Some practical aspects of the use of lognormal models for confidence limits and block distributions in South African gold mines

by D.G. Krige*

Synopsis

For the purpose of determining confidence limits for ore grade estimates, an extensive simulated database was used, which included 1728 ore blocks with 'actual' grades and the corresponding estimates based on ordinary and simple kriging. The practical results show that conditional unbiasedness is essential and that acceptable confidence limits can then be set using Normal theory for the simulated error distribution based on the logarithms of the ratios of 'actuals'/estimates with a variance related to the kriging variance. The mean of the data used to prepare the variogram and the estimated block grade are to be used properly in this approach. A practical analysis was also done of the actual point and block data from a very large mined out Carbon Leader area. The distributions of the actual grades for two levels of block sizes with different levels of variability were fitted using three models, ie. two three and compound lognormal. The trends of tonnages, grades and relative profits for a range of pay limits, as well as the usual tonnage-grade curves, show a practical insensitivity for the choice between the three lognormal models considered.

Introduction

Distribution models for observed grade distributions in the South African mining environment have received attention repeatedly, and deservedly so, since the inception of geostatistics more than 50 years ago, and the process is still continuing. Such models are essential for point values and for larger support values such as ore blocks (usually *Selective Mining Units—SMUs*) as well as for estimator models and the related error distributions of, and confidence limits for, grade estimates for ore blocks and for larger areas such as mine sections still to be exploited.

For *point distributions* it is essential to compare alternative distribution models for large bases of actual data, and in South Africa this led to the use of mainly three lognormal models, i.e. *the two, three and compound lognormal distributions —2PLN, 3PLN and CLN* (Sichel, 1947; Krige, 1960; and Sichel *et al.* 1992). Experience has shown that before any model is chosen and used, the subdivision of the orebody into geologically homogeneous sub-populations is essential. Usually the particular model best suited to the point data has also been accepted for the corresponding *block distributions* but this practice will be verified for the examples used in this paper.

As far as *error distribution* models are concerned, Sichel covered the small sample theory for random sampling from the 2PLN (Sichel, 1952) with application also to the 3PLN. Some aspects of error distributions for kriged estimates were dealt with by Dowd (1988). However, the cases dealt with by Dowd covered kriging estimates which are seldom of much practical use, i.e. estimates of point grades using very limited search routines. These are subject to unavoidable conditional biases and the effects of these were not dealt with. Parkin *et al.* (1990) compared five methods for estimating confidence intervals for *random* samples drawn from a 2PLN distribution but, as for Sichel's limits referred to above, these will not be applicable to kriging estimates which cater for data with a significant spatial structure and where the error variance is estimated by the *kriging variance* (KV). An attempt will be made in this paper to address some of the aspects of error distributions and confidence limits based on the KV indicated for kriging estimates.

Error distributions for kriged estimates

In South African gold mines ore reserve blocks are generally valued using a variogram of untransformed grades and *ordinary kriging (OK) or simple kriging (SK)*. This approach is preferred to the use of lognormal kriging which, although theoretically more efficient, can introduce significant biases mainly during

* *Private Consultant, Florida Hills, Gauteng, South Africa.*

 The paper was presented at the SAIMM Symposium: Applications of Computers and Operations Research in the Minerals Industries, *14–16 May 2003.*

the back-transformation process (Sinclair and Blackwell, 2002). The OK or SK block estimates incorporate the corresponding KVs and their use in estimating the corresponding theoretical confidence levels will be investigated. If properly interpreted, such confidence levels can assist in the classification of resources and reserves into the appropriate categories. In many circles, however, the use of the KV has been discredited and is still viewed by some as misleading and inappropriate, an issue which requires clarification.

For this purpose, a massive 2PLN database of 13 824 closely spaced point values with a realistic spatial structure, as simulated for a paper by Krige and Assibey-Bonsu (2000), was used. On subdivision into 1 728 ore blocks, each block was kriged on the data inside the block as well as sufficient surrounding data to ensure a negligible level of error variances, and thus to provide a set of 'actual' block values for comparisons with corresponding block estimates. For these block estimates, a set of 216 point values was selected on a regular grid from the 13 824 available values, and these data were used to prepare OK estimates with a limited search as well as SK estimates for each block together with the corresponding KVs.

The errors between these estimates and the 'actuals' can be defined either as:

- *differences*, i.e. 'actuals' less estimates, or
- *ratios*, i.e. 'actuals'/estimates.

The corresponding error distributions were analysed in total, i.e. for 1 728 blocks and also for the more homogeneous set of 1 331 blocks remaining after peeling off the peripheral blocks where the available data were more limited. The results were analysed after subdivision into ten grade estimate categories from the lowest estimates, 0 to 10 per cent cumulative frequencies, to the highest estimates, 90 to 100 per cent cumulative frequencies. Such a split is essential to ensure that the confidence levels indicated do not vary significantly with variations in the block grade levels as estimated.

Errors defined as differences

Figure 1 shows the error distributions based on 'actuals' less estimates for the OK estimates for several of the ten per cent grade categories and for all 1 728 estimates, as well as the theoretical Normal distribution based on the indicated average KV level. It is obvious that the OK distributions for individual grade categories depart radically from each other and from the overall distribution, as well as from the theoretical Normal error distribution. Figure 2 shows the same analyses for the SK estimates; the departures are somewhat less than for the OK estimates, but still unacceptable for the purpose of applying a consistent model for confidence intervals.

Errors defined as ratios

The alternative model based on ratios, i.e. 'actuals'/estimates, should lead to more logical confidence limits expressed as percentage, and not absolute, deviations from the estimates. For the lognormal distributions of the basic block 'actuals' and estimates, the log-transformed distributions will be Normal as well as the differences between the log values. The log-differences correspond to the ratios of the untransformed values and thus the distribution of the ratios will also be lognormal with a variance defined as:

$$\sigma_k^2 = m^2\left[\exp\left(\sigma_{\ln}^2\right) - 1\right], \qquad [1]$$

where σ_k^2 = untransformed error variance = KV
m = *mean of data used for the variogram*
σ_{ln}^2 = equivalent logvariance of errors
= variance of deviations of 'actuals' from estimates on a log basis.

The 'logvariance' can thus be solved for and confidence limits estimated on Normal theory as follows:

$$\text{Central 90 per cent limits} = \exp\left[\ln(x) - \sigma_{\ln}^2/2 + -1.645\left(\sigma_{\ln}\right)\right] \qquad [2]$$

where x = the block estimate, and σ_{ln}^2 = natural logarithmic variance.

Note that in formula [1] 'm' is not the estimated block grade but the mean of the data on which the variogram is based. The KV (or error variance) does not depend on the grades accessed for the kriging of a block, but only on the data *configuration* and the relevant covariances for the data and the ore block. The covariances, in turn, depend on the parameters of the spatial structure as modelled by the

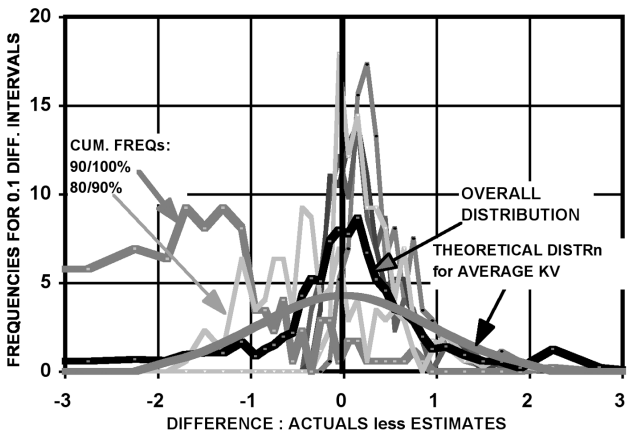

Figure 1—Showing the error distributions as 'actuals' minus OK estimates, for some of the ten grade categories; also the 'actual' and theoretical overall distributions

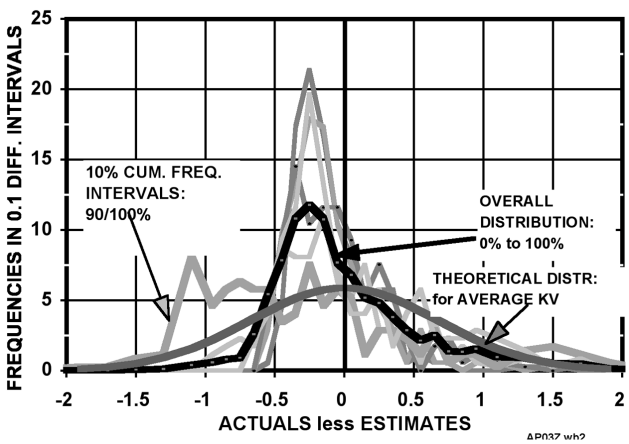

Figure 2—Showing the error distributions as 'actuals' minus SK estimates for some of the ten grade categories; also the 'actual' and theoretical overall distributions

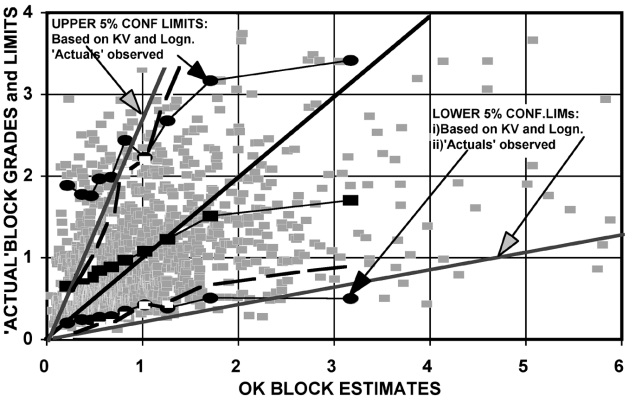

Figure3—Showing the correlation between OK block estimates and 'actuals' as well as the 'actual' and theoretical 90 per cent central confidence limits in ten grade categories

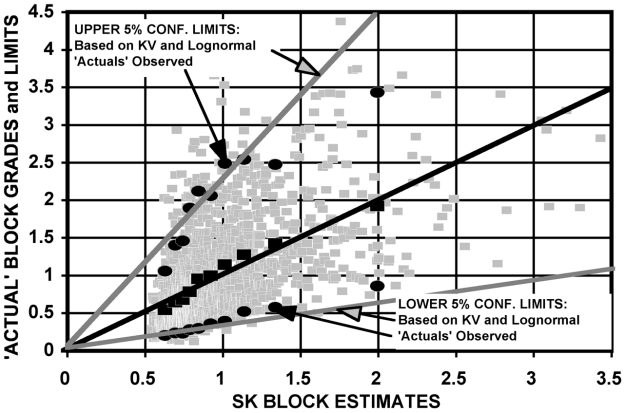

Figure 4—Showing the correlation between SK block estimates and 'actuals' as well as the 'actual' and theoretical 90 per cent central confidence limits in ten grade categories

variogram and specifically on the total sill (i.e the population variance) which, owing to the proportional effect, will vary with the overall mean grade of the data (see Formula [1]). *The KV is thus directly linked to this mean grade and not to the block estimate*. The validity of this approach will be tested in some detail on the data set mentioned above. The overall grade distributions for the basic point data, the 'actual' block grades and the SK block estimates, can all be accepted as 2PLN. The distribution of the OK ratios of 'actuals'/estimates are also close to 2PLN. These lognormal patterns were confirmed via log-probability plots (not shown) and supported the so-called 'permanence' of the lognormal.

A practical analysis of the block estimates was done on the correlation graphs of the OK and SK estimates versus the corresponding 'actual' block values as shown by grey squares on Figures 3 and 4 respectively. The regression trends observed, i.e. the averages of the 'actual' grades for the conditional distributions in a range of ten estimate categories, are shown as filled-in squares, as well as the ideal linear regression with a slope of unity. Within each of the ten grade categories used, the observed 90 per cent central confidence limits for the 'actual' distributions are also shown as filled-in circles. Also shown as straight grey lines are the corresponding 90 per cent theoretical limits for the conditional 'actual' (Y) distributions based on the lognormal model for the ratios of 'actuals'/estimates with an untransformed variance accepted at the recorded average KVs for the OK and SK estimates respectively, using formula [2] on the previous page.

OK estimates

The deviations of the OK regression trend from a slope of unity shown on Figure 3 clearly stresses the significant conditional biases present owing to the limited data search used. The same trend of biases also appears in the actual 90 per cent central confidence limits (black circles). When these conditional biases are removed, these confidence limits change to the heavy broken lines. The serious problems which can be encountered when confidence limits for OK estimates are not based on routines with an adequate data search are evident and probably contributed historically to the scepticism of the KV as a useful measure of uncertainty.

SK estimates

Figure 4 shows the corresponding results for the SK block estimates. The regression trend is close to unity and the confidence limits agree reasonably well with the theoretical limits as would be calculated using the logarithmic equivalent of the KV and the lognormal distribution model as covered by Formulae [1] and [2] on the previous page. Bearing in mind the high level of the data variability and the wide average 90 per cent central confidence levels of about –70 per cent and +125 per cent, this ratio model, although not perfect in this example, is recommended for use in cases where the two- or three-parameter lognormal models apply. In the latter case, the estimate 'x' is replaced by '$x + b$' (with b = 3rd parameter) and 'b' is deducted from the limits as calculated.

Distributions for block grade estimates

Effects of change of support

In changing from point values to block grades, actual or estimated, the change of support results in a significant reduction in the variance level, as well as a reduction in the practical differences between the shapes of distribution patterns for the three lognormal models under consideration. Apart from the mean and variance, the 2PLN has no other parameter, the 3PLN has a third 'b' parameter, and the CLN has a third and a fourth parameter for the logarithmic skewness and kurtosis respectively.

For the purpose of comparing the 2PLN and 3PLN models with the CLN, a detailed analysis was done on actual data from a Carbon Leader mined-out area which is also referred to by Assibey-Bonsu and Krige (2003). The 2PLN proved unsuitable for the point values and the 3PLN required a 'b' parameter of 170 cm g/t. This third parameter did not change significantly for the corresponding block distribution. Figure 5 shows the log-probability plots of the point and 50 x 50 m block values for the 'actuals' and the fitted 3PLN and CLN models. The reduction in the differences between these two models for the block values is evident. Changes in the critical parameters for various block sizes are shown in Table I.

From Table I it is evident that for these data the critical three parameters, i.e. the logarithmic variance, skewness and kurtosis decrease significantly as the support size increases

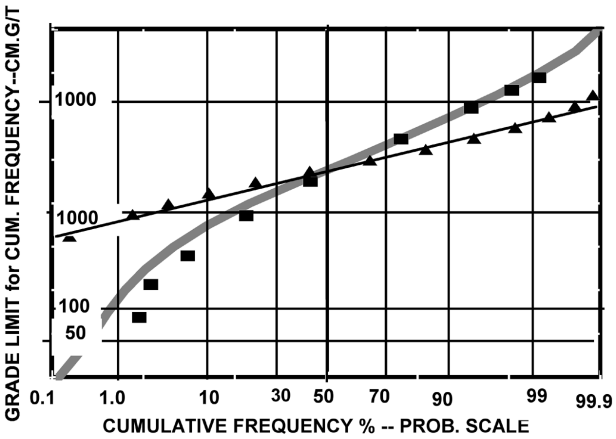

Figure 5—Showing the CLN model for 'actual' point values (grey curve), the corresponding 3PLN model with *b* = 170 (black squares) and the 'actual' grade distribution for 50 × 50 m blocks (black triangles)

Table I

Points and mining blocks comparison

Support	ln mean	ln variance	ln skewness	ln kurtosis	Data number
Points	7.715	1.035	-0.670	5.430	59 951
20 x 20 m blocks	8.008	0.323	-0.221	5.640	4 965
50 x 50 m blocks	8.058	0.202	-0.162	4.189	1 315
100 x 100 m blocks	8.069	0.139	-0.044	3.140	403
200 x 200 m blocks	8.106	0.098	+0.021	3.021	117

and effectively the latter two parameters reach Normal levels at the 200 × 200 m block size. Similar trends were observed by Krige and Dohm (1994) for a VCR distribution but not conclusively for a Vaal Reef distribution. The overall effects of these changes for the Carbon Leader in the present case will be studied in some detail.

The observed 'actual' distributions in Table I for the 20 × 20 m and 50 × 50 m block sizes show untransformed *coefficients of variation (C of V)* of 0.65 and 0.47 respectively, and will be used to compare the effects of using the 2PLN, the 3PLN and the CLN models for these distributions in the paragraphs that follow.

Two-parameter lognormal

As covered in the example used previously, it is generally accepted that where the point data follow a 2PLN model, the same model will also apply to corresponding distributions for block grade estimates. It should be noted that the latter distributions have much lower variance parameters than those for point data, particularly for deep level mining areas where the SMUs will be relatively large and block estimates will involve the significant smoothing effect resulting from the essential conditional unbiasedness. In any particular application where individual blocks are valued on an SMU basis (usually for measured and indicated resources) the kriged results provide estimates of the KV and *dispersion variance (DV)* for each block. If the data used for the block valuations are in their final form, i.e. if no further data will be forthcoming before the final selection of payable blocks is made, the distribution of the individual block estimates will form the basis for the tonnage-grade curve to be used for planning purposes. No mathematical model is then necessary but if a model is required for the observed distribution, the 2PLN model should normally be suitable. For the two 'actual' distributions of the 20 m and 50 m blocks in Table I, this model also proved suitable even though the point distribution was distinctly CLN (see Figure 5). This is discussed in the next section.

For ore resource blocks where the data available are more limited than that which will be available later for the corresponding measured resources, the early SMU estimates will be smoothed and some post-processing is required (Assibey-Bonsu and Krige, 1999). The use of the 2PLN model is also indicated for the post-processed grade-tonnage curve(s) in the corresponding category(ies) of resources.

Three-parameter lognormal

Where the point data follow this model, it is generally also accepted that the block distributions follow the same pattern with the same third parameter '*b*'. However, in practice the third parameter could change for the block distribution and it is thus instructive to observe the sensitivity of the model to any such change. This is shown in Figures 6 to 8 where the block distributions are based on the 20 m and 50 m block supports from Table I and cover the relevant two levels of variability, i.e.

C of V = 0.47, the lower variability applicable to 50 m blocks, and

a C of V = 0.65, i.e. the higher variability for the 20 m blocks; also

for *b* = 0, i.e. the 2PLN, and for an upper limit of

b = 30% of the overall mean grade, i.e. the 3PLN.

The two variability levels are those for the distributions of 'actual' grades and are thus higher than can be expected from estimates. For a range of *pay limits* (PL) the results of the fitted 2PLN and 3PLN models are compared in Figures 6 and 7 with each other and with the 'actual' block distributions (grey curves) for the tonnage percentages and the corresponding grades above each PL. The 'relative profits' are also shown, where

$$\text{relative profit} = \%\ \text{tons} * (\text{grade above PL} - \text{PL}). \quad [3]$$

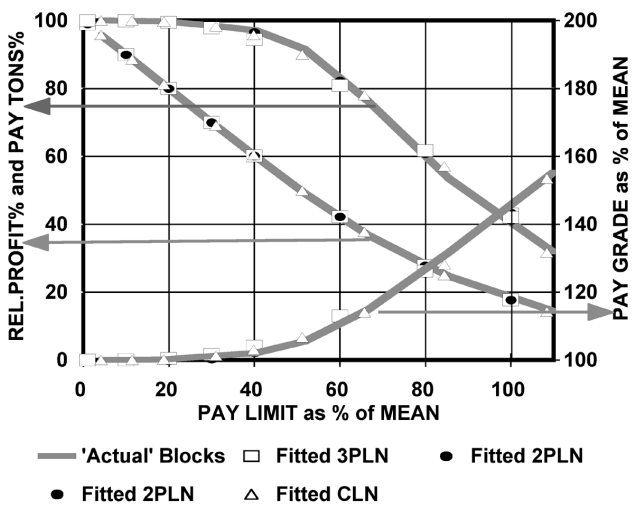

Figure 6—Showing the 'actual' pay tonnage percentages, pay grades and profit for 50 × 50 m blocks with a grade coefficient of variation of 0.47 and the fitted three lognormal models with the same C of V

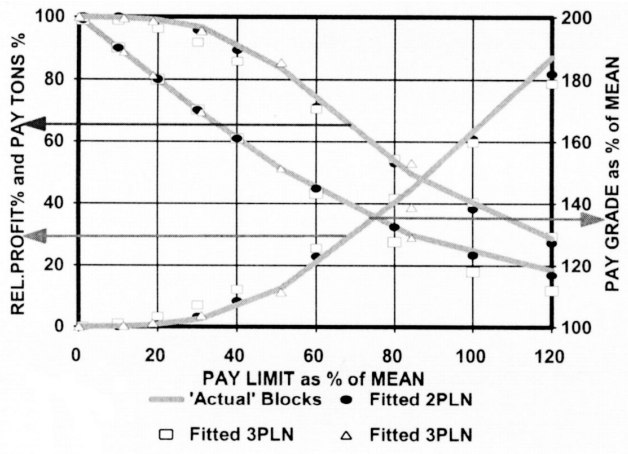

Figure 7—Showing the pay tonnage percentages, pay grades and profits for 'actual' 50 × 50 m blocks with a grade coefficient of variation of 0.65 and the three fitted lognormal models with the same C of V

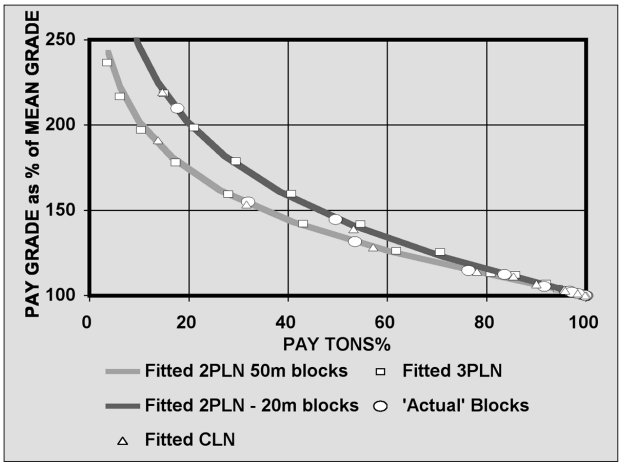

Figure 8—Showing the tonnage grade curves for the three lognormal models

Figure 8 shows the usual corresponding tonnage-grade plots.

It is clear that within the range of 'actual' grade variability levels covered, the results are not sensitive to variations in the '*b*' value between zero and 30%, i.e. between the 2PLN and 3PLN models respectively. This conclusion will apply even more forcibly to the usual case where block 'actuals' are not known and modelling is done on SMU estimates. Note also that the results for the 50 m 'actuals' are expected to be closer to the position expected for block estimates in deep level gold mines.

Compound lognormal model

Apart from the fitted 2PLN and 3PLN models, Figures 6 to 8 also show the results of fitting the CLN model to the 'actuals'. It is evident that the three fitted models agree well with the 'actuals', particularly for the 50 m block case. It is thus concluded that the modelling of these block distributions is not sensitive to the choice between the three lognormal models, particularly within the practical limits of selective mining percentages. For deep level areas selective mining decisions will be made on the basis of larger SMUs, the KVs will be higher, and lower variability levels can be expected for block estimates. This will result in the choice of the relevant model being even less significant than in the practical examples dealt with.

Conclusions

The following conclusions are drawn from these analyses:

- In estimating confidence limits for grade estimates on a lognormal model the mean of the data used for the variogram modelling must feature correctly as covered in a previous section, the limits should be on a per cent basis and will be skew
- Any distribution model has its own limitations and can never be accepted as a perfect and final representation of the behaviour of the variable concerned
- Where alternative models present themselves some practical tests of sensitivities and applicability are required. For SMU block distributions in deep level gold mines the choice of a specific model does not appear to be critical
- The best validation of a model, apart from any theoretical considerations, is to be obtained from practical follow-up tests wherever possible.

Acknowledgements

The author expresses his appreciation for the assistance provided by Dr W. Assibey-Bonsu in handling of the Compound Lognormal Model and for permission by Gold Fields Limited for the use of the relevant database.

References

ASSIBEY-BONSU, W. and KRIGE, D.G. Use of direct and indirect distributions of selective mining units for estimation of recoverable resource/reserves for new mining projects. *28th APCOM Symposium*, Colorado School of Mines, Golden,. Co, Oct. 1999. pp. 239–248.

ASSIBEY-BONSU, W. and KRIGE, D.G. An analysis of the practical and economic implications of systematic underground drilling in deep South African gold mines. *31st APCOM Symposium*, Cape Town. South Africa, May 2003.

DOWD, P.A. Some observations on confidence intervals and kriging errors, *Third International Geostatistics Congress*, Avignon, Sept. 5-9, 1988, vol. II. pp. 124–149.

KRIGE, D.G. On the departure of ore value distributions from the lognormal model in South African gold mines. *J. S. Afr. Inst. Min. Metall.*, 61:231–244, 333-338, 401 and 62:63–64. 1960.

KRIGE, D.G. and DOHM, C.E. The role of massive grade data bases in geostatistical applications in South African gold mines. *Geostatistics for the next century.* Kluwer Academic Publishers. 1994. pp 46–54.

KRIGE, D.G. and ASSIBEY-BONSU, W. Limitations in accepting repeated simulations to measure the uncertainty in recoverable resource estimates particularly for sections of an ore deposit. *6th International Geostatistics Congress.* South Africa. April 2000.

PARKIN, T.B., CHESTYER, S.T., and ROBINSON, J.A. Calculating confidence intervals for the mean of a lognormally distributed variable. *Soil Sciences Society of America Journal*, vol. 54, no. 2. March–April 1990.

SICHEL, H.S. An experimental and theoretical investigation of bias error in mine sampling with special reference to narrow gold reefs. *Trans. Inst. Min. Metall.*, London, vol. 56. 1947. pp. 403–473.

SICHEL, H.S. New methods in the statistical evaluation of mine sampling data. *Bull. Inst. Min Metall.*, London, vol. 56, 1952. pp. 403–473

SICHEL, H.S., KLEINGELD, W., and ASSIBEY-BONSU, W. A comparative study of the frequency distribution models for use in ore valuation. *J. S. Afr. Inst. Min. Metall.* April, 1992. pp. 91–99.

SINCLAIR, A.J. and BLACKWELL, G.H. *Applied Mineral Inventory Estimation*. Cambridge University Press. 2002. ◆