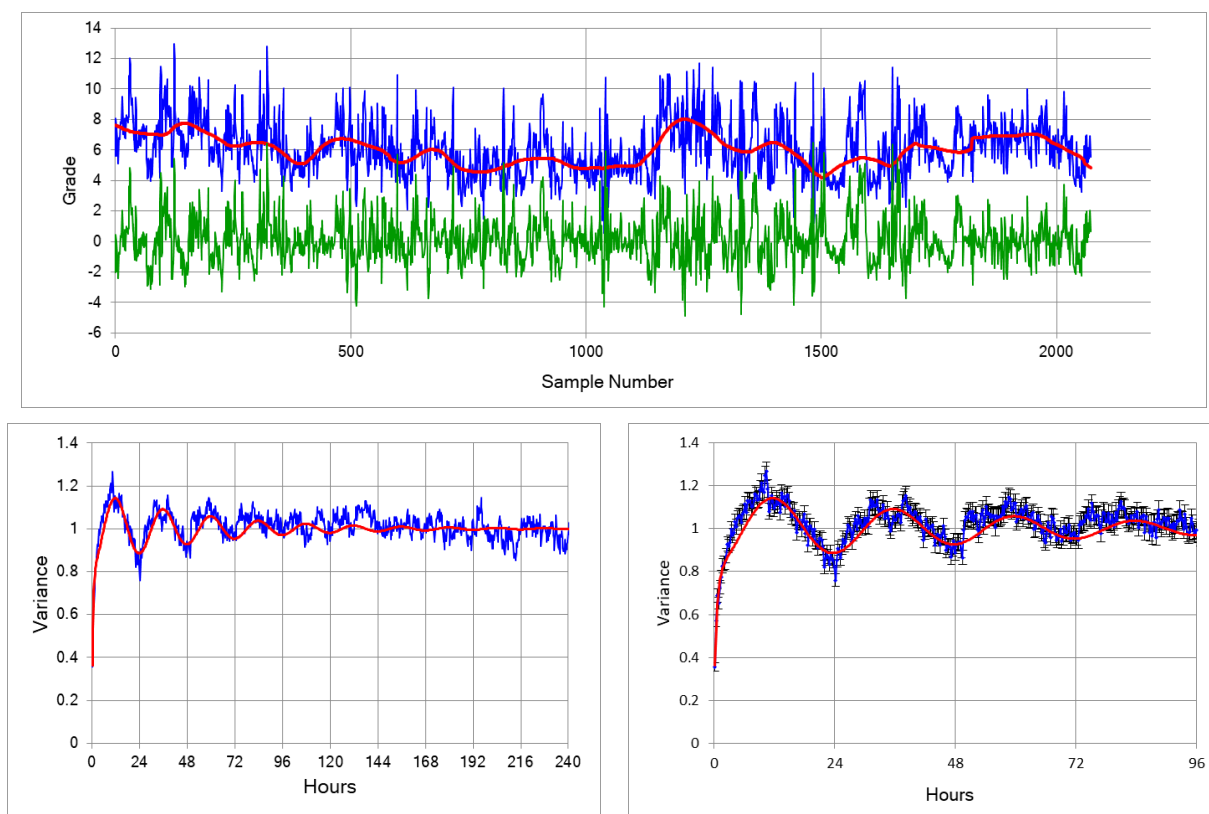


Figure 10. Raw and detrended data for a flotation concentrate from an on-line analysis system (top) with the corresponding variogram. The bottom plot shows the full variogram and the part of the variogram with up to 4 days of logging, with the variogram with error bars of 1 SD and fitted variogram function.

In a plant with on-line analysis, where all streams around a processing unit are measured, it is possible to make a statistical material balance around the unit using the assays. Using measurement precisions derived from variographic analysis of the grade variations, it is possible to calculate metal recoveries on-line and attach an estimate of the uncertainty in the recovery. There is substantial value in such information; the cost of the information is low to acquire, given it usually derives from better use of existing data.

6. Process Variance Estimates and Total Sampling Variance

A statistically sound method for making process variance estimates of a given sampling/measurement scheme is to start with the point variogram/covariance for the stream and fit a model to the experimental data, ideally by maximum likelihood but secondarily by weighted least squares fitting of the Matheron variogram, if the inherent uncertainties can be deemed acceptable and the data set is large. Under no circumstance should the raw variogram be used for sampling variance estimation as the raw variogram is not an admissible function. Following a good representation of the variogram with an admissible function, its covariance matrix is calculated at an appropriate time step and the process variance determined from the method leading to (29). The total sampling variance is then found by adding to this the sample preparation and analysis variances.

Caution should be exercised to check for significantly non-normal grade distributions. When the grade variation about the mean is not normally distributed, the fact that a substantial number of increments have been collected to make up the composite sample will usually cause the distribution of the composite sample (true) assay to be normally distributed. Moreover, the uncertainty in the final assay may not be normally distributed when the sample preparation and analysis variance is large. In such cases, the final

distribution of the total sampling variance can be found by combining the component due to process variance and that due to preparation and analysis by the characteristic function method [1,7], since the two variance components are statistically independent.

7. Summary

The analysis presented herein provides a mathematical basis for the statistical analysis of process flows by characterizing the process dynamics with the covariance function or variogram. The methods here improve and substantially expand on those provided by the current sampling literature. For estimation of the covariance function, the optimal method of maximum likelihood is described in detail. The use of process dynamical models as a basis for covariance function estimation is introduced and shown to lead to covariance function structures that differ from those used by geostatisticians. The relationship between the covariance function for samples that are composites of individual increments collected from a process stream and the point covariance function is derived with an example of the difference between the two. The work is supported by several examples based either on simulations or plant data.

Funding: This research received no external funding.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Lyman, G.J. A Statistical Theory for Sampling of Particulate Materials. *Minerals* **2023**, *13*, 905. [CrossRef]
2. Journel, A.G.; Huijbregts, C.J. *Mining Geostatistics*; Academic Press: Cambridge, MA, USA, 1978.
3. Chiles, J.-P.; Delfiner, P. *Geostatistics, Modeling Spatial Uncertainty*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2012.
4. Cressie, N.A.C. *Statistics for Spatial Data*; Wiley: Hoboken, NJ, USA, 1991.
5. Gy, P.M. *Sampling of Particulate Materials, Theory and Practice*; Elsevier: Amsterdam, The Netherlands, 1982.
6. Pitard, F.F. *Theory of Sampling and Sampling Practice*, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2019.
7. Lyman, G.J. *Theory and Practice of Particulate Sampling, an Engineering Approach*; Materials Sampling & Consulting: Southport, Australia, 2019. Available online: www.materials-sampling-and-consulting.com (accessed on 2 July 2023).
8. Lyman, G.J. The Impact of the Weighting Error. In Proceedings of the Eighth World Conference on Sampling and Blending, Perth, Australia, 9–11 May 2017; AusIMM: Melbourne, Australia, 2017; pp. 311–318.
9. Lyman, G.J. Variograms: Properties and Estimation. In Proceedings of the Sixth World Conference on Sampling and Blending, Lima, Peru, 19–22 November 2013; Gecamin: Santiago, Chile, 2013; pp. 185–206.
10. Cleveland, W.S. Robust Locally Weighted Regression and Smoothing Scatterplots. *J. Am. Stat. Assoc.* **1979**, *74*, 829–836. [CrossRef]
11. Lyman, G.J. Optimal Sampling for Variogram Estimation. In Proceedings of the Third World Conference on Sampling and Blending, Porto Alegre, Brazil, 23–25 October 2007; Fundacao Luiz Englert: Porto Alegre, Brazil, 2007.
12. Press, W.H.; Teukolsky, S.A.; Vetterling, W.T.; Flannery, B.P. *Numerical Recipes, the Art of Scientific Computing*, 2nd ed.; Cambridge UP: Cambridge, UK, 1992.
13. Priestley, M.B. *Spectral Analysis and Time Series*; Academic Press: Cambridge, MA, USA, 1981.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.