



# Application of variography to the control of species in material process streams: %Fe in an iron ore product

by R.C.A. Minnitt\* and F.F. Pitard†

## Synopsis

The use of a time series variogram as an aid in controlling the quality of a product was pioneered by Pierre M. Gy and is described. The ability to analyse and interpret covariance relationships time series data is possible using the variogram. The concept of stream heterogeneity allows sources of variability in process streams to be identified and isolated. Principally this is distributional heterogeneity that is due to a short range effect of segregation in the process stream related to size or composition distribution of the material on a local scale and due to a long-range change in the size and grade of the material stream. The variogram is simply a tool for representing these sources of variability in a way that is explicit, allowing identification of the source and providing important insights to temporal continuity between samples. This paper describes the chronologically related large-scale variability in an iron ore process stream (%Fe), and illustrates the control and mitigation of variability at various stages within the process using a variogram. Three main sources of variability are identified including the random variability  $V[0]$ , the process variability  $V[1]$ , and the variability due to cyclical phenomena  $V[\text{cyclic}]$ , arising from human routine or mechanical defects. These sources of variability can be represented on the control chart that provides the plant superintendent with insights as to the sampling capability of the systems in place. Maintaining levels of variability within customer specification limits requires that production systems be controlled without overcorrecting for the trends in changing quality. This can be done only by representing the limits of control relative to the limits of specification on a control chart. Where the specification limits encroach on the control limits, one or the other, or both, may need to be changed; specification limits that constrain the control limits lead to reactive decision making that is expensive and stressful for plant operators.

## Introduction

Applications of variography to understanding large-scale variability in process plants, process monitoring and process control were pioneered by Pierre M. Gy, a metallurgist, in the early 1950s (Pitard, 1993, 2007). Correlations between samples in any time series or product stream are known to exist and the ability to analyse and interpret such relationships is vastly improved through the use of the variogram. For the purist the name is 'semi-variogram', but common parlance has led the change to the term variogram. Gy

(1996) suggested the term 'chronostatistics' for the analysis of time series data, a term which Pitard (2007) uses to describe this 'powerful, pragmatic, new science for metallurgists' (Pitard, 2007). Poor understanding of the powerful analytical capabilities of the variogram means that it has not been widely applied in process control. According to Pitard (2006), conventional statistics and statistical process control (SPC) fail to address the concept of stream heterogeneity, and therefore fail to identify and distinguish between the various sources of variability in a process stream. Because of the range in types of heterogeneity, there is a matching range in variety of process variability, but principally this is a form of distributional heterogeneity that is due to segregation in the process related to size or composition distribution of the material on a local scale (Lyman, 2007). The variogram is simply a tool for representing the sources of variability in a way that is explicit, allowing identification of the source and providing important insights to temporal continuity between samples. The variability in processes is both dynamic and complex with the likelihood of stable and simple processes that are easy to manage being a rarity. In terms of fluctuations in product quality, Pitard (1993, 2007) identified long-range random fluctuations, short-range random fluctuations, and periodic or cyclical variations.

The aim of this paper is to provide a description of chronologically related large-scale variability in sampling systems and chemical process streams, and to illustrate the control and mitigation of variability at various

\* School of Mining Engineering, University of the Witwatersrand.

† Francis Pitard Consulting Services, Colorado, USA.

© The Southern African Institute of Mining and Metallurgy, 2008. SA ISSN 0038-223X/3.00 + 0.00. Paper received Aug. 2007; revised paper received Nov. 2007.

# Application of variography to the control of species in material process streams

stages within the process using a variogram. Such methods have been investigated by Holmes (1986), who compared the variogram and ISO methods of assessing the sampling variance for iron ores. He concluded that the two methods gave comparable results if the slope of the variogram is relatively small. When the slope is significant the ISO method considerably overestimates the sampling error. In the specific example used here, an iron ore product between -8 mm and +5 mm is collected as an incremental sample every four hours and has a lower specification (LS) limit of 65 %Fe; no upper specification (US) limit or target average (TA) is defined. The average grade of the iron ore product is 66.1 %Fe. Spangenberg (2001) also used variography to investigate the run-of-mine go-belt sampling precision at two AngloGold mines, and to estimate the appropriate sampling frequency at the mines. Analyses of 30 samples taken at 1-hour intervals from the go-belt at each mine provided a stable variogram, but very little interpretation of the results was undertaken.

Lyman (2007) suggests that any process stream may demonstrate a lack of 'mixedness', or distributional heterogeneity, that can manifest itself, firstly, as a short-range effect of segregation in the material stream, and secondly as a long-range change in the size and grade of the material stream. The segregation heterogeneity is usually small compared to the heterogeneity arising from changes in the size and grade of material in the process over time. Furthermore, because the sampling variance due to segregation is inversely proportional to the number of increments, the latter can be minimized by taking as many increments as possible.

In order to maintain levels of variability within customer prescribed specification limits, it is essential to control production systems without overcorrecting for the trends in changing quality. This can be done only by representing the limits of control relative to the limits of specification on a control chart. Where the specification limits encroach on the control limits, one or the other, or both, may need to be changed; specification limits that constrain the control limits lead to reactive decision making that is expensive and stressful for plant operators.

The procedure used in the analysis of a material stream consists of a number of steps:

- Firstly, an analysis of the data in a control chart or variability diagram that shows the size and intensity of variations in the item of interest in the stream of material being considered. This diagram should show the mean, the target average, and the specification limits (upper and/or lower)
- Secondly, an analysis of the data using moving average points to identify the main cycles within the material stream
- Thirdly, an analysis of the data using an absolute variogram in order to identify the components of variability, including the random variability  $V[0]$ , the process variation  $V[1]$ , and the cyclical variation  $V[\text{cyclic}]$ , and
- Fourthly, the transfer of the variability information to the control chart showing the upper (UCL) and/or lower control limits (LCL) as determined from the

components of variability  $V[0]$ ,  $V[1]$ , and  $V[\text{cyclic}]$ . This allows the capability of the process to deliver to the specification limits and target average to be determined. It also provides an indication of the constraining factors in the process stream and allows one to identify the areas or items within the plant that are in need of attention.

## Time series data analysis

Descriptive statistics and a histogram for 1 511 analyses of an iron ore product (%Fe) collected as stratified samples in a continuous stream at 4-hour intervals are shown in Table I and Figure 1, respectively. These data have also been analysed using:

- Moving average analysis, including the moving average plot, the corrected data for the moving average and the control chart for each ore type, and
- Variographic analysis, including the absolute and Relative variograms, the control chart, and the pie diagram showing the relative contribution of each type of error in the variogram.

## Descriptive statistics

Descriptive statistics for the iron content of the iron ore product shown in Table I assume that the process stream has an average grade and size distribution. The product has a mean of 66.10%Fe and a standard deviation of 0.68%Fe.

The histogram of iron ore grades (%Fe in Figure 1) indicates a slightly negatively skewed distribution that is typical for ores of ferrous metals. Although statistics are important at a global level, they provide no information about how to make meaningful process control decisions. In addition, it is noteworthy that the average grade of the iron ore (66.10%Fe) is 1.1% higher than the lower specification (LS) limit for the product at 65%Fe. Note that in this particular case for the iron ore product there is no upper specification limit.

## Variability control chart

Plots of random variables in moving material streams as a function of time, sometimes referred to as control charts, may indicate the presence of either random or periodic fluctuations

Table I

### Descriptive statistics of %Fe in an iron ore product

Statistic	Value
Mean	66.10
Standard error	0.02
Median	66.15
Mode	65.93
Standard deviation	0.68
Sample variance	0.46
Kurtosis	2.40
Skewness	-0.81
Range	7.22
Minimum	61.18
Maximum	68.40

# Application of variography to the control of species in material process streams

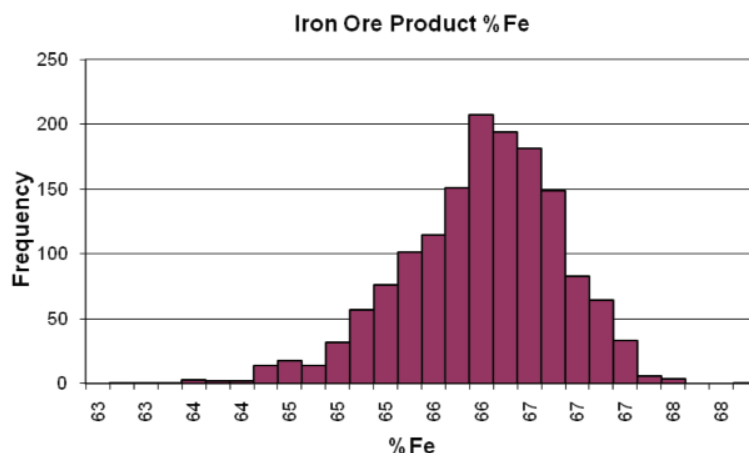


Figure 1—Histogram of 1 511 analyses of %Fe in an iron ore product with a mean of 66.10%Fe and a standard deviation of 0.68%Fe

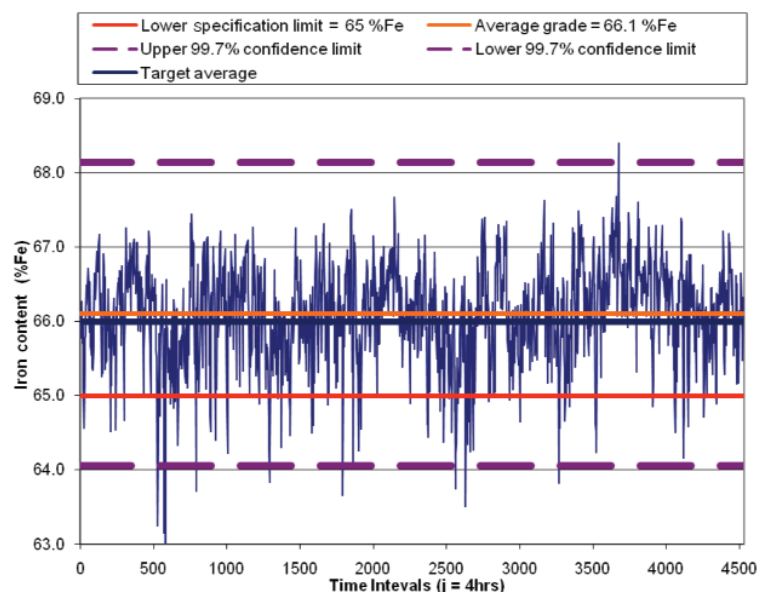


Figure 2—Control chart showing variability in iron ore grades (%Fe) with time. Sample interval = 4 h. Target average, average grade, lower specification (LS) limit and upper and lower 99.7% confidence intervals (mean  $\pm 3sd$ ) are shown; there is no upper specification (US) limit

in the data. Individual analyses are point values in a much wider continuum, with the possibility that much larger and much smaller grades may have passed between the points that were measured. Such information can provide a plant superintendent with a more or less effective means of monitoring the variability in the material stream, depending on the variability of the stream and the frequency of measurements.

The form of the control chart may vary depending on the scale of observation, the length of time over which the random function is measured, the stationarity, and the nature of the variability (Lyman, 2007). The problem with such representations of plant behaviour is that the superintendent's decisions to adjust the process are based on his interpretation of the direction of the trends (either up or down) and their intensity.

The control chart is especially useful when control limits that represent the existing sampling protocol are shown in

relationship to the total variability in the product, as shown in Figure 2. Such plots indicate the volatility of the iron ore product with an average grade of 66.1 %Fe. It also provides an indication of the ability of operators to maintain stability of the product grade and periods when control of product grade has been inadequate. With a lower specification limit (LS) of 65 %Fe, there have been occasions when it has not been possible to maintain the grade of the iron ore product above the LS (Figure 2). One may argue that the specification limits are too narrow, and that the supplier probably has good grounds for lowering the LSL to 64%Fe to include a greater proportion of the variability. The upper and lower 99.7% confidence interval in Figure 2 encloses a large proportion of the variability, but there is a regularity that suggests that the lower confidence limit is breached at regular intervals. In general, the variability is erratic, suggesting that attempts to adjust the system are easily subject to over-correction.

# Application of variography to the control of species in material process streams

## Large-scale variability: the moving average

Large-scale variability, usually in the form of product composition cycles and chemical cyclic variations may be manifest throughout a sampling system or a process plant, but this type of variability is most easily discernable in one-dimensional process streams. Pitard (2006, 2007) has provided examples that include stationary or moving streams of solid fragments, liquids, liquid suspensions, slurries, or aerosols. Such streams may be contained on conveyor belts, in pipes or chutes or carried in the open air. In addition, a chronologically ordered set of stationary or moving railroad cars, trucks, jars or barrels can be viewed as just such a stream. A downhole sequence of analyses of borehole core or a long-term list of commodity prices can also be considered as a stream of information.

Without a moving average, the variability plot of the data is too dense for a visual appreciation of the cyclical variability in the data. If there are sufficient data, a wide moving average window is a simple means of identifying large-scale,

cyclical variability in the sampling data. Such cycles may have a regular period, but somewhat irregular amplitude, as in the case of %Fe in the iron ore product, as shown in Figure 3. A 60-point moving average for the total data-set (1511 analyses) indicates strong, well-developed cycles in the data with a period of 208 four-hourly lags, which is about 35 days, as shown in Figure 3. This cyclical characteristic embedded in the data also appears in the variogram and confirms the 35-day period. The operational feature that is considered to be responsible for this cyclical behaviour is difficult to identify with certainty, but is likely to be linked to a combination of the mining operations and the different mining benches that are accessed, as well as the routine involved in building and reclaiming various product stockpiles.

The residual random component, which is subtracted from the original data, is shown in Figure 4. The absence of any cyclical features in these data suggests that there is no underlying trend in the data and that the application of variographic techniques is appropriate for this data-set.

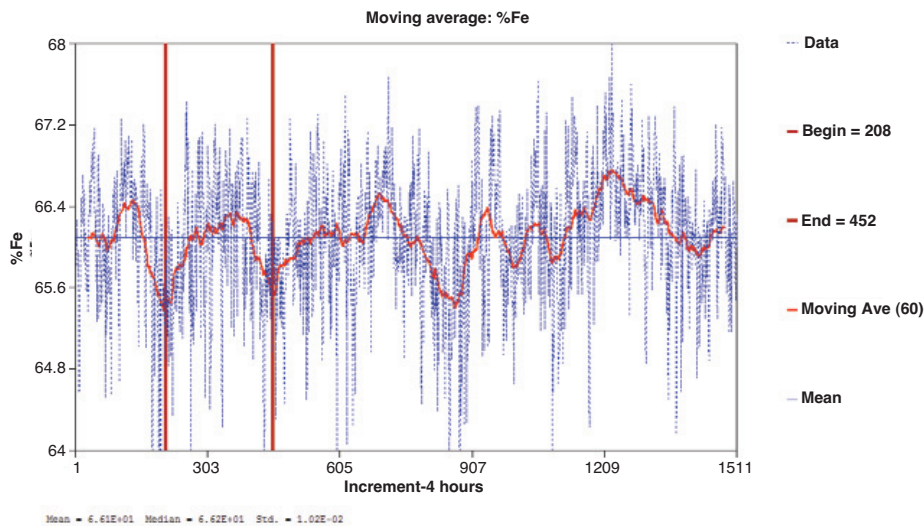


Figure 3—A 60-point moving average of 1 511 %Fe analyses indicating cycles of about 208 lags (832 hours, or 35 days)

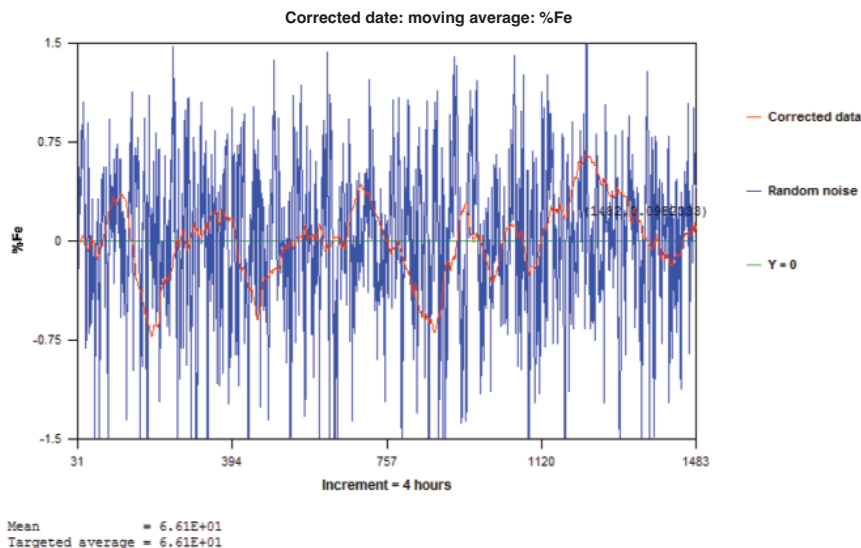


Figure 4—Random noise subtracted from the 60-point moving average. Moving average corrected with respect to the mean



## Application of variography to the control of species in material process streams

Superimposed on the large-scale cycles are numerous shorter period cycles. Using all the data makes the variability chart too crowded to identify cycles at narrow windows, so only 600 data have been used to compile the 5-point moving average, as shown in Figure 5. The 17-lag interval (68 hours) is evident in the data-set irrespective of where the group of data is chosen from in the full data set.

By applying moving average windows of different widths, the 5-point moving average is the shortest interval at which the cyclical behaviour emerged. The component of random noise subtracted from the original data in order to calculate the moving average is shown in Figure 6, and indicates that the data are evenly spread around the horizontal zero line.

This type of spread in the random noise suggests that the parent data are from a normal distribution, that there is no underlying large-scale trend, that the window for the moving average is not too wide, and that the 5-point moving average is an acceptable representation of the cyclicity in the data. The symmetry of the variation of random noise around the

zero line (Figure 6) provides an indication of the appropriateness of the width of the moving average window. If the random noise shows a positive or negative trend, and strong cyclical patterns, the window is probably too wide and should be reduced in width until the random noise becomes symmetrical about the zero line. The absence of trends or cycles in the random noise around the zero line shown in Figure 6 suggests that the width of the moving average window is appropriately set for these data.

### Principles of variography

Variability in process streams arises due to heterogeneity in a system or stream and the time dependence of sampling events is best handled using a variogram. Different components of variability may be difficult to resolve since one kind of variability could mask or include another. According to Pitard (2007) chronostatistics covers the entire domain of variability, but he presents a number of caveats

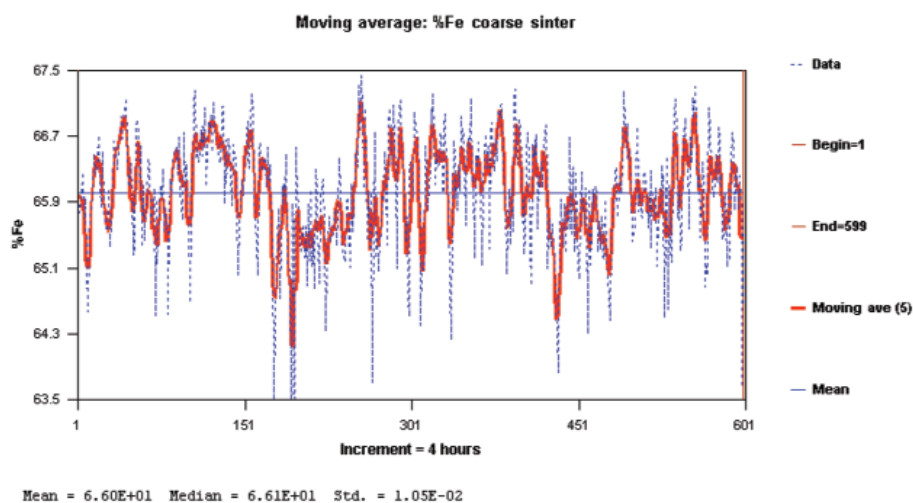


Figure 5—A 5-point moving average of 601 %Fe analyses indicates cycles of about 17 lags (68 hours, or 2.8 days) probably related to small stockpile building and reclamation

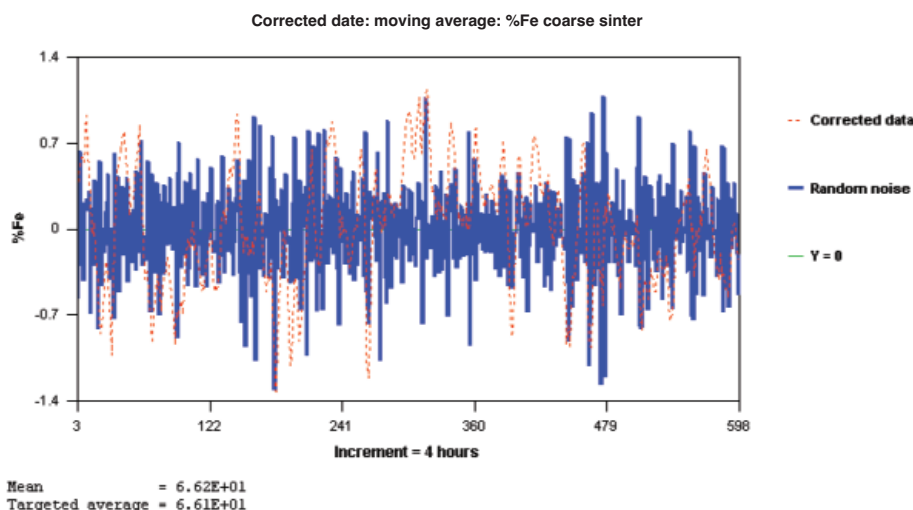


Figure 6—Random noise subtracted from the 5-point moving average. Corrected with respect to the mean

# Application of variography to the control of species in material process streams

for the input data: the lag must be reasonably constant through time, no important changes should be made to the process while analytical data is being captured, data integrity must be preserved, and the basic sampling interval should not be too long (days or weeks). Furthermore, he explains that the components of variability around a Targeted average in a control chart are the integrated accumulation of stochastic phenomena that include random, functional and continuous elements, each with independent sources. For example, for size distribution in a stream the heterogeneity could be expressed as:

$$f[t] = f_L + f_A[t] + f_B[t] + f_C[t] \quad [1]$$

where:

- $f_L$ : is a constant term, e.g. the average proportion of a given size fraction of a stream during period  $f[t]$
- $f_A[t]$ : is a random, discontinuous component of heterogeneity characterized by chaotic and unpredictable changes in the size fraction at very short time intervals, despite the fact that fragments are sourced from the same place earlier in the process
- $f_B[t]$ : is an essentially continuous, non-random, non-periodic, and inherent property of the feed such as changing ore density or hardness
- $f_C[t]$ : is a continuous periodic component arising from a cyclical imprint imposed on processes due to mechanical equipment and people (procedural routine), e.g. cycles in the feed to a SAG mill due to coarse-fine fragment segregation in a stockpile.

Since the variance of a constant is zero, the term  $f_L$  is obviously zero, and the total heterogeneity affecting the parameter of interest in the stream is fully characterized by a linear combination of  $f_A[t]$ ,  $f_B[t]$ , and  $f_C[t]$ .

The variability, and by analogy the heterogeneity, in a moving stream of material can best be represented by a variogram, a model for which is shown in Figure 7. Generally, variography is a graphical method of representing the difference in grade as a semi-variance between samples at different spatial locations. In the case of a process plant, the spatial separation is simply the time lag between collecting one sample and the next, as illustrated in the top row of Figure 7. The semi-variance between sample values at increasing distance (time) between samples is calculated for the continuous stream of data. The semi-variance at increasing sample intervals, known as lags, along the stream of data is then plotted against the lag distance to produce the variogram. In the first step, the variance between the closest consecutive samples  $a_i$  and  $a_j$  is calculated and averaged according to Equation [2] and provides the first data point on the variogram.

$$\gamma_1^2 = \frac{1}{2N} \sum (a_i - a_j)^2 \quad [2]$$

The second step is to calculate and average the variance between every second sample, and then every third sample and so on. A typical absolute variogram compiled from such information is shown in Figure 8. Two variograms, the absolute variogram and relative variogram are defined. The absolute variogram ( $V_a(j)$ ) has the following equation:

$$V_a(j) = \frac{1}{2(N_u - j)} \sum_i [a_{i+j} - a_i]^2 \quad [3]$$

where  $a_i$  is the grade of the sample and  $N$  is the number of pairs of samples used for calculating the variance. By dividing the average variance for each point in the variogram by the mean squared of the points used in the variogram, the relative variogram ( $V_r(j)$ ) is derived:

$$V_r(j) = \frac{1}{2(N_u - j)a_L^2} \sum_i [a_{i+j} - a_i]^2 \quad [4]$$

The relative variogram ( $V_r(j)$ ) provides a means of comparing the information from one variogram with information from another variogram for data at a different lag, for instance. The calculations for a variogram for any metal  $X$  shown in Table II are plotted in Figure 9. The main parameters of a variogram include:

- The random variability referred to as  $V[0]$  (Figure 8) is a very short-range, random, and discontinuous term representing the inherent variability in the sample data. In rock materials mineralized with precious metals this is due to the presence of large nuggetty grains or the clustering of small grains. In process streams the short range random variability  $V[0]$  is a function of the

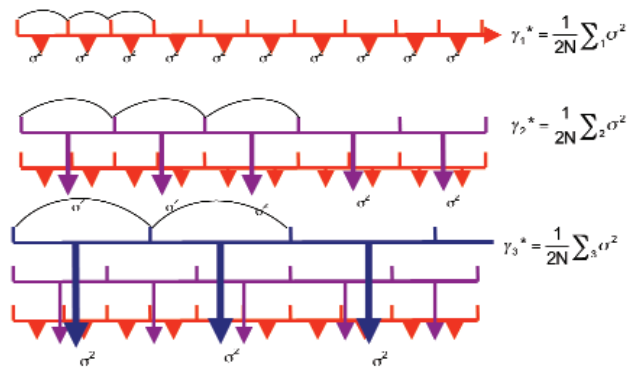


Figure 7—Top row: A series of sample points separated by distance  $j$  (the lag); second row: a series of points separated by distance  $2j$ ; third row: a series of points separated by distance  $3j$ , etc.

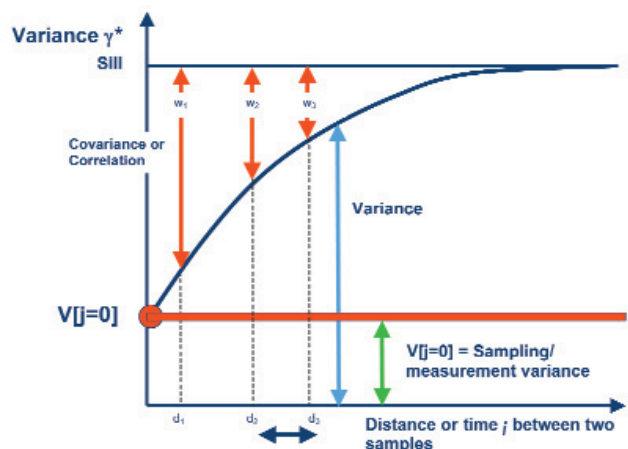


Figure 8—Components of the variogram indicating  $V[j=0]$ , commonly referred to as the nugget effect

# Application of variography to the control of species in material process streams

Table II

Example of a template for calculating the values plotted in a variogram

4 h	%Fe	Lag (J)	1	2	3	4	5	6	7	8
1	66.3	Avg (5)								
2	65.9		0.32							
3	65.2	65.4	0.06	0.10						
4	65.1	65.2	0.64	0.31	1.27					
5	64.6	65.1	0.34	1.91	1.30	2.91				
6	64.6	65.1	0.00	0.30	1.81	1.22	2.79			
7	65.4	65.2	0.62	0.68	0.06	0.31	0.10	0.77		
8	65.8	65.5	0.15	1.37	1.45	0.39	0.03	0.00	0.25	
9	65.8	65.9	0.00	0.18	1.49	1.57	0.45	0.02	0.01	0.20
10	66.0	66.1	0.05	0.08	0.43	2.09	2.20	0.81	0.01	0.12
11	66.5	66.2	0.20	0.45	0.52	1.21	3.58	3.71	1.81	0.30
12	66.2	66.3	0.06	0.04	0.18	0.22	0.72	2.69	2.81	1.20
13	66.3	66.4	0.01	0.02	0.09	0.28	0.34	0.92	3.06	3.19
14	66.4	66.4	0.01	0.04	0.00	0.15	0.38	0.44	1.09	3.36
15	66.7	66.5	0.08	0.13	0.23	0.05	0.45	0.80	0.89	1.76
Mean	65.8	Variance	8.9E-06	2.0E-05	3.3E-05	4.0E-05	4.4E-05	4.2E-05	4.2E-05	4.5E-05
Std Dev	1.0									

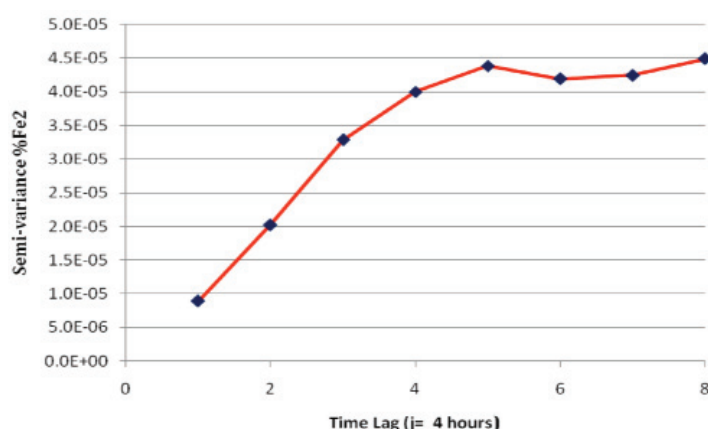


Figure 9—An absolute variogram plotted from data shown in Table II

fundamental error (FE), the grouping and segregation error (GSE), and any other spurious or unaccounted for errors:

- *The sill*—a measure of the total variability in the system or the set of data beyond the point where samples are correlated with one another
- *The range of influence*—the point beyond which there is no correlation between data points.
- The number of components defined by the shape of the variogram between 0 and the total sill. Many variograms have one component that has a single sill and a single range of influence. Kinks in the variogram that divide it into portions with different gradients indicate the presence of more than one component. In such cases each portion of the variogram comprises a different component with its own sill and own range of influence.

The principal difference between the variogram as applied in geostatistics and chronostatistics is the stationarity of the

different functions. In geostatistical analyses stationarity refers to the absence of a trend, meaning that the mean across the area being estimated may be considered to be more or less constant. In dealing with moving streams, however, there is no stationarity because the average grade over a given period of time may change. Nevertheless, the chrono-variogram provides a means of identifying a variety of problems related to the sampling of the material stream. Importantly the chrono-variogram provides a tool for identifying the nature and range of variability related to the random variability, as well as long-range and short-range cycles in the process stream that can be related back to a process variability chart (Shewart chart). This allows the plant superintendent to identify sampling events in the process that are 'out of control', compared to those that are easily controlled. This provides logical, pragmatic and definitive solutions to improving the process as well as saving money at the same time (Pitard, 2007).

# Application of variography to the control of species in material process streams

The benefit of the variogram is that it graphically represents all types of variability in a sampling stream. Any single calculation of the variance in a sampling stream is an aggregate of all types of variability, but the variogram allows the components of variability to be identified and resolved. As a consequence of studying many variograms, the four major components of the variogram have been identified and these are explained in detail by Pitard (2007) as:

$$V[j] = V_A[j] + V_B[j] + V_C[j] + V_D[j] \quad [5]$$

where:

$V_A[j]$  is a very short-range, random, and discontinuous term. At the limit when  $j = 0$ , the term  $V[j=0]$ , simplified to  $V[0]$ , is referred to as the short range random variability, the variability from sampling, subsampling, and measurement that does not exist in the process.

$V_B[j]$  is a long-range, usually non-random, and continuous term that is a direct consequence of bias in the process. This variability can be controlled and reduced by implementing an optimized sampling protocol in a process.

$V_C[j]$  is a periodic, continuous term, linked to the routine procedures that people employ or to the cyclical processes mechanical equipment impose on processes. This source of variability is usually poorly understood, leading to over-corrections in system processes by plant personnel, and inducing losses of process efficiency. The economic impact of this source of variability, if misunderstood, is enormous.

$V_D[j]$  is a random residual variability tied to the variogram precision when the variance  $V[j]$  is calculated with too few pairs  $N-j$ .  $V_D[j]$  tends to zero when the number of pairs increases. It is not good practice to calculate any point on a variogram with fewer than 20 pairs; 30 pairs or more are strongly recommended.

## Extrapolation of the variogram to time or distance zero

An accurate extrapolation of the variogram to time or distance zero is required for many of its applications. The most effective solution as demonstrated by Pitard (2007) is to extrapolate either the first order integral  $W[j]$  or the second order integral  $W'[j]$  of the variogram. Indeed,  $W[j]$  and  $W'[j]$  smooth out residual noise due to  $V_D[j]$ , cycles, and other features interfering with a good extrapolation. As demonstrated in the sampling theory, the variogram  $V[j]$ ,  $W[j]$ , and  $W'[j]$  all have the same intercept  $V[0]$ ,  $W[0]$ , and  $W'[0]$  at time or distance zero (Gy, 1988; Pitard, 1993).

The most effective way to calculate  $W[j]$  and  $W'[j]$  is by using Gy's point-by-point interpretation of the variogram. A surface area  $S[j]$  under any given value of the lag  $j$  of the variogram is estimated as follows: give each point  $j = 1, 2, 3, 4, \dots, n-1$  an interval of influence  $j = 1$ ; then, give the extreme values  $V[0]$  and  $V[j=n]$  an interval of influence equal  $j = 1/2$ . The generalization of this approach led to a formula convenient to program in a spreadsheet for any value of  $j \geq 1$ :

$$W[j] = \frac{1}{j} EST. \left[ \int_0^j V(j') dj' \right] = \frac{S[j]}{j} = \quad [6]$$

$$S[j-1] + \left[ \frac{V(j=1)}{2} \right] + \left[ \frac{V(j)}{2} \right]$$

For  $j=0$ ,  $S[0] = 0$ , and  $W[0] = V[0]$ .

Following a similar approach, calculation of the second order integral leads to the following pragmatic formula:

$$\begin{aligned} W'[j] &= \frac{1}{j} \int_0^j W[j'] dj' = \frac{2}{j^2} \int_0^j dj' \int_0^{j'} V[j''] dj'' \\ &= \frac{2}{j^2} \int_0^j j' W[j'] dj' = \frac{2}{j^2} \int_0^j S[j'] dj' \end{aligned} \quad [7]$$

If we define  $S'[j] = \int_0^j S[j'] dj'$ , then

$$W'[j] = \frac{2}{j^2} S'[j] \quad [8]$$

Then for  $j > 0$ :

$$W'[j] = \frac{2}{j^2} \left[ S'(j-1) + \frac{S(j-1)}{2} + \frac{S(j)}{2} \right] \quad [9]$$

## Important limitations for the variogram

The use and interpretation of the variogram is constrained by a number of important limitations identified by Pitard (2007). These include:

- The term  $V_D[j]$  of the variogram (i.e., precision of the variogram) must remain as small as possible. Under no circumstance should any point on the variogram be calculated with fewer than 20 pairs
- In a chronological series of data, there is a problem associated with the central values of that series. If the selected lag  $j$  of the variogram goes beyond half of the data available, the calculation of the variance for that lag can no longer use data located at the centre of the series. Therefore, the minimum number of pairs  $N-j$  should remain larger or equal to  $N/2$
- A process is not stationary. It is in a permanent state of change. Most processes in the mining and chemical industries are dynamic processes. The general trend observed today, or this week, may be different from the one observed tomorrow or next week. When looked at on a large scale of time, these trends may actually carry a random component. Therefore, before calculating a variogram, it is good practice to take a good look at the chronological data, and select a window within which the general trend is reasonably consistent: some would say, in a variographic analysis 'do not mix oranges and apples'.

The basic tools needed by managers and supervisors in order to build an effective, pragmatic strategy for analysing existing chronological data have been described. Existing sampling data collected at great expense during production are invaluable for identifying the source and origin of annoying problems plaguing the daily process optimization. Now, let's enter the heart of chronostatistics. Like a skilled detective, the variogram has the ability to disaggregate the components of process variability from one another. A complex problem divided into its basic components is much easier for managers who wish to set priorities for continuous improvement.

## Variographic analysis: interpretation of the absolute variogram

The absolute variogram for 800 points is shown in Figure 10, indicating that very large low amplitude cycles occur within



# Application of variography to the control of species in material process streams

the data. The first and most pronounced cycle ends at approximately 208 lags. Although the cycle continues to 248 lags, the fact that the sill and trend meet at 180 lags suggests that beyond 208 lags the variability may be purely random (Figure 10). The variogram confirms what is already known about the data-set from the moving average plot (Figure 3) and the process performance that there is a cycle in the metallurgical stream with a period of about 35 days.

Figure 11 represents only the portion of the absolute variogram between 0 and 250 lags. The vertical red dashed line ( $V_{trend\_j}$ ) shows the 35 day period for a single large cycle. Superimposed on this large cycle are numerous short-period cycles with reasonably regular amplitude (see also Figure 5). The smaller cycles have a regular period of about 4

days (3.92) shown by the vertical green line ( $V_{cycle\_j}$ ) in Figure 11. The cyclical behaviour is due to a combination of mining activity in iron ores with varying chemical characteristics and the way in which blending stockpiles are built and reclaimed over the period. Details of the components of variability in the absolute variogram shown in Figure 11 are listed in Table III with a description of how they arise and how they are calculated.

The value for  $V[0]$  is the intersection of the first order integral  $V[trend]$  (the green line) on the y-axis. Pitard (1993) suggested that such extrapolation may not be valid if:

- $j = 1$  is long, i.e. hours or days, then  $V[j = 1]$  includes the true random variability, the unknown non-random variability, and the unknown periodic variability.

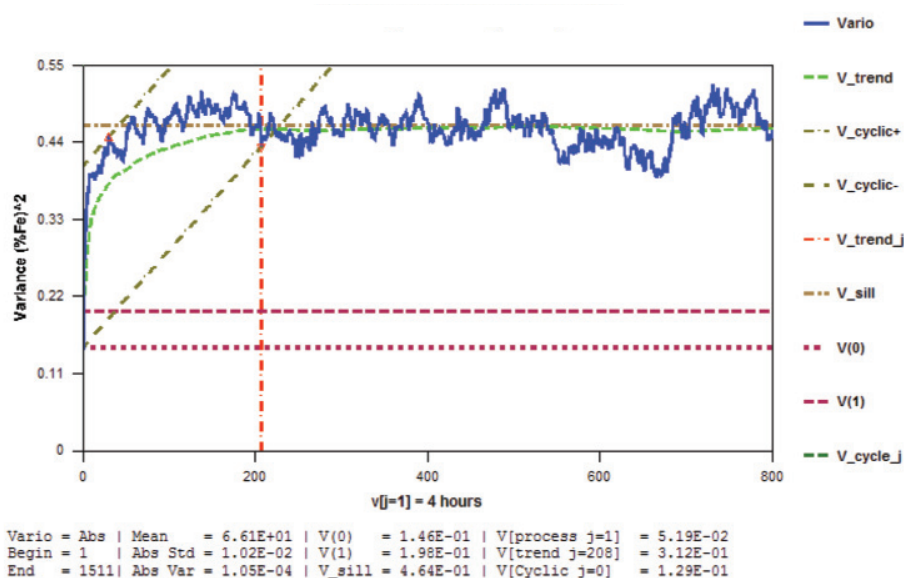


Figure 10—Absolute variogram of %Fe in an iron ore product up to 800 lags; the variogram reaches stability after 208 lags or 35 days

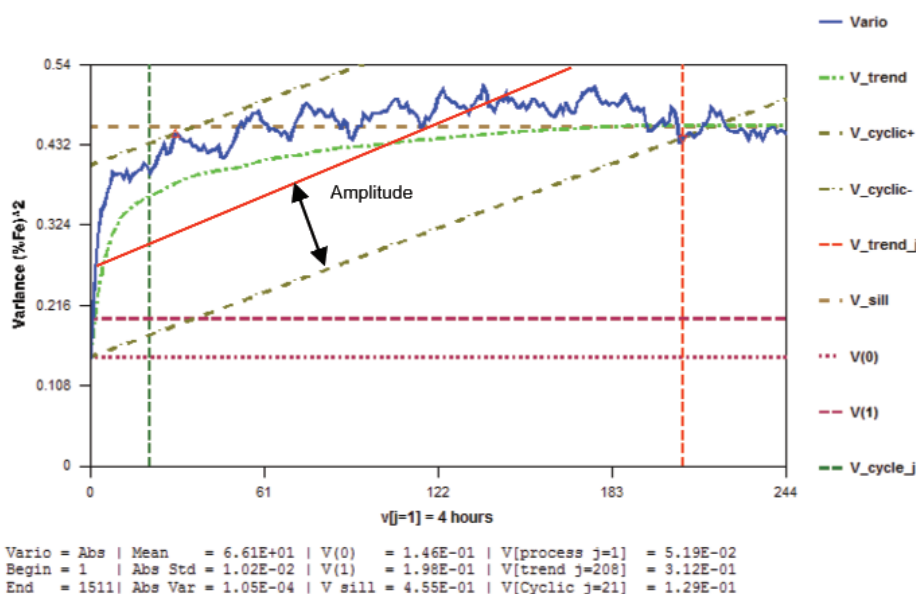


Figure 11—Absolute variogram for an iron ore product showing extremely strong sub-cycles with regular period (4.16 days) and regular amplitude. Three components comprising  $V[trend]$  are evident

# Application of variography to the control of species in material process streams

Table III Symbol, source, and amount of variability in the absolute variogram (Figure 11)*		
Symbol	Source of variability	Variance
Vario	Represented by a thick blue line in Figure 11, is the variogram over a period of 250 lags or 42 days	
Mean	66.10 %Fe	(%Fe <sup>2</sup> )
Rel std:	Relative standard deviation	0.0102
Rel var:	Relative variance	0.000105
V[0]	Random variability; shown as the lower, horizontal, thick red dotted line in Figure 11. It is the total sampling and measuring variability of the process	0.146
V[1]	Total process variability; shown as the purple dotted line in Figure 11, just above the line for V[0]	0.198
V[process]	Process variability is measured by the difference between V[0] and V[1]	0.0519
V[cyclic]	V[cyclic] + and V[cyclic]—represent the boundaries along the upper and lower limits of the cyclical variability. Half the total cyclical variability is the amplitude of the cycle	0.129
V[sill]	V[sill] is the average variability of the process measured across all the data. It is represented by a brown horizontal line in Figure 11 and coincides with the flattening out of the variogram (blue line)	0.455
V[trend]	Is shown on Figure 11 by a green dotted line and represents the first order integral W that is used to extrapolate the variogram back to V[0]. V[trend] can be measured at any lag distance, but usually at the specific lag point where the variogram reaches a maximum, i.e. the range of influence.	0.312

\*Sourced from Pitard 1993 and 2007.

However, extrapolation may be acceptable if:

- $j = 1$  is short, i.e. seconds or a few minutes, then  $V[1]$  probably includes only the random variability.

The control chart (Figure 2) indicates three data quality objectives (DQOs, EPA 2007), namely the target average, the process average, and the lower specification limit (LSL), the latter being set by the purchaser of the product. These DQOs define the limits to acceptable variability in the process and provide the basis for identifying and evaluating the source and components of sampling and process variability.

- *Random variability  $V[0] = 0.146\%Fe^2$* —This short-range random or irrelevant variability is not related to the plant process, but is a combination of inherent heterogeneity (random variability), fundamental error (FE) and grouping and segregation errors (GSE)) and all uncontrolled sampling errors arising from a poor sampling protocol, and is a function of the total sampling and measurement variability (Pitard, 1993, 2007). In the rest of this text  $V[0]$  is simply referred to as the random variability or ‘sampling variability’. If  $V[0]$  is large, other errors are also likely to be large. The nugget effect accounts for about 32% of the total sill, which is a relative measure of the overall variation. This component of error could be reduced if appropriate attention is paid to the sampling protocol.

Upper and lower control limits (UCL and LCL) are set in the control charts (Figures 13 and 14) by multiplying the standard deviation of  $V[0]$  by 3 in order to cover the 99.7% confidence interval; UCL and LCL are calculated as follows:

$$\begin{aligned} V[0] &= 0.146 \%Fe^2 \\ \sqrt{V[0]} &= 0.3821\%Fe \\ \text{Std dev} &= \pm 0.3821 \%Fe \\ \text{UCL} &= \text{Mean} + 3\sqrt{V[0]} = 66.10 + 3 \times 0.382 = 67.25 \%Fe \\ \text{LCL} &= \text{Mean} - 3\sqrt{V[0]} = 66.10 - 3 \times 0.382 = 64.95 \%Fe \end{aligned}$$

In this particular case the actual mean (66.10 %Fe) rather than the target average (66.0 %Fe) has been used to define the limits.

- *Total process variability  $V[1] = 0.198 \%Fe$* —This variability cannot be controlled unless the routine sampling interval is reduced. It is the typical value of  $V[1]$  at the first lag point in the variogram and is a combination of the total sampling and measuring variability of the process and the total non-random variation that occurs in the plant between any two consecutive analyses. The non-random component of variability is due to bias in the sampling process related to the delimitation error (DE), the extraction error (EE), the preparation error (PE) or the analytical error (AE) and can be eliminated through implementing an optimized sampling protocol.
- *Process variability  $V[\text{process}] = 0.0519\%Fe^2$* —The process variance,  $V[\text{process}]$ , is simply the difference between  $V[1]$  and  $V[0]$ , in this case  $0.0519\%Fe$ . The ratio  $V[0]$  to  $V[1]$  is 0.74 with  $V[1]$  being about 26% higher than  $V[0]$ , suggesting that considerable variability exists in the process between analyses. Process variability cannot be controlled unless the sampling interval is reduced, but in this particular case the sampling interval is appropriate and process variability can be controlled.

UCL’ and LCL’ include both the random variability due to the random, and the non-random process variability that takes place between two consecutive samples. These upper and lower control limits combine three standard deviations of the random variability  $V[0]$ , and the process variability  $V[\text{process}]$ , to give a 99.7% level of confidence (Pitard, 1993) in the control charts shown in Figures 13 and 14.

The upper and lower control limits (UCL’ and LCL’) are set in the control chart by multiplying the standard deviation of  $V[0]$  by 3 in order to cover the 99.7% confidence interval, and then adding the contribution from  $\sqrt{V[\text{process}]}$ . The process variability is calculated as follows:

$$\begin{aligned} V[1] - V[0] &= V[\text{process}] = 0.198 - 0.146 = 0.0519\%Fe^2 \\ \sqrt{V[\text{process}]} &= 0.2278\%Fe \end{aligned}$$

## Application of variography to the control of species in material process streams

$$UCL' = \text{Mean} + 3\sqrt{V[0]} + \sqrt{V[\text{process}]} = 66.10 + 1.1463 + 0.2278 = 67.47 \% \text{Fe}$$

$$LCL' = \text{Mean} - 3\sqrt{V[0]} - \sqrt{V[\text{process}]} = 66.10 - 1.1463 - 0.2278 = 64.73 \% \text{Fe}$$

The position of the sampling, measuring and process variability relative to the iron ore product stream is shown in the control chart as UCL' and LCL', the long-dash, light-green lines in Figure 13 and 14.

- *Cyclical variability*  $V[\text{cyclic}] = 0.129 \% \text{Fe}^2$ —The value for  $V[\text{cyclic}]$  (as reported on Figure 11) is half the total amplitude of the process cycle, between the highest and lowest points on the variogram, usually associated with the first cycle. It is a non-random variable related to specific activity in the process. In the case of the short-range analysis, the amplitude is spread over three cycles. The periods for the large-scale cycle is 208 lags or about 35 days (Figure 10), while the short-range cycles have a period of about 23.5 lags or about 4 days (Figure 11).  $V[\text{cyclic}]$  is calculated as follows:

$$V[\text{cyclic}] = 0.129 \% \text{Fe}^2 \text{ (this is the total amplitude shown in Figure 11)}$$

$$\sqrt{V[\text{cyclic}]} = 0.3592 \% \text{Fe}$$

$$UCL'' = UCL' + \sqrt{V[\text{cyclic}]} = 67.47 + 0.3592 = 67.73 \% \text{Fe}$$

$$LCL'' = LCL' - \sqrt{V[\text{cyclic}]} = 64.73 - 0.3592 = 64.27 \% \text{Fe}$$

This variability is introduced as a direct consequence of interventions on, or interactions with, the process stream. It may be related to mechanical or human interventions, but is usually due to periodic changes in the diurnal performance of, or maintenance interference with equipment, or due to changes in manpower behaviour or material inputs on the plant. The regularity of both the period and the amplitude of the short-range cycles in the variogram suggest that this effect is introduced by mechanical equipment. The reason for the cycles should be identified and adjustments made to the sampling equipment.

- $V[\text{sill}] = 0.455 \% \text{Fe}^2$ —This is a measure of the average variability of the process measured across all the data.  $V[\text{sill}]$  is measured across the entire variogram and is a measure of the total variance in the data set.
- $V[\text{trend}] = 0.312 \% \text{Fe}^2$ —This is the first order integral of the variogram at any given lag; it is the difference between  $V[0]$  and the first integral of the variogram at any lag (Figure 12).  $V[\text{trend}]$  with a variance of  $0.312 \% \text{Fe}^2$  is slightly above the sill at 208 lags, the difference between  $V[\text{sill}]$  and  $V[0]$  being  $0.455 - 0.146 = 0.309 \% \text{Fe}^2$ . Beyond the range,  $V[\text{trend}]$  is more or less constant.

The trend, being the first order integral of the variogram, rises sharply in the short-lag portions of the variogram and flattens out as it approaches the range of influence. The trend provides an explanation of the variogram behaviour during the period of the large-scale cycle. This component of variability is also due to some mechanical or human intervention that occurs every 35 days ( $208 \times 4 \text{ h}$  lags), and introduces variability into the system. Generally the trend of the variogram is upwards until a point (the range) is reached, beyond which the variogram is level or declines. The range in this case is about 35 days.

Between  $V[0]$  and the range of influence three components in  $V[\text{trend}]$  contribute to the curve as shown in the absolute variogram (Figure 11). The first component comprises the very steeply rising portion of the variogram between 0–8 lags, equivalent to the first 32-hour period. The second component is the moderately rising portion of the variogram between 8–40 lags, equivalent to 108 hours or 6 days. The third component of  $V[\text{trend}]$  is the gently rising portion of the variogram between 36 and 208 lags (35 days). The lag and sill values for the three individual components of  $V[\text{trend}]$  are listed in Table IV. Beyond 208 lags,  $V[\text{trend}]$  is horizontal, indicating that the variogram is stable. The different gradient of the three components of  $V[\text{trend}]$  between 0 and 180 lags indicates the rate at which the system is moving towards stability; a high rate for steep gradient and slow rate for the gentle gradient.

At 180 lag points, the variogram  $V[\text{trend}]$  flattens out and remains constant at  $0.312 \% \text{Fe}^2$  over the full set of data; this means that there is no correlation between data points beyond about 208 lag points. The sill and variogram trend meet at about 208 lags, the range of influence. The main insight provided from the absolute variogram is that the semi-variance only stabilizes beyond 208 lags (35 days), a long time for the process to reach stability. The variogram varies around the sill until 540 lags at which point another cycle appears. However, beyond about 400 lags, the behaviour in the variogram is due only to random residual variation and is probably not relevant to this analysis (Figure 10).

### Description of the components of variability

The average grade (%Fe), the targeted average (TA) and the lower specification (LS) limit set according to the customers (or suppliers) product requirements is shown on the control chart (Figure 12). The sources of variability listed in Table III are successively shown on control charts (Figures 13 and 14) and allow the overall process variation to be examined relative to the control limits imposed by the DQOs.

### Establishing upper and lower sampling guidelines (USG and LSG)

An indication of acceptable levels of process variability is established by compiling upper and lower sampling guidelines (USG and LSG; dotted green lines) on the control chart as shown in Figure 12. These guidelines indicate the maximum acceptable total variability due to sampling, measurement and the four-hourly sampling interval. Pitard (2006) has suggested that these be arbitrarily located one-

Table IV

#### Lag distance and sill values for the components of $V[\text{trend}]$ in the absolute variogram

Components	Lag distance (lag = 4 hours)	Sill ( $\% \text{Fe}^2$ )
First component	8 (32 hours)	0.313
Second component	36 (6 days)	0.390
Third component	208 (35)	0.455

# Application of variography to the control of species in material process streams

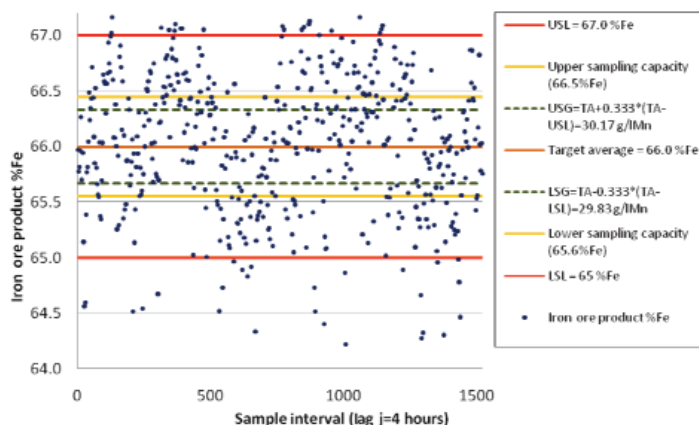


Figure 12—Control chart for an iron ore product showing the upper and lower sampling guideline limits, and upper and lower sampling capacity limits that lie between the upper and lower specification limits

third of the distance between the target average and the USL and LSL. It is important to remember that these are only suggested guidelines, but they do come with the years of extensive practical experience and insight from Francis Pitard. The upper specification limit is natural since iron ores do not get any richer in iron than 67%Fe.

## Establishing upper and lower sampling capacity (USC and LSC)

The first step is to define the so-called upper (USC) and lower sampling capacity (LSC) DQOs in the control chart (Figure 12). The upper and lower DQOs for the existing protocol, shown by the solid yellow lines (USC and LSC) in Figure 12, are derived by placing  $\sqrt{V[1]}$  around the target average. These limits indicate the current maximum acceptable total variability induced by sampling, analytical measurement and the four-hourly variability between sampling events in the plant. They describe the variability of the process  $\sqrt{V[1]}$  around the target average, so:

$$USC = TA + \sqrt{V(1)} = 66.45 \%Fe$$

$$LSC = TA - \sqrt{V(1)} = 65.56 \%Fe$$

The sampling guidelines lie between the limits set by the sampling capacity (USC and LSC; solid yellow lines) respectively, indicating that these behaviours are not compatible. This is a clear indication that the entire existing protocol is incapable of providing a means of controlling the process variability. Typically one would want the USC and LSC limits to be as close to the target average as possible since that would reflect very little variability and bias in the sampling process. The fact that both the upper and lower sampling capacities lie outside the limits set by the sampling guidelines as shown in Figure 12, suggests that the sampling and measurement protocol is in need of attention in order to function appropriately.

In order to avoid redundancy, the upper and lower limits for the sampling guidelines and sampling capacities are not shown in the following diagrams because all the available information from these control charts has been derived.

## Control limits from $V[0]$

The upper and lower control limits (UCL and LCL) are shown

by the short-dash dark-green lines in Figure 13. They are placed in the control chart using the mean  $\pm 3\sqrt{V[0]}$  (to give a 99.7% confidence interval) and indicate the total variability due to the random variability, sampling and analytical measurement. These lines lie beyond the upper and lower specification limits (USL and LSL), indicating that the existing sampling and measuring protocol is not performing well, and that the protocol and the DQOs need to be redefined if control over the process variability is to be restored.

## Capability of the 4-hourly sampling interval (UCL' and LCL')

The upper (UCL') and lower control limits ( $= \text{mean} \pm 3\sqrt{V[0]} + \sqrt{V[\text{process}]}$ ) are shown by the long-dashed light-green line in Figure 13. They are interpreted from  $\pm\sqrt{V[\text{process}]}$  and correspond to the small contribution or addition of variability to UCL and LCL arising from the typical process variability between two consecutive 4-hourly sampling events. The extrapolation of the variogram from a 4-hour interval to  $V[0]$  could mask additional components of variability in the period. The only sure way of identifying these potentially hidden sources of variability that can interfere with the most carefully prescribed DQOs is to carry out a short-term experiment by sampling at a shorter interval, say every 15 or 30 minutes.

The short-term experiment should reveal if the following problems are concealed in the short range unsampled variability:

- The sampling and sub-sampling protocol
- The number of increments taken
- A cycle in-phase with the sampling interval
- Sampling stations that are poorly maintained and cleaned
- Sampling systems that are not correctly proportional
- Excessive variability in the laboratory,
- A combination of all the items listed above.

## Process variability contribution from a cycle

The upper and lower control limits ( $LCL'' = LCL' \pm \sqrt{V[\text{cyclic}]}$ ) for process variability as a consequence of cyclic variation are shown in Figure 14. They are defined by the variability due



# Application of variography to the control of species in material process streams

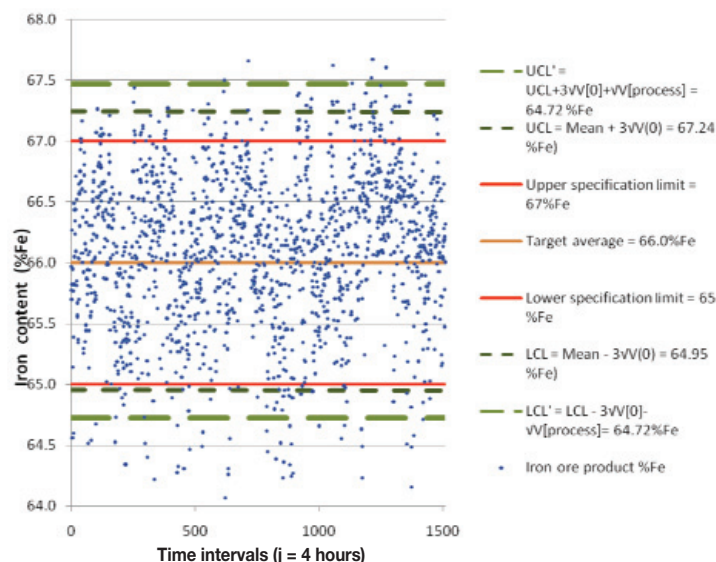


Figure 13—Control chart for an iron ore product showing the upper and lower control limits for testing the validity of the entire protocol for sampling and measurement (UCL and LCL) as well as for sampling, measurement and 4-hourly sampling interval (UCL' and LCL'). These lie outside the variability of the iron ore product defined by the specification limits

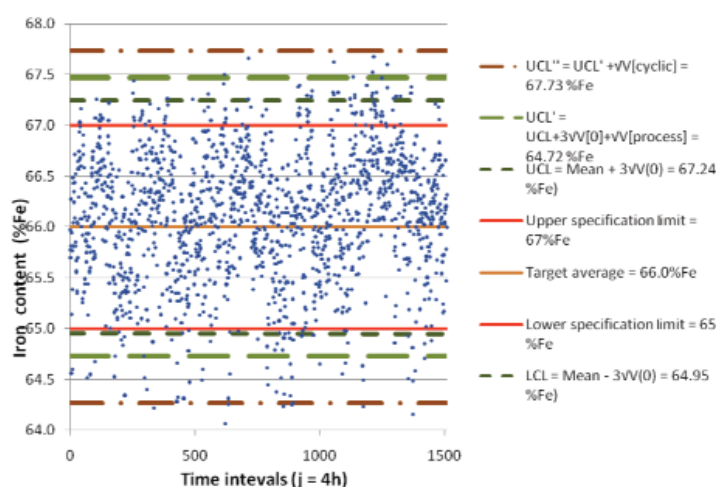


Figure 14—Contribution to process variability due to the large-scale cyclical component

to sampling, analytical measurement, the 4-hourly sampling interval, and the amplitude of the large-scale cyclical component of variability ( $V[\text{cyclic}]$ ). The interpretation of the process from the control chart indicates that the total variability contribution exceeds the variability prescribed by the DQOs, specifically the upper and lower specification limits. This clearly indicates that the process is not under control with the current protocol. In addition, the reasons for the 35-day cycle are poorly understood. If such a cycle were understood, the opportunity exists to reduce the variability and control the process much more rigorously.

## Recommendations for the control limits

Pitard (1993) has made the following recommendation about the positioning of the control limits relative to one another and the specification limits saying that 'Any of the newly defined control limits UCL and LCL, UCL' and LCL' and UCL''

and LCL' should lie less than halfway between the targeted Average (TA) and the Specification Limits (SL)' (p. 448). Figure 14 indicates that all of the control limits lie well beyond the limits he has suggested. The magnitude of  $V[0]$  due to the inherent variability of the ore is such that the grade of the ore simply cannot be controlled within these limits; it is too variable. With the added variability due to the process (UCL' and LCL'), the control lines lie beyond the suggested range. The lower specification limit for the iron ore product is of particular concern because low grade ores require a careful process of blending before they can be sold.

The pie diagram shown in Figure 15 indicates how the relative distribution of variability among the components has been identified. According to this analysis, the most important contributor to the variability is the trend that accounts for 49.4 % of the variability. This is followed by the cyclicity in the system at 20.45% and the uncontrolled

# Application of variography to the control of species in material process streams

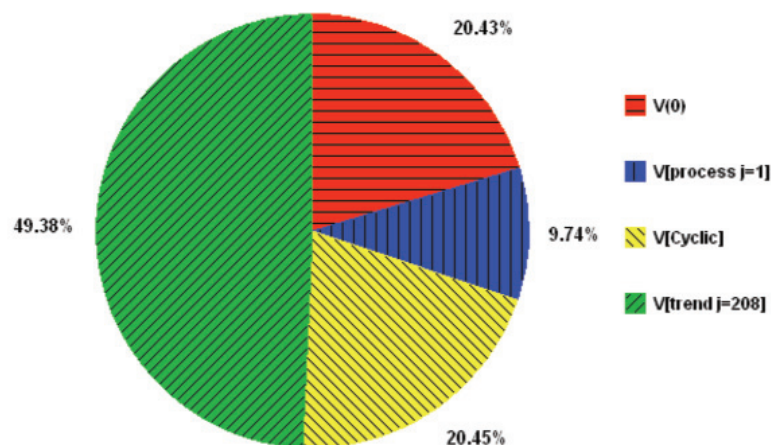


Figure 15—Pie diagram showing the distribution of the components of variability in the system for %Fe

random variability due to sampling error contributes 20.43%. Together the random, measurement, analytical, and cyclical variability contribute just on 50% of the total variability.

## Conclusions and recommendations

Using the 99.7% level of confidence, it is safe to suggest that the lower specification limit is too stringent. The positioning of the control limits on the chart of Figure 14 suggests that the sampling protocols, the variability of the process and the large-scale cyclical behaviour in the overall process are such that the allotted lower specification limit cannot be satisfied. It appears to be too ambitious for the current protocols and the process cannot be controlled under the present conditions. The high value of  $V[0]$  suggests that the sampling variability is a function of the sampling error and that the sampling characteristics of the iron ore product are not properly understood. These characteristics must be more accurately determined by performing heterogeneity tests. This is currently underway. The options are to change LSL and find the dynamic characteristics of the process by performing short-range variographic experiments..

The cyclical nature of the data in the moving average plot and the variograms suggests that the sampling interval could be further investigated. This could be done on an experimental basis, where iron ore samples are collected every 30 minutes for a period of about 3 days. This would provide 144 pieces of data, which would be sufficient to analyse the very short range variability and identify any additional cycles within the currently observable cycles. It is essential to find the source of the cycle and try to minimize it.

Future research should focus on reducing the variability associated with the sampling, measurement and processes in the plant. Reduction of the large-scale variability will provide significant opportunities to improve the product specifications and probably improve costs effectiveness through a less demanding blending routine. In addition the determination of more appropriate specification limits will improve throughput and resource utilization.

## References

- CARRASCO, PEDRO, CARRASCO, PABLO, and JARA, E the economic impact of correct sampling and analysis practices in the copper mining industry. Published in a special issue by Chemometrics and Intelligent Laboratory Systems, Elsevier: *50 years of Pierre Gy's Theory of Sampling Proceedings, First World Conference on Sampling and Blending (WCSB1)*, 2003. pp 209–231.
- EPA. Guidance on Systematic Planning Using the Data Quality Objectives Process EPA QA/G-4. United States Environmental Protection Agency, Office of Environmental Information Washington, DC 20460. EPA/240/B-06/001 February 2006. <http://www.epa.gov/QUALITY/qs-docs/g4-final.pdf>. 2007. 121 pp.
- GY, P. *Sampling of Particulate Materials, Theory and Practice*. Elsevier Scientific Publishing Company. Developments in Geomathematics 4. 1979 and 1983.
- GY, P. *Heterogeneity – Echantillonnage – Homogenisation: Ensemble coherent de theories*. Masson Editeur, Paris. ISBN: 2-225-81313-2. ISSN: 0765-0221. 1988.
- GY, P. *Sampling of Heterogeneous and Dynamic Material Systems: Theories of Heterogeneity, Sampling and Homogenizing*. Amsterdam, Elsevier. 1992.
- GY, P. [Echantillonnage des los de matiere en vue de leur analyse. English] *Sampling for Analytical Purposes*. The Paris School of Physics and Chemistry. Translated by A.G. Royle. 1998 John Wiley and Sons, New York. 1996. 153 pp.
- HOLMES, R.J. Application of variograms to sampling iron ore. CSIRO Institute of Energy and Earth Resources, Division of Mineral Engineering. Port Melbourne Laboratory, Mineral Engineering Communication, October 1986. 41 pp.
- PITARD, F.F. *Pierre Gy's Sampling Theory and Sampling Practice. Heterogeneity, Sampling Correctness, and Statistical Process Control*. Second Edition. CRC Press, New York. 1993. 488 p.
- PITARD, F.F. A short course in The Theory and Practice of Sampling. Presented through the Geostatistical Association of South Africa (GASA) at the University of the Witwatersrand. October 2006.
- PITARD, F. F. Chronostatistics: A powerful, Pragmatic, New Science for Metallurgists *AusIMM Proceedings Metallurgical Plant Design and Operating Strategies Conference*, Perth, Australia, September 18–19, 2006. 22 pp.
- PITARD, F. *Pierre Gy's Sampling Theory and Sampling Practice*. CRC Press, Inc., 2000 N.W. Boca Raton. Second edition, July 1993.
- SPANGENBERG, I.C. An investigation into ROM go-belt sampling at two AngloGold Mines. Mine Metallurgical Managers Association of South Africa. MMMA Circular No. 1/2001 Interim Report of the Council for the year 2001. pp. 1–25. ♦