Process simulation with spreadsheet software

J.W. Merks

Abstract

Spreadsheet software is an effective tool for simulating dynamic stochastic systems such as mineral processing plants. The two-product formula is used to show how to set up a Monte Carlo simulation in a spreadsheet template and how to calculate variances of dependent variables. The key to a reliable and realistic simulation model is to obtain unbiased estimates for all independent variables and variances in the set that defines the stochastic system. Monte Carlo simulation is also a powerful tool for optimizing sampling procedures, for applying sensitivity analysis and for designing and evaluating experiments.

Key words: Modeling and simulation, Computer software, Monte Carlo simulation, Sampling.

Introduction

Spreadsheet software has evolved into an effective, if not indispensable, tool for many applications in science and engineering. A case in point is the Monte Carlo simulation, which is a statistical tool used to assist in understanding the behavior of dynamic stochastic systems or processes. The simulation is used in estimating the variances of functionally dependent variables as fundamental measures for variability and precision. The properties of variances play an essential role in simulation models. Monte Carlo simulations show how sets of independent variables and variances interact, and they determine values and variances of dependent variables.

The @RAND function in Lotus spreadsheet software generates pseudorandom numbers of the standard uniform probability distribution. The term pseudorandom implies that the same seed gives the same sequence of standard uniform random numbers (SURNs). Seeding the @RAND function with some internal variable of the microprocessor would reduce the probability for unwanted periodicities to occur. It is highly improbable that Monte Carlo simulations with spreadsheet software, when applied to a dynamic stochastic system such as a mineral-processing plant, will be flawed due to associative dependence between ordered sets of SURNs.

The proposed simulation model is based on the premise that constant input variances underestimate the intrinsic variability of the stochastic variable of interest (Merks, 1991). Therefore, constant input variances are used to obtain variance estimates by simulation, which, in turn, are used to obtain normally distributed random numbers (NDRNs) for dependent variables. Monte Carlo simulation eliminates the need to obtain the partial derivatives towards more complex dependent variables associated with the three-product formula, such as recoveries in percent or metal contents of concentrates and tailing.

The formula for the variance of a general function, as defined in probability theory, can be used to derive the formula for the variance of the mass of metal contained in a quantity of mill feed, concentrate or tailing (Merks 1985; International Organization for Standardization, 1996). A simplified version of the same formula is used to convert measured values of independent variables and their coefficients of variation into variances of contained metals or metal contents.

Population parameters and sample statistics

Mathematical probability deals with the properties of fundamental distributions such as Gaussian or normal, binomial, Poisson and standard uniform. The latter distribution and the standard normal distribution play essential roles in Monte Carlo simulations. Population parameters such as the mean and the variance are defined for each of those probability distributions.

The Central Limit Theorem implies that the statistics of samples selected from the standard uniform distribution tend towards those of the standard normal distribution, as the number of samples increases. Under the same conditions, the standardized statistics of samples selected from binomial or Poisson distributions tend towards those of the standard normal distribution.

Mathematical probability and sampling theory deal with population parameters such as the population mean (μ) and the population variance (σ^2), whereas applied statistics and sampling practice deal with sample statistics such as $\bar{x}$, the mean of a sample selected from a sampling unit (the wet mass of mill feed processed during a shift), and with *var(spa)*, the sum of the variance of selecting a primary sample, the variance of preparing a test sample of the primary sample, and the variance of selecting and assaying a test portion of a test sample (Merks, 1985).

Standard uniform distribution

The @RAND function in Lotus spreadsheet software gives randomly distributed values of the standard uniform distribution (SURNS). SURNS are equiprobable within the range from zero to unity, so that they can be used to simulate

J.W. Merks is president of Matrix Consultants Ltd., Vancouver, British Columbia, Canada. Nonmeeting paper number 98-328. Original manuscript submitted for review May 1998. Paper accepted for publication September 1998. Discussion of this peer-reviewed and approved paper is invited and must be submitted to SME prior to Nov. 30, 1999.

Table 6a — Sets of 60 on-stream measurements.

Independent variables	Symbol	Value	CV, %
Wet mass of mill feed, t	Mw	50,000	0.5
Moisture content, %	$\%H_2O$	1.8	10.0
Metal grades on dry basis:			
Mill feed	%Me(f)	0.408	1.5
Concentrate	%Me(c)	30.52	1.2
Tailings	%Me(t)	0.037	4.6
No. of on-stream measurements	n	60	–

Dependent variables	Symbol	Feed	Conc.	Tailings
Dry mass, t	Md	49,100	598	48,502
Contained metal, t	Me	200.3	182.4	17.9

Simulated variables	Symbol	Feed	Conc.	Tailings
Wet mass, t	Mw	49,970	–	–
Moisture content, %	$\%H_2O$	1.85	–	–
Dry mass, t	Md	49,045	596.8	48,449
Metal grade, %	%Me	0.408	30.52	0.037
Contained metal, t	Me	200.1	182.2	17.9

Table 6b — Confidence limits for recovery in percent..

Dependent variable	Symbol	Calc.	Simulated
Recovery, %abs	%R	91.04	91.03
Variance, $\%abs^2$	var(R)	–	0.0027
Standard deviation, %abs	sd(R)	–	0.0524
Coefficient of variation, %rel	CV	–	0.06
95% confidence interval*, %	95% CI	–	0.10
95% confidence interval	95% CR		
Lower limit, %	95% CRL	–	90.94
Upper limit, %	95% CRU	–	91.15

*Based on 2*sd(R)

Table 6c — Calculated and simulated variance terms for contained metal.

Variance terms for:	Symbol	Calculated t^2	Calculated %	Simulated t^2	Simulated %
Mass term	var(Me | Mw)	1.00	77.9	0.69	70.5
Moisture term	var(Me | MF)	0.13	10.5	0.13	13.3
Grade term	var(Me | GF)	0.15	11.7	0.16	16.3
Contained metal	var(Me)	1.29	100.0	0.98	100.0

Pairs of interleaving primary samples

The first and second frame of Table 7a lists the same variables as Table 6a with the exception of *n*, the number of interleaving primary samples selected from mill feed, concentrate and tailing during the shift. The third frame gives the NDRNs for the dry mass of mill feed and its metal content. In this case, the NDRN generator requires only two rows, a column for each independent variable and a column for each dependent variable under examination. The generator is not displayed.

Table 7b lists the calculated recovery of 91.0%, a simulated recovery of 90.87% and various precision estimates for the simulated recovery. The variance of $var(Rs) = 0.0726\%^2$ gives a standard deviation of $sd(Rs) = 0.2694\%$, the coefficient of variation of $CV = 0.30\%$rel, a 95% confidence interval of 95% $CI = +0.54\%$ and a symmetric 95% confidence range with a lower limit of 95% $CRL = 90.50\%$ and upper limit of 95% $CRU = 91.58\%$.

Table 7a — Pairs of interpenetrating primary samples.

Independent variables	Symbol	Value	CV, %
Wet mass of mill feed, t	Mw	50,000	0.5
Moisture content, %	$\%H_2O$	1.8	10.0
Metal grades on dry basis:			
Mill feed	%Me(f)	0.408	1.5
Concentrate	%Me(c)	30.52	1.2
Tailings	%Me(t)	0.037	4.6
No. of on-stream measurements	n	2	–

Dependent variables	Symbol	Feed	Conc.	Tailings
Dry mass, t	Md	49,100	598	48,502
Contained metal, t	Me	200.3	182.4	17.9

Simulated variables	Symbol	Feed	Conc.	Tailings
Wet mass, t	Mw	49,970	–	–
Moisture content, %	$\%H_2O$	1.89	–	–
Dry mass, t	Md	48,776	586.3	48,189
Metal grade, %	%Me	0.403	30.48	0.037
Contained metal, t	Me	196.7	178.7	18.0

Table 7b — Confidence limits for recovery in percent..

Dependent variable	Symbol	Calc.	Simulated
Recovery, %abs	%R	91.04	90.87
Variance, $\%abs^2$	var(R)	–	0.0726
Standard deviation, %abs	sd(R)	–	0.2694
Coefficient of variation, %rel	CV	–	0.30
95% confidence interval*, %	95% CI	–	0.54
95% confidence interval	95% CR		
Lower limit, %	95% CRL	–	90.50
Upper limit, %	95% CRU	–	91.58

*Based on 2*sd(R)

Table 7c — Calculated and simulated variance terms for contained metal.

Variance terms for:	Symbol	Calculated t^2	Calculated %	Simulated t^2	Simulated %
Mass term	var(Me | Mw)	1.00	17.7	0.82	6.9
Moisture term	var(Me | MF)	0.13	2.4	0.14	1.2
Grade term	var(Me | GF)	4.51	79.9	10.91	91.9
Contained metal	var(Me)	5.65	100.0	11.87	100.0

Table 7c lists the mass, moisture and grade terms in metric tons squared (t^2) and in relative percent. The mass term is $var(Me|Mw) = Me^2*(CV*0.01)^2 = 200.3^2*(0.5*0.01)^2 = 1.00$ t^2 or 17.7%rel; the moisture term is $var(Me \mid MF) = Me^2*(\%H_2O*CV*0.01)^2/(100-\%H_2O)^2 = 200.3^2*(1.8*10.0*0.01)^2/(100-1.8)^2 = 0.13$ t^2 or 2.4%rel; and the grade term is $var(Me \mid GF) = Me^2*(CV*0.01)^2/2 = 200.3^2*(1.5*0.01)^2/2 = 5.65$ t^2 or 79.9%rel.

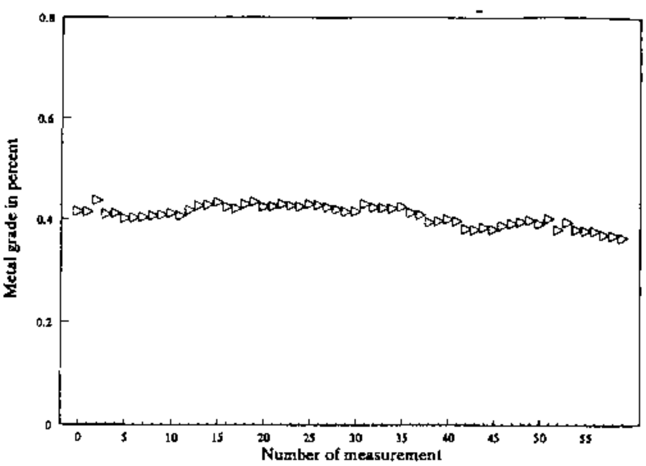

Figure 1 — On-stream measurements of mill feed grade.

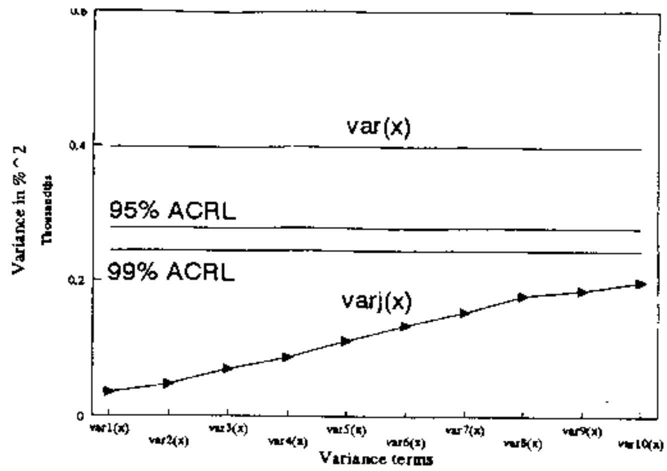

Figure 2 — Sampling variogram for metal grades of mill feed.

most to the variance of content and, thus, to the risk associated with the measurement chain.

Two simulation models will be examined. Both are based on the premise that the wet mass of mill feed is measured with a belt scale. One model is based on measuring the metal grades of cyclone overflow, tailing and concentrate with an on-stream analyzer, and the other is based on measuring the metal grades of pairs of interleaving primary samples. Monte Carlo simulation can also be applied to monthly production data, irrespective of whether daily production data are based on on-stream measurements or on test results for dewatered, dried and prepared slurry samples. For simplicity, reference will be made to mill feed, concentrate and tailing.

Sets of on-stream measurements

This simulation model is based on the arithmetic means of sets of 60 on-stream measurements obtained at 12-min intervals during a 12-h shift and on the first variance terms for the ordered sets. Figure 1 displays a plot of on-stream measurements for the metal grade of mill feed processed during the shift under examination.

The question of whether the ordered set of on-stream measurements in Fig. 1 displays a significant degree of associative dependence is resolved by applying analysis of variance (ANOVA) to the variance of 0.000397%^{2}, for the randomly distributed set, and the first variance term of 0.000035%^{2}, for the ordered set (see Table 5a).

Fisher's F-test indicates that the calculated F-ratio of $F = var(x)/varj(x)$ = 0.000397/0.000035 = 11.34 exceeds the tabulated values of $F0.95;59;118$ = 1.44 at 95% probability and $F0.99;59;118$ = 1.66 at 99% probability by a wide margin.

Hence, the first variance term of 0.000035%^{2} for the ordered set is significantly lower than the variance of 0.000397%^{2} for the randomized set. In other words, the probability exceeds 99% that the degree of associative dependence between ordered measurements is statistically significant and that the set of on-stream measurements for the metal grades of mill feed displays a significant degree of spatial dependence in the sample space of time (in this case, a 12-h shift).

Table 5b lists the first ten variance terms for the ordered set. Plotting these variance terms against the variance of the randomized set and the lower limits of its 95% and 99% confidence ranges gives the sampling variogram in Fig. 2. Not only is the sampling variogram a graphic interpretation of Fisher's F-test, but it also shows where associative dependence dissipates into randomness. The on-stream analyzer still generates valid NDRNs if the variances of randomly distributed and ordered sets are statistically identical, but the process is in a chaotic state that defies control.

The first section of Table 6a, the wet mass, the moisture content and the metal grades of the mill feed, concentrate and tailings are listed. This section also gives the coefficients of variation for these variables and the number of on-stream measurements on which the mean grades are based. The second section of Table 6a gives the dry mass and metal content of mill feed, and the "Simulated variables section" gives the NDRNs for the same variables. The NDRN generator requires 60 rows, a column for each independent variable and a column for each dependent variable under examination. The generator is not displayed.

Table 6b lists the calculated recovery of 91.04%, a simulated recovery of 91.03% and the various precision estimates for the simulated recovery. The variance of $var(Rs)$ = 0.0027%^{2} is obtained by applying the @VARS function to the set of 60 NDRDs for recovery in percent. Derived measures for precision are the standard deviation of $sd/Rs) = \sqrt{0.0027\%^2}$ = 0.0524%, the coefficient of variation of CV = 0.0524*100/91.03 = 0.06%rel, the 95% confidence interval of 95% CI = 0.0524*2 = ± 0.10% and the symmetric 95% confidence range with its lower limit of 95% CRL = 90.94% and upper limit of 95% CRU = 91.15%.

Table 6c lists the mass, moisture and grade terms in metric tons squared (t^2) and in relative percent. The mass term is $var(Me \mid Mw) = Me^2*(CV*0.01)^2 = 200.3^2*(0.5*0.01)^2 = 1.00\ t^2$ or 77.9%rel. Given that the moisture factor is $MF = 1 - 0.01*\%H_2O$, it follows that the moisture term is $var(Me \mid GF) = Me^2*(\%H_2O*CV*0.01)^2/(100 - \%H_2O)^2 = 200.3^2*(1.8*10.0*0.01)^2/(100-1.8)^2 = 0.13\ t^2$ or 10.5%rel. The grade term is $var(Me \mid GF) = Me^2*/CV*0.01)^2/n = 200.3^2*(1.5*0.01)^2/60 = 0.15\ t^2$ or 11.7%rel.

The variance terms in Table 6c show that the variance associated with the measurement of the wet mass of mill feed adds most to the variance of metal content. Evidently, a higher degree of precision for the recovery in percent would require a more precise belt scale. Yet, a coefficient of variation of 0.5% for a belt scale is low indeed. CVs within the range from 1% up to 2% are common, and CVs of more than 2% have been observed.

where

$\sigma^2(y)$ is the population variance of the function,
$\sigma^2(x1)$ is the population variance of the first variable,
$\sigma^2(x2)$ is the population variance of the second variable and
$\sigma^2(xn)$ is the population variance of the nth variable.

In applied statistics, population variances are replaced by sample variances estimated from finite sets of measured values so that variances of $x1, x2, ..., xn$ are $var(x1)$, $var(x2)$, ..., $var(xn)$. The formula for the variance of a general function is often referred to as the additive property of variances. In applied statistics, the additive property is not necessarily commutative, because the difference between two variances itself is a valid variance estimate only if the ratio between the variances is statistically significant. Otherwise, the difference is merely a random number to which mathematical analysis should not be applied. The sum of two or more variances can be partitioned into its components by applying analysis of variance (Merks, 1985).

The transition from probability theory to applied statistics is not merely a matter of changing some symbols, but it also involves scores of tables in which values are listed as a function of degrees of freedom. Such tables reflect that finite sets of measured values determined in samples selected from populations do not instill as much confidence as those elusive population variances would.

Variance of contained metal

The variance of the mass of metal contained in a quantity of mill feed is a function of its wet mass, moisture content and metal grades, and of the variances of these stochastic variables. In formula

$$var(Me) = (MF*GF)^2*var(Mw) + (Mw*GF)^2*var(MF) + (Mw*MF)^2*var(GF) \quad (2)$$

where

Me is the mass of contained metal (t),
Mw is the wet mass (t),
MF is the moisture factor: $1-0.01*\%H_2O$ (dimensionless),
GF is the grade factor: $0.01*\%Me$ (dimensionless),
$var(Me)$ is the variance of contained metal or content,
$var(Mw)$ is the variance of wet mass (t^2),
$var(MF)$ is the variance of moisture factor and
$var(GF)$ is the variance of grade factor.

Multiplying each term with Me^2/Me^2 gives a variance of metal content of $var(Me) = Me^2*[var(Mw)/Mw^2 + var(MF)/MF^2 + var(GF)/GF^2]$. Given that $CV = sd(x)*100/\bar{x}$, it follows that $(CV*0.01)^2 = var(x)/\bar{x}^2$. By implication, the variance formula for the mass of contained metal can also be formulated as $var/Me) = Me^2*\Sigma[(CVk*0.01)^2/nk]$, in which CVk is the coefficient of variation of the k^{th} variable, and n^{th} the number of measured values used to calculate the k^{th} coefficient of variation.

Sampling theory and practice

The validity of simulation models depends critically on the procedures applied to interrogate the stochastic system under examination. The ideal sampling model is the equiprobable model in which the ratio between the mass of a primary increment (Δm) and the mass that it represents (ΔM) is a constant probability. The potential for bias can be addressed by examining the closeness of agreement between theory and practice.

The sampling of slurries at constant time intervals only conforms to the ideal model if a constant volume with a constant percentage of solids is processed and if turbulence prevails near the sampling point. In practice, both the volume and the percentage solids vary so that additional components of uncertainty are added to the primary sampling stage. A bias would occur if periodicities, either in the slurry flow or in its percentage solids, were to synchronize with the sampling interval but the probability is low to encounter this type of bias.

Associative dependence between slurry density and metal grade may cause an insidious bias. The degree of associative dependence between slurry density and grade is not likely to be statistically significant for cyclone overflow and tailing but it may well be significant for concentrate. The magnitude of this sampling bias is difficult to quantify but its sign can be assessed by examining how the various minerals are distributed over the particle size spectrum of the concentrate.

Gy's sampling (Gy, 1977) formula gives an estimate for the composition component of the sampling variance, but its distribution component cannot be estimated on an *a priori* basis. In fact, the very process of estimating the distribution component of the sampling variance tends to reduces it (Heisenberg's uncertainty principle in sampling practice!).

The question of whether Gy's sampling formula can be used to obtain reliable variance estimates for simulation models is intriguing. Because Gy's sampling constant (C) is a function the mineralogical composition factor (c), the liberation factor (l), the particle shape factor (f) and the size range factor (g), it follows that the variance of his constant is $var(C) = C^2*\Sigma[CVk*0.01)^2/nk]$, in which nk is the number of measured values used to calculate CVk.

Duplicate test results for each mineralogical factor are required to determine which factor adds most to the variance of Gy's constant. A cost-effective alternative to determine the variances of metal grades is to select a pair of independent samples, to prepare a test sample of each independent sample and to select and assay a test portion of each test sample.

Visman's sampling theory is based on the additive property of variances. His sampling variance is the sum of the composition variance and the distribution variance (Visman 1962, Merks 1985). Visman replaces Gy's *a priori* premises with *a posteriori* inferences based on measured values determined in primary samples selected under practical conditions. Visman's sampling theory gives a sound scientific basis to practical sampling applications in mineral processing.

Simulation model

The simulation model for the two-product formula requires reliable estimates for the wet mass, moisture content and metal grade of the mill feed, the metal grades of the concentrate and tailings and the variances of these variables. The simulation model shows how the variables and variances interact and determine, not only the masses of concentrate and tailing, their metal contents and the recovery in percent, but also the confidence limits for these variables.

The variance of a mass of contained metal is a function of wet mass, moisture content (mass loss on drying in ISO terminology) and metal grade, and of the variances of the variables (Merks, 1991; International Organization for Standardization, 1996).

The variance of this mass is the sum of three components, namely the mass, moisture and grade terms. Converting these variance terms into percentages shows which variable adds

Table 5a — Variances of randomized and ordered on-stream measurements.		
Statistics for mill feed	**Symbol**	**Value**
Mean grade (dry basis), %	xbar	0.408
Number of on stream measurements	n	60
Variance of randomly distributed set	var(x)	0.000397
Standard deviation	sd(x)	0.0199
Coefficient of variation	CV	4.9
Degrees of freedom	df(r)	59
First variance term for ordered set	var1(x)	0.000035
Standard deviation	sd1(x)	0.0059
Coefficient of variation	CV	1.5
Degrees of freedom	df(o)	118

Table 5b — Sampling variogram for on-stream measurements		
Statistics for mill feed	**Symbol**	**Variance**
Variance of randomly distributed set	var(x)	0.000397
Asymmetric lower limits at:	ACRL	
95% probability	95% ACRL	0.000278
99% probability	99% ACRL	0.000244
Variance terms for ordered set:	var j(x)	
First term	var1(x)	0.000035
Second term	var2(x)	0.000048
Third term	var3(x)	0.000069
Fourth term	var4(x)	0.000087
Fifth term	var5(x)	0.000111
Sixth term	var6(x)	0.000134
Seven term	var7(x)	0.000153
Eighth term	var8(x)	0.000178
Ninth term	var9(x)	0.000186
Tenth term	var10(x)	0.000199

tabulated values of $F0.95;df;\infty$ and $F0.95;\infty;df$, the lower and upper limits of the symmetric 90%-confidence range and the corresponding coefficients of variation.

Table 4 shows that the variances of small sets of test results are extremely imprecise and should, therefore, be interpreted with caution. It also shows the astounding confidence limits that infinite degrees of freedom afford. In fact, any variant of applied statistics unencumbered with degrees of freedom ought to be contemplated with suspicion and trepidation. How to estimate the sum of all variance components in a measurement chain, in an unbiased manner and at affordable cost, is an important element of measurement technology as it applies to mineral-processing plants.

Selecting, preparing and assaying a pair of interleaving or interpenetrating primary samples gives only one degree of freedom, but selecting, preparing and assaying a single primary sample gives none. Interleaving primary samples give the sum of the sampling, sample preparation and analytical variances at the lowest possible cost. A pair of interleaving or interpenetrating primary samples is obtained by storing the set of primary increments (either a mass of solids or a volume of slurry) in two sample receptacles such that one contains the subset of odd-numbered increments and the other the subset of even-numbered increments (Merks 1985).

When based on a set of 60 randomly distributed on-stream data (see Table 5a), the 90% confidence range for a variance estimate of 0.000397%^{2} has a lower limit of 90% $CRU = var(x)/F0.95;59;\infty = 0.000397/1.31 = 0.000303\%^2$ for a coefficient of variation of $CV = \sqrt{var(x)}*100/\bar{x} = \sqrt{0.000303}*100/0.408 = 4.3\%$, and an upper limit of 90% $CRU = var(x)*F0.95;\infty;59 = 0.000397*1.40 = 0.000556\%^2$ for a CV of 5.8%.

If the same variance estimate were based on a pair of interleaving primary samples, it would have only one degree of freedom, and its 90% confidence range would have a lower limit of 90% $CRU = var(x)/F0.95;1;\infty = 0.000397/3.84 = 0.000103\%^2$ for a CV of 2.5% and an upper limit of 90% $CRU = var(x)*F0.95;\infty;1 = 0.000397*254 = 0.1008\%^2$ for a CV of 77.8%.

Generally, a set of n randomly distributed measured values has $df(r) = n - 1$ degrees of freedom, whereas the first variance term for the ordered set has $df(o) = 2(n - 1)$ degrees of freedom. The variance of a set of n randomly distributed measured values is $var(x) = \sum(xi - \bar{x})^2/n - 1$, whereas the variance terms for the ordered set are $varj(x) = \sum(xi - xi + j)^2/2(n - j)$, in which j is the spacing or lag between measured values.

Properties of variances

The variance is the fundamental measure for the intrinsic variability of a stochastic variable within its sample space, and the sum of all random variations associated with selecting primary samples, preparing test samples of primary samples, and selecting and assaying test portions of test samples. How to obtain unbiased variance estimates, either for the entire measurement chain or for a single stage, are elements of metrology — the science of measurement. The advantage of the variance as a measure for precision and variability is that variances are amenable to mathematical analysis. The additive property of variances in particular plays an essential role in the calculation of confidence limits for metal contents.

Many properties of variances are derived from the formula for the variance of a general function. With its roots in probability theory, this formula has a multitude of practical applications in science and engineering that became collectively known as applied statistics. For example, the formulas for the variance of the arithmetic mean, the variance of the weighted average (count, distance, length or mass) and the variance of the mass of contained metal are all derived from the variance of a general function.

Variance of a general function

The formula for the variance of a general function between a dependent variable and a set of independent variables finds its origin in calculus and probability theory (Volk, 1980). The variance of a general function is equal to the sum of n terms, each consisting of a squared partial derivative towards a stochastic variable multiplied by the variance of that variable. In formula

$$\sigma^2(y) = \left(\frac{\partial y}{\partial x1}\right)^2 * \sigma^2(x1) + \left(\frac{\partial y}{\partial x2}\right)^2 * \partial^2(x2) + \dots + \left(\frac{\partial y}{\partial xn}\right)^2 * \sigma^2(xn) \quad (1)$$

Table 3 — Measured and simulated variable and variance.

Measured and simulated variable and variance:

Description	Symbol	Measured	Simulated
Wet mass, t	Mw	50,000	50,469
Variance, t^2	var(Mw)	62,500	83,662

NDRN Generator:

Number	Mw	Number	Mw
1	50,171	6	50,658
2	49,690	7	50,143
3	50,020	8	50,091
4	50,020	9	49,743
5	49,808	10	49,617

sheet template and pressing *F9* gives another set of 144 SURNs and another set of 12 SNRNs.

The next step in Monte Carlo simulations is to convert SNRNs into normally distributed random numbers (NDRNs) for the variable of interest. Process simulations should preferably be based on unbiased variance estimates, but realistic guesstimates are quite useful when applying sensitivity analysis.

Normally distributed random numbers

If a wet mass of 50,000 t of mill feed were estimated with a belt scale and if the precision of the belt scale were equivalent to a coefficient of variation of 0.5%, then the variance of that wet mass would be $var(Mw) = (Mw*CV*0.01)^2 = (50{,}000*0.5*0.01)^2 = 62{,}500\ t^2$ for a standard deviation of $sd(Mw) = \sqrt{62{,}500} = 250$ t, a 95% confidence interval of 95% $CI = sd(Mw)*z0.95 = 250*2 \approx \pm 500$ t or 95% $CI = 500*100/50{,}000 = \pm 1.0$%rel, and a 95% confidence range with a lower limit of 95% $CRL = Mw - 95\%\ CI = 50{,}000 - 500 = 49{,}500$ t and an upper limit of 95% $CRU = Mw + 95\%\ CI = 50{,}000 + 500 = 50{,}500$ t.

Multiplying the standard deviation with $z0.95 = 1.96 \approx 2$ gives a symmetric 95% confidence interval (the area under the standard normal probability distribution between $z = -1.96$ and $z = 1.96$). The factor $z0.95 \approx 2$ implies that the belt scale's CV was estimated from a large set of calibration data in a moving database (Merks and Merks, 1992). If this CV were estimated from a small set of calibration data, the factor $z0.95 \approx 2$ from the normal or z-distribution is replaced with the factor $t0.95;df$ from the t-distribution, in which df is the number of degrees of freedom. For example, a set of ten calibration data gives nine degrees of freedom for a 95% confidence interval of 95% $CI = sd(Mw)*t0.95;9 = 250*2.262 = \pm 566$ t or 95 % $CI = 566*100/50{,}000 = +1.1$%rel

The next step is to generate a normally distributed random number (NDRN) with an expected value of $Mw = 50{,}000$ t and an expected variance of $var(Mw) = 62{,}500\ t^2$. Applying the @VARS function to a suitable set of NDRNs for the wet mass of 50,000 t gives a simulated variance of $var(Mws) = 83{,}662\ t^2$, which, in turn, gives an NDRN of $Mws = Mw + (\sqrt{var(Mws)}*SNRN) = 50{,}000 + (\sqrt{83{,}662}*0.586) = 50{,}169$ t. The simulated NDRN is a member of the infinite set of NDRNs defined by the same expected value and variance. It is based on the SNRN = 6 – 5.414 = 0.586 listed in Table 1.

Table 3 lists, in addition to the measured wet mass of $Mw = 50{,}000$ t and the calculated variance of $var(Mw) = 62{,}500\ t^2$, the same simulated variance of $var(Mws) = 83{,}662\ t^2$, but

Table 4 — 90% confidence limits for variance*.

Degrees of Freedom	Lower limit F0.95;df;∞	90%	CV, %
1	3.84	0.000103	2.5
2	2.99	0.000133	2.8
4	2.37	0.000168	3.2
9	1.88	0.000211	3.6
16	1.64	0.000242	3.8
25	1.51	0.000263	4.0
49	1.36	0.000292	4.2
∞	1.00	0.000397	4.9

Degrees of Freedom	Upper limit F0.95;∞;df	90%	CV, %
1	254	0.100838	77.8
2	19.5	0.007742	21.6
4	5.63	0.002235	11.6
9	2.71	0.001076	8.0
16	2.01	0.000798	6.9
25	1.71	0.000679	6.4
49	1.44	0.000572	5.9
∞	1.00	0.000397	4.9

*See Table 5a: mean and variance randomized set of on-stream data for mill feed.

with another simulated wet mass of Mws = 50,469 t. NDRNs for the other independent variables are generated in the same manner. Setting up the template in Table 3 and pressing "F9" would generate different *Mws* and *var(Mws)*, but each would be valid within its own probabilistic domain.

Monte Carlo simulations should not be based on constant variances. Not only do they underestimate the variances of dependent variables (Merks, 1991), but they are also at variance with the fundamental concept of degrees of freedom in applied statistics. Indeed, the question of how many NDRNs make a suitable set can be solved by examining the relationship between the degrees of freedom associated with variance estimates and their confidence limits.

Confidence limits for variances. The calculation of confidence limits for variance estimates is based on analysis of variance (ANOVA). Tabulated values of the F-distribution at 95% probability give lower limits (90% CRLs) and upper limits (90% CRUs) of symmetric 90% confidence ranges (90% CRs) as a function of the number of degrees of freedom associated with the variance estimate. Similarly, tabulated values of the F-distribution at 99% probability give lower and upper limits of symmetric 98% confidence ranges. The chi square distribution can be used to calculate confidence limits at other probability levels (Volk, 1980).

The formulas for the lower and upper limits of a symmetric 90% confidence range are 90% $CRL = var(x)/F0.95;df;\infty$ and 90% $CRU = var(x)*F0.95;\infty;df$. Table 4 gives the lower and upper limits of the symmetric 90% confidence range for a variance of $var(x) = 0.000397\%^2$, estimated from a set of 60 randomly distributed on-stream measurements (see Table 5a).

The mean grade of $\bar{x} = 0.408\%$ is used to convert the lower and upper limits of this variance estimate into coefficients of variation. Table 4 lists the numbers of degrees of freedom, the

Table 1 — Standard uniform random numbers (SURNs).

Standard uniform random numbers:

Description	Population	Sample
Sum	6.000	5.414
Mean	0.500	0.451
Variance	0.0833	0.0877
Standard deviation	0.2887	0.2962
Coefficient of variation	57.7	65.6

SURN generator:

Number	@RAND	Number	@RAND
1	0.3454	7	0.1833
2	0.1315	8	0.9593
3	0.2091	9	0.5953
4	0.3246	10	0.6656
5	0.0479	11	0.5629
6	0.9029	12	0.4864

games of chance. For example, the function @ROUND(@RAND*6 + 0.5,0) gives the outcome of a single cast with an unbiased die. It is a condition *sine qua non* that the die is unbiased. In fact, random-number generators can be tested for bias based on *a priori* probabilities for games of chance such as casting dice, drawing cards and playing roulette.

The properties of the standard normal probability distribution are explored based on the sum of a set of 12 SURNs and of the expected mean and variance of such sets. The population mean and variance of the standard uniform probability distribution are $\mu = 0.5$ and $\sigma^2 = 1/12$, which implies that the sum of 12 SURNS is $\sum xn \rightarrow 6$ for $n \rightarrow \infty$. Sample statistics differ from population parameters due to randomly distributed variations intrinsic to the stochastic system under examination and due to uncertainties associated with the stages of the measurement chain applied to interrogate the system.

The third column of the first section of Table 1 gives the expected values for the sum $\sum xn \rightarrow 6$ for $n \rightarrow \infty$, the mean ($\mu = 0.5$), the variance ($\sigma^2 = 1/12$ or 0.0833), the standard deviation ($s = \sqrt{[1/12]}$ or 0.2887) and the coefficient of variation ($CV = \sigma*100/\mu = 57.7\%$)). The last column in the first section of Table 1 displays the measured values calculated from a sample of 12 SURNs. The second section of Table 1 gives each SURN in the set of 12 on which the sample statistics in the last column of the first part of Table 1 are based.

When set up in a spreadsheet template, the simulation model in Table 1 gives a different set of sample statistics in the last column whenever *F9* is pressed. Each of the statistics is valid within the probabilistic domain defined by the properties of the standard uniform probability distribution.

The variance of $\sigma^2 = 1/12$ or 0.0833 for the standard uniform distribution shows why sets of 12 SURNs are such an effective choice for obtaining standard normally distributed random numbers. After all, any other number but twelve requires that an additional factor be applied to obtain an estimate for the variance of the standard uniform probability distribution.

The difference of 0.586 between the expected sum of 6.000 and the observed sum of 5.414 is the intermediate output of the Monte Carlo simulation. It is a member of the infinite set of standard normally distributed random numbers (SNRNs) and the fundamental building block for process simulations with ubiquitous spreadsheet software.

Table 2 — Standard normal random numbers (SNRNs).

Standard normal random numbers

6.467	6.419	3.967	7.385
5.989	5.408	5.592	6.761
7.401	4.988	5.910	5.235

SRN generator

0.467	0.419	-2.033	1.385
-0.011	-0.592	-0.408	0.761
1.401	-1.012	-0.090	-0.765

Standard normal distribution

The standard normal probability distribution has a zero mean ($\mu = 0$) and a unit variance ($\sigma^2 = 1$). Each independent stochastic variable in the set that defines the stochastic system requires a single randomly selected value of the standard normal distribution, and, thus, a set of 12 randomly selected values of the standard uniform distribution (Yakowitz 1977; Graybeal and Pooch 1980).

It is of critical importance that each standard normal random number (SNRN) obtained by summating a set of 12 standard uniform random numbers (SURNs) and by deducting their expected sum of six is indeed a member of the infinite set of SNRNs. The @RAND function is designed to meet this criterion. Its performance was tested under conditions that were more rigorous than what Monte Carlo simulations of mineral-processing plants demand.

Even though the expected sum of 12 SURNs is six, random variations associated with the sample-selection stage cause the observed sum to be either lower or higher than six, but never exactly identical to the integer six. The term random errors should not be used in reference to randomly distributed variations (random variations) for which the expected sum is zero. After all, Karl Pearson, who introduced the coefficient of variation as a measure for variability and precision long ago, did not call it "coefficient of error."

The coefficient of variation (CV) in percent is defined as the standard deviation of a measured value as a percentage of that value or of the mean of a set. Another advantage of working with CVs in simulations models is that it simplifies the calculation of variances of dependent variables such as the mass of metal contained in mill feed.

When a measured value is biased, uncertainties associated with measuring the variable of interest can be partitioned into random variations and a bias for one or more stages of the measurement chain. A case can be made that the use of the term "error," even with proper adjectives such as "sampling error" or "systematic error," be restricted. How to partition the sum of the sampling, preparation and analytical variances into its components, how to test for bias, how to test for normality and how to test for statistical significance, the degrees of kurtosis and skewness are subjects beyond the scope of this paper.

The bottom part of Table 2 lists a set of 12 SNRNs, each generated by summating a set of 12 SURNs (see top part of Table 2) and deducting the expected value of six. Not shown in Table 2, but contained in the template, are 144 SURNs from which 12 SNRNs are calculated. Setting up the spread-

Table 7c indicates that the variance associated with the measurement of the metal grade of mill feed adds most to the variance of metal content. Thus, pairs of interleaving samples give a significantly lower degree of precision than sets of 60 on-stream measurements do. Modules are available to routinely select pairs of interleaving samples, either from slurry flows at low capacity plants or from primary sample flows at high capacity plants.

Summary

Monte Carlo simulation is an intuitive and powerful tool to obtain precision estimates for dependent variables, such as the recoveries at mineral processing plants and the metal contents of mill feed, concentrate and tailings. The proper application of process simulations with spreadsheet software demands some understanding of the properties of variances and of the limitations that the concept of degrees of freedom imposes on the confidence limits for variances.

For a given set of independent variables, the 95% confidence limits for a recovery of 91.0% range from a low of 90.9% to a high of 91.2% when the metal grades of mill feed, concentrate and tailings are measured with an on-stream analyzer. Under those conditions, the variance of the wet mass of mill feed adds most to the variance of metal content. However, for the same set of independent variables and variances, the confidence limits range from a low of 90.5% to a high of 91.6% when the metal grades of mill feed, concentrate and tailings are measured in pairs of interleaving samples. In the latter case, the variance of the grade of mill feed adds most to the variance of content.

Unlike deterministic systems of linear equations, which adequately describe dynamic stochastic systems (such as mineral processing plants) only during infinitesimal instants, the proposed methodology is based on the properties of the variances of grades and contents of mill feed, concentrate and tailings. Monte Carlo simulations for two- and three-product formulas are simple to set up with spreadsheet software.

Monte Carlo simulations also make it simple to solve "what-if" questions by replacing a suspected value of a stochastic variable with a value corrected for the presence of bias. Yet, the real power of Monte Carlo simulations is in generating confidence limits for dependent variables such as recoveries and metal contents of concentrate and tailing.

References

Box, G.E.P., and Jenkins, G.M., 1976, *Time Series Analysis*, Holden-Day, San Francisco.

Graybeal, W., and Pooch, U.W., 1980, *Simulation: Principles and Methods*, Winthrop Publishers, Cambridge, MA.

Gy, P.M., 1977, *Sampling of Particulate Matter, Theory and Practice*, Elsevier Scientific Publishing Company, Amsterdam.

International Organization of Standards (ISO), 1996, "Determination of Mass of Contained Metal in the Lot," DIS 13543, Geneva, Switzerland.

Merks, J.W., 1985, *Sampling and weighing of Bulk Solids*, Trans Tech Publications, Clausthall-Zellerfeld, Germany.

Merks, J.W., 1991, "Simulation models for mineral processing plants," *CIM Bulletin*, Volume 84, No 965, Sept. 24-24.

Merks, J.W., and Merks, E.A.T, 1992, "Precision and Bias for Mass Measurement Techniques," Matrix Consultants, Vancouver, British Columbia, Canada.

Visman, J., 1962, "Towards a Common Basis for the Sampling of Materials," CANMET Research Report R93, July.

Volk, W., 1980, *Applied Statistics for Engineers*, Krieger Publishing Co., Huntingdon, NY.

Yakowitz, S.J., 1977, *Computational Probability and Simulation*, Addison-Wesley Publishing Company, Reading, MA.