Paper for the UNISA conference in honour of Prof.F.Steffens, May 2002

A CLASSICAL STATISTICAL PERSPECTIVE ON THE BIRTH AND SOME CRITICAL DEVELOPMENTS OF MINING GEOSTATISTICS

D.G.KRIGE, CONSULTANT

ABSTRACT

The critical basic problems of ore valuation in South Africa which led to the birth and the subsequent world-wide development of geostatistics since the middle of the last century, are reviewed and placed in a classical statistical perspective. The review covers the fundamental aspects of the available distributional models; effects of changes of support; estimating the mean grade of a new project from borehole results; grade estimations for individual ore blocks based on the spatial stucture as modeled, on peripheral ore samples plus other data in the neighbourhood; confidence limits for these estimates; the dangers of conditional biases; and the introduction, where practical, of Bayesian principles

1. INTRODUCTION

Geostatistics, specifically for mining, originated in South Africa as a branch of Applied Statistics (Krige, 1951) some 50 years ago during a study of practical problems inherent in the valuation of gold ores in the Witwatersrand geological basin. This basin is a unique deposit in the world consisting of layered sediments and volcanics with a total thickness of some 8 km. It covers an area stretching from the Far East Rand to the Southern Orange Free State and measuring some 300 by 200 km. Gold bearing conglomerate bands, or reefs, occur at intervals within these sediments and are remarkably consistent and continuous.

The distribution of gold within a reef shows a high variability even within small areas and this variability increases within larger areas. While many differences exist within and between reefs, there are also similarities which provided a fruitful field for statistical applications. A further crucial factor in this regard was the extensive physical sampling of these reefs during exploration and mining operations leading to the accumulation of massive data records dating back to the end of the 19th century.

2. GOLD FREQUENCY DISTRIBUTIONS

The gold reefs are relatively thin sheets of ore which can, from a practical point of view, generally be seen as two-dimensional occurrences. Ore samples are obtained underground at regular intervals along exposed ore faces as well as borehole cores from holes drilled during the exploration stage. The gold grade, expressed as a cm.g/t figure, measures the gold content per unit area of reef, e.g. 36.6 cm.g/t = 1 gram per sq. metre. In a particular mining area, the statistical ‘population’ is, therefore, represented by all the cm.g/t values from the almost infinite number of physical samples which can theoretically be obtained within the area. The statistical ‘sample’ would be the grades of the physical samples actually available for analysis.

The first attempts to analyse large numbers of gold values statistically date back to 1919, and 1929. The positively skew nature of the frequency distributions was observed, but it was only in 1947, (Sichel, 1947), that the log-normal model was proposed. This 2-parameter model is strictly applicable only if the log-transformed values follow a Normal distribution exactly.

With the subsequent development of geostatistical applications it was shown (Krige, 1960) that, in fact, most of the log-transformed gold distributions were significantly negatively skew and that the 3-parameter log-normal, which required an additive constant before transformation, was a more suitable model. This model is still in extensive use throughout the gold mining industry. More recently, and to cater for gold distributions which cannot be covered properly by the 3-parameter model Sichel proposed the Compound Log-normal Model (Sichel H.S.,1972; Sichel, et al, 1992),, originally developed for diamond distributions.

For ore valuation applications it is essential to study not only the frequency distribution patterns of the gold grades of individual small ore samples ('point' grades) but also those for much larger ore units such as blocks or ore. In geostatistics this principle is referred to as a 'change of support' where the 'point support' is the physical ore sample assayed for its gold grade, and its grade is a member of the corresponding statistical population. The ore block constitutes a larger 'support' and is further qualified by the how the corresponding block grades are defined, e.g. the actual average grades, or estimates thereof based on the available data and the valuation technique used.

For a specific gold area, there would thus be both a 'point' distribution, or population of grades, as well as a series of block grade distributions covering the actual grades and a distribution of block estimates for each level of data concentration and each valuation technique used. In the absence of any global biases in the assaying and in the block estimates, all these distributions will have the same global mean grade but with a lower variance level for the actual block grades and even lower variance levels for the block estimates. This important aspect is discussed in some detail in par. 5 below. In addition there is also a 'point' distribution within each block. The definition of such block distributions is critical in the mining scenario where selectivity is essential so as to eliminate the uneconomic mining of low grade blocks. Even in cases where, for practical reasons, no selectivity can be practised, it is still necessary to estimate individual ore block grades as well as global mean grades for virgin areas (using borehole 'point' grades). Block grades as estimated directly, or their distribution pattern as inferred, e.g. from borehole grades, are essential for ore reserve estimations, selective mining, mine planning, and grade control.

3. THE FIRST GEOSTATISTICAL ANALYSES

The birth of geostatistics is closely linked to South Africa and more specifically to gold mining in the Witwatersrand basin. These gold mining operations called for intensive and regular sampling of development ends and of all advancing stope faces. This set the scene for the accumulation of extensive data sets conducive to geostatistical analyses. Extensive follow-up data from sampling inside ore reserve blocks provided a massive data base for the application of classical statistical correlation techniques.

The first geostatistical techniques originated from the application of classical correlation and regression techniques to data from Witwatersrand gold mines (Krige, 1951). Orthodox block estimates were based on the averages of all available physical samples around the peripheries or part peripheries of ore blocks. Using the log-normal model, these were correlated with the follow-up averages of much larger numbers of samples obtained from stope faces advancing through the blocks as these were mined out, i.e. with the 'actual' internal block grades. The correlation and the resultant regression trend, explained clearly the reason for the disturbing feature observed from the early days of the Witwatersrand, namely, the under and over valuation of blocks valued as low and high grade respectively. The results observed were satisfactorily reproduced theoretically by simulating an 'actual' log-normal block distribution with appropriate global mean and log-variance, superimposing on this a log-error variance for each block estimate and correlating the block estimates with the 'actual' block grades.

The persistent bias errors in the orthodox estimates complicated grade control on the mines and gave misleading grade tonnage trends over a range of cut off grades. No rational explanation for this feature of conditional biases was forthcoming and no action was taken until 1951, when, as explained above, it was shown to be a straightforward and unavoidable regression effect present when estimates, subject to error, are correlated with the corresponding 'actual' values. The practical solution implemented on many mines in the 1950's and 1960's, was to value blocks on a regression basis which, effectively, corresponded to estimates based on weighted averages of the peripheral averages and of the mean grade of the mine or section concerned. These regressed estimates can, therefore, be regarded as the first application of what later became known as *Kriging,* more specifically S*imple or Elementary Kriging*. This technique is equivalent to the classical statistical regression of estimates.

This estimator soon developed (Krige, et al,1963), to a multidimensional correlation and regression estimator. In order to reduce the number of individual data point values to be used in the inversion of the data matrix, these were averaged, or regularised, into a much smaller number of data blocks around the ore block to be estimated. The levels of correlation and co-variance between such data block grades, and ore blocks were analysed in adjacent mined out parts of the mine section concerned where a detailed spread of data was available. In this way data 'outside' the ore block, catered for previously by the population mean, were now allowed to carry individual weights. Thus data blocks close to the ore block and with a higher correlation level, could carry more weight than data blocks further away with a lower correlation level. This simple kriging technique still uses the mean grade of the section, but is more efficient than the earlier elementary kriging and is, in effect*, a typical classical multidimensional regression estimator*.

4. THE SPATIAL STRUCTURE

This early work clearly indicated the presence of a spatial structure in the gold grades. The correlation between peripheral block estimates and the follow-up inside grades as well as the increase in variability of grades with increases in the physical size of the area covered by the data (Krige, 1952), provided direct evidence of a spatial structure. In addition, early work also showed that the correlation level between pairs of 'point' data depended on the lag distance (and direction) between them and decreased as the lag increased. As geostatistics developed, the necessary spatial correlations were modeled by the now commonly used variogram.

In parallel with the original analyses referred to above and published in 1951, the author also studied the behaviour of borehole grades in the gold fields being explored at the time, particularly from the main sector of the Orange Free State field (Krige, 1952). This paper detailed the global valuation of the Basal Reef (the main gold carrier) in the main sector of the field, covering ten potential mining properties. This global valuation was based on and covered;

i) The geological identification of the Basal Reef in each drill hole and the evidence of the geological continuity of this reef throughout the field.

ii) The observed log-normal distribution of the intersected grades (accumulations) in 91 drill holes, with deflections, on these properties. The log-normal model was introduced earlier by Sichel (1947), as well as the more efficient ‘t’ estimator based on this model (Sichel, 1952). This was compared with the orthodox arithmetic mean.

iii) Confidence limits based on Sichel’s ‘t’ estimator.

The corresponding estimates and confidence limits for the global mill tonnage and recovery grade were based on the estimated spatial structure for the 91 intersection values using a variance-size of area analysis as shown in Fig. 1 (Krige, 1952). This model can readily be transposed into the later more regularly used variogram model.

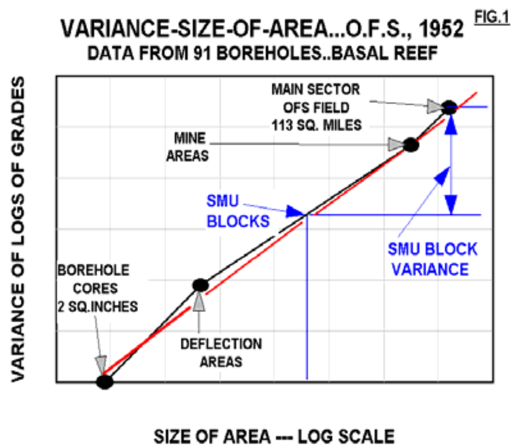

Fig. 1 : First spatial structure for gold

This model can directly provide an estimate of the variance of ore blocks (i.e. the SMU block variance) within the main sector; also, via the log-normal model a tonnage-grade curve and thus an estimate of the tonnage and grade above any pay limit. These estimates are summarised in Table 1, together with follow-up grades from subsequent underground exposures, and production results.

TABLE 1

	Drill hole estimate	First 11 shaft intersections
Global average grade (cm.g/t)		
‘t’ estimate:	1790	1777
Central confidence limits :	104/3286	1123/2926
Arithmetic mean :	2256	1777
	Estimate 1952	Production to 1980
Mill tons (millions) :	680	365
Recovery grade (g/t) :	16.8	14.5
Lower 1/100 limit : Tons	454	
g/t	11.7	

The mill tonnage and recovery estimates included estimates of the pay limit (cut-off), mining widths and losses, and also plant recoveries. The production figures refer to mining of the Basal Reef as well as of some lower grade reefs.

These estimates accepted the log-normal model for the distribution of SMU's. Where the distribution departs from this model, various advanced techniques are now available to overcome this problem (Sichel et al, 1992). In addition, various problems and procedures for the valuation of the tonnage grade estimates are discussed in paragraph 5 below.

The Free State Geduld mine, included in the above global estimates, was floated on the evidence of the geological and grade continuity of the Basal Reef on the mine and throughout the Main Sector of the field, and on the exposure from only 5 deep drill holes, which gave the following cm.g/t values:

988, 4720, 5234, 7607, 100310

The last value was the highest encountered in the whole field, with visible gold in the core. The summary in Table 2 shows how, with the development of techniques – partly Bayesian – which take full account of the larger perspective, the problem of such an extraordinary high value can be handled without any arbitrary cutting or capping of the value.

TABLE 2
FREE STATE GEDULD ESTIMATES

	Mean Cm.g/t	Central 90% Limits Lower	Upper
Arithmetic Mean	233772	9534	97795
# Sichel's 't' estimate	18486	6248	625830
## Macro co-kriging with Bayesian 't' estimate	5399	2952	8836
First 20000m of development	5351	-	-

#Unrealistic because the log-variance of the 5 values is well outside all known limits because of the hgh value of 100310.
Based on the spatial structure for the whole Main Sector, for the mine concerned as a member of a population of 10 mines, the average log-variance within mines for a Bayesian type t" estimate of the mine and on the co-kriging of this estimate with those of the other 9 mines.

The results show the advantages gained by introducing the additional knowledge of the grade distribution model, and the broader Bayesian perspective on the relevant data together with the data's variability.

5. CONDITIONAL BIASES, SMOOTHING AND POST-PROCESSING

At the production stage when selective mining decisions are made on the data then available, ore reserve blocks have to be valued on the most efficient technique, i.e. on a basis which ensures the minimum error variance. This is impossible without, at the same time, eliminating all conditional biases and this, in turn, can only be ensured with *Simple Kriging(with the population mean) or with Ordinary Kriging(without the mean) but based on an adequate data search routine.* Originally such estimates based on ordinary or simple kriging were called Best Linear Unbiased Estimates (BLUE) with 'Unbiased' clearly covering the elimination of Conditional Biases.

Any estimates at this stage, which still incorporate conditional biases are unacceptable, should not be called 'kriging', and should not be passed by a 'competent person' for the *publication of reserves in terms of any of the codes at present being formulated or updated world wide.*

At the exploration stage kriged block estimates with a proper search routine will be conditionally unbiased and will have the lowest level of uncertainty, but will, unavoidably, be *'smoothed'* because of the level of data then available. This means they will have a lower 'dispersion' variance than that of the final SMU distribution at the production stage when more information will be available.

Any block estimates at the earlier exploration stage when much less data will be available must, thus, be less efficient than at the final selection stage. They will, in effect, be 'smoothed' and will generally overestimate the tonnage above the economic cut-off and underestimate the corresponding grade. The reason for this 'smoothing' effect is that proper kriging is, in fact, a regression estimate and it is well known in classical statistics that regressed estimates have a variance equal to the variance of the dependent variable 'y' less the conditional variance of the 'y' values (or error variance of the regressed estimates). This error variance reduces as more data become available; at the same time the 'smoothing' effect will decrease, the 'dispersion' variance of the estimates will increase and the efficiency of the estimates will improve.

Various post-processing techniques are available to correct for this smoothing feature such as e.g. *uniform or direct conditioning* (Assbey-Bonsu, 1999) which, effectively, can provide for each 'smoothed' block, or set of blocks, the average percentage of ore and the corresponding grade, which can be expected to be selected for mining when the more closely spaced data become available. Note that such techniques can be applied only to block estimates which are conditionally unbiased. These *processed estimates* can then be used in the life of mine programme for feasibility studies. However, it does not provide, and no technique can ever fully provide, at the exploration stage, for a practical detailed definition of the in situ SMU blocks which will eventually be selected or discarded. An alternative interim procedure for mine planning in order to minimise the misclassification of ore blocks until the final selective mining decisions have to be taken, has also been suggested (Krige,1999).

It is in trying to overcome these problems that some geostatisticians have used, and still use, the fallacious procedure of limiting the data search routine for the 'kriging' of block valuations to a level which is intended to provide a set of unsmoothed block estimates, which will correspond to a global grade-tonnage pattern close to that expected at the production stage. Although the global tonnage and grade estimates could thus be reasonable, the individual block estimates and even the averages of the combinations of blocks to be mined over limited time periods, will be conditionally biased and unacceptable. This aspect has been covered recently in detail in various papers (Krige, 1999a, Krige and Assibey-Bonsu, 2000, and 2001). Similarly, it has been shown that simulation whilst also providing reasonable global estimates if based on an adequate global data level, will also fail to provide conditionally unbiased estimates for individual blocks and for local areas where the data coverage is inadequate (Krige and Assiby-Bonsu, 2001).

6. CONCLUSIONS

This short review cannot do full justice to all the developments in geostatistics over more than 50 years, including all the variations in kriging techniques, simulation applications and extensive developments in non-mining fields, e.g. animal abundance in the Kruger Park(Steffens,F.C., 1992), and weather modification (Steffens, F.E. 1988). The world-wide spread of the geostatistical community covers a multi-disciplinary make up of mining engineers, mathematical statisticians, geologists, surveyors, metallurgists, hydrologists, etc. The discipline will continue to develop new techniques and to cover new fields on a foundation of fundamental concepts most of which have been refeered to in this publication.

6. REFERENCES

Assibey-Bonsu W. and Krige D.G., (1999). Use of direct and indirect distributions of selective mining units for estimation of recoverable resource/reserves for new mining projects. *APCOM'99 Symposium, Colorado School of Mines, Golden, Oct. 1999.*

Krige, D.G., (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *J.Chem.Metall.Min.Soc.S.Afr., Dec.1951, 119-139.*

Krige, D.G., (1952). A statistical analysis of some of the borehole values in the Orange Free State goldfield. *J.Chem. Metall. Min. Soc. S.Afr., Sept. 1952, 47-64*

Krige, D.G., (1960). On the departure of ore value distributions from the lognormal model in South African gold mines. *J.S. Afr. Inst. Min. Metal, 61, 231-244.*

Krige, D.G. and Ueckermann, H.J., (1963). Value contours and improved regression techniques for ore reserve valuations. *J.S.Afr. Inst. Min. Metall., May 1963, 429-452..*

Krige D.G.(1999a). Conditional biases and uncertainty of estimation in geostatistics. *Proc. APCOM'99 Symposium, Colorado School of Mines, Oct. 1999.*

Krige D.G. and Assibey-Bonsu W.(1999b). Practical Problems in the estimation of recoverable reserves when using simulation or block kriging techniques. *Conditional Simulation Conference, Perth Australia, October, 1999.*

Krige D.G. and Assibey-Bonsu W.(2000). 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.*

Krige D.G. and Assibey-Bonsu W.(2001). Valuation of recoverable resources by kriging, direct conditioning or simulation. 29th APCOM, Beijing, China, April, 2001, Balkema Publishers.

Sichel, H.S.,(1947). An experimental and theoretical investigation of bias error in mine sampling with special reference to narrow gold reefs. *Trans. Inst. Min. Metall. Lond., 56, 403-473.*

Sichel H.S., (1952). New methods in the statistical evaluation of mine sampling data. *Bull. Inst. Min. Metall., Lond., June, 1952, 261-288.*

Sichel H.S.(1972). Statistical valuation of diamondiferous deposits. *Proc. 10th APCOM Symp., S.Afr.Inst.Min.Metall., 1972: pp. 17-25.*

Sichel H.S.,Kleingeld,W.J. and Assibey-Bonsu,W. (1992). A comparative study of the frequency distribution models for use in ore valuation. *J.S.Afr. Inst. Min. Metall. April, 1992, 91-99.*

Steffens,F.E.(1992). Geostatistical estimation of animal abundance in the Kruger National Park, South Africa. *Proceedings Geostatistics Troia '92, Kluwer Academic Publishers, 1992.*

Steffens,F.E. (1988). Design of a raingauge netwotk for a weather modification experiment. *Proceedings 3rd International Geostatistics Congress, Avignon,France, Geostatistics, Sept. 1988 Kluwer Academic Publishers..*