ESSENTIAL BASIC CONCEPTS IN MINING GEOSTATISTICS AND THEIR LINKS WITH GEOLOGY AND CLASSICAL STATISTICS.

by

D.G.KRIGE

ABSTRACT

The basic concepts involved in the birth and the historical development of mining geostatistics over more than half a century, are reviewed. The review concentrates on the main fields of:

the definition of the statistical population(s), i.e. the ore body parameter(s) of interest, e.g. grade,

their frequency distribution(s),

the geological model of mineralisation,

the types of support of the members of the population(s),

the spatial structures involved and, where applicable,

the subdivision of the ore body into more homogeneous sub-populations,

the impact of these factors on the confidence levels for the relevant estimates, e.g. of block grades, and

the need for practical validation checks on the quality and unbiasedness of the estimates and for anticipating and eliminating any global or conditional biases.

SAMEVATTING

Die onderliggende begrippe betrokke by die ontstaan en historiese ontwikkeling van mynbou-geostatistiek oor meer as ‘n half eeu word in oënskou geneem. Die oorsig konsentreer op die hoof aspekte van:

die definisie van die statistiese populasie(s), dws. die ertsliggaam se parameters van belang, bv. die ertsgraad ,

die frekwensie verdelings,

die geologiese model van mineralisasie,

die tipes van ondersteuning van die lede van die populasie(s),

die betrokke ruimtestrukture en, waar toepaslik,

die onderverdeling van die ertsliggaam in meer homogene subpopulasies,

die impak van hierdie faktore op die foutperke van die betrokke skattings, bv. van die ertsblokgraad, en

die noodsaaklikheid van praktiese bevestiging van die kwaliteit en ewewigtigheid van die ertsblokskattings, en om enige globale of voorwaardelike onewewigtighede te antisipeer en te elimineer.

INTRODUCTION, GEOLOGY AND GEOSTATISTICS

This paper is intended to place the roles of geology and geostatistics in ore valuation in perspective, to highlight the basic geostatistical concepts as they developed historically in South Africa, and to address some current problems and the future challenges of avoiding these problems.

In the field of ore valuation, geology and geostatistics are the two inseparable sides of the same coin. On the one side geology concentrates on the physical features of the orebody,

such as structures, source and type of mineralisation, and all parameters associated with the levels and patterns of mineralisation. Ideally, geology will produce a logical model of the mineralisation which in turn should form the foundation of any geostatistical analysis. Geostatistics is the other side of the coin and provides mathematical, statistical and geostatistical models for the analytical sampling data available in order to introduce efficient valuation techniques for resource and reserve estimates and to attach confidence limits to these.

FREQUENCY DISTRIBUTION MODELS

The initial efforts in applying classical statistical procedures to ore valuation date back to 1919 (Watermeyer) and 1929 (Truscott). The grade distributions of underground samples on the gold mines were studied in some detail. Contrary to studies in classical statistics, where the symmetrical Normal distribution was in common use, these gold distributions were very skew with long tails towards the higher grade ranges. This led Watermeyer and Truscott to ignore *a basic statistical concept, i.e. that regardless of the frequency distribution pattern, the arithmetic mean of a representative or random sample set of data is an unbiased estimator of the true mean.* Truscott had the misconception that by using a weighting system based on the square of the relative frequencies, the 'problems' caused by the skewness would be overcome. Skewness was mistakenly blamed for low Mine Call Factors, i.e. for the differences observed between gold called for by the sample data and that accounted for in the plant.

It was only in the late 1940's and early 1950's that Sichel(1947,1952) introduced *the lognormal model for gold values*, and using this model developed the 't' estimator. This

estimator is more efficient than the arithmetic mean but is strictly only valid for a random set of data and which follows the lognormal model exactly. Departures from the usual lognormal model were largely overcome with the introduction in 1960 (Krige,1960) of *the 3-parameter lognormal* which requires an additive constant before taking logarithms. However, there were still cases which could not be covered by the 3-parameter lognormal, and Sichel (1992) recently introduced *the more flexible Compound Lognormal Distribution,* originally developed by him for diamond distributions. This model is very flexible and caters specifically for a tail of high values which is much longer than that for earlier models. However, the small sampling theory for this model is not yet available and in practice, some of the parameters have to be accepted on a Bayesian basis; also the presence of a spatial structure in the data still imposes a restriction as in the earlier models.

SPATIAL PERSPECTIVES, ELEMENTARY KRIGING

Geostatistics as such did not really originate until the basic concept of ore grades as a spatial variable with a spatial structure was introduced in 1951/2 (Krige,1951,1952). This arose firstly in an endeavour to explain the experience on all the gold mines for many decades, of *ore reserve block estimates during subsequent mining consistently showing significant undervaluation in the lower grade categories and the reverse for estimates in the higher grade categories, i.e. what is now known as conditional biases.* To explain this problem, it was essential to no longer view the peripheral data used for individual block estimates and the ore blocks themselves in isolation. In a broader perspective, *the peripheral data were seen as part of an extensive spread of data (the data population) in stopes and development ends in the relevant mine section, and the ore block concerned as part of a collection of blocks (both intact and already mined out), i.e. as a member of a population of ore blocks.*

In this way, *the spatial concept was introduced*, a theoretical mathematical model was set up of the distribution of actual block values, and the error in assigning the peripheral grades to the blocks were super-imposed to yield the corresponding distribution of block estimates. On correlating these two sets of block values, the averages of the actual block values corresponding to specific grade categories of block estimates could be observed, i.e. the regression of actuals on estimates. This was a theoretical 'follow-up' exercise and provided the geostatistical explanation of purely natural phenomenon, i.e. the unavoidable under- and over-valuation features as mentioned above. In statistical termonology these features are called conditional biases. At the same time, this justified the so- called *regression adjustment to block estimates* which eliminated these conditional biases. Also, as such a regressed estimate was in effect, a weighted average of the peripheral estimate and the global mean of the mine section, it was *the first application of kriging. It can be labelled Simple Elementary Kriging, being based on the spatial correlation between the peripheral values and the actual grades of the ore inside the blocks, and giving proper weight to the data outside the block periphery via the mean.*

During the 1950's several large gold mines introduced regression techniques for their ore reserve estimates on a routine basis. *It is instructive to observe that on our gold mines, the improvement in the standard of block valuations due to the elimination of conditional biases, accounts for some 70% of the total level of improvement achievable today with the most sophisticated geostatistical techniques. It is for this reason that the author places so much stress on the implementation of conditional unbiasedness.*

By 1963/66 (Krige,1963,1966) the spatial patterns were defined in far more detail. These studies covered the spatial correlations between individual "point' sample values, as well between regularised data blocks, into which the same point data was summarised. The

corresponding correllograms or covariograms were used on a *simple kriging* basis for block valuations. Kriging on a routine basis for ore reserve valuations was, therefore already in use on some Anglovaal gold mines more than 30 years ago. *These estimates were then still called 'weighted moving averages' until Matheron's insistence in the middle 60's on the term kriging prevailed and became generally accepted.* The weighting system was also soon adjusted to cater for the fact that the spatial structures proved to be stronger in the direction of significant grade 'shoots', i.e. anisotropies were already catered for.

Matheron, also then proposed the use of the *variogram to define the spatial structure.* This model is an extension and refinement of the concept covered by De Wijs (1951/3). An alternative model was already used in 1952 (Krige, 1952) in an analysis of the borehole values in the Orange Free State field. This model showed the trend of increasing logarithmic variances of data within a range of sizes of areas from the small deflection areas up to the field as a whole. *This 'variance - size of area' graph shown in figure 1 can be translated directly into a De Wijsian variogram.*

CONFIDENCE LIMITS

The classical statistical concepts used in deriving error variances and confidence limits for estimates based on a set of data drawn from a population also apply in geostatistics, with the main difference of *having to incorporate the effects of skew distributions, the presence of a spatial structure and of dealing with non-randomly spaced data.*

The inaccuracy of grade estimates based on a small number of values is, even today, not fully appreciated. For example it is a widespread practice in opencast mining to take final grade control decisions on the evidence of individual blast hole grades. This is often done in spite of a variogram analysis of such grades showing a significant nugget effect. Thus, if

additional blast holes were to be drilled very close to existing ones, significantly different grades could be found. In principle, the problem is identical to that of 1951, (of block valuations based on peripheral values only), as well as the solution, i.e. to also take into account the background of available data in the vicinity via the appropriate kriging technique.

The basic geostatistical principle involved is that of reverting to the original 1951/2 data perspective of also using the additional background data plus other support data such as blasthole assays.

The correct error distribution to use for calculating skew confidence limits for any grade estimate is a difficult theoretical problem. However, my experience has shown that in general, the use of a 2- or 3- parameter lognormal model for the error distribution (with the same natural error variance as that for the estimate) provides acceptable practical confidence limits.

THE CUTTING OF HIGH GRADES

As a direct result of the positive skewness inherent in most ore grade distributions, it does happen from time to time that in a small set of data, e.g. in a set of borehole values or a set

of stope face values, a value is present which appears to be *an 'outlier', i.e. a value which appears to be alien to the general pattern of the other values.* This is usually a high grade value which can have a very significant influence on the average of the small set of values. The practice of cutting such an apparent 'outlier' usually follows some arbitrary procedure or may attempt more respectability by adjusting the value down to that which could be

expected by fitting a lognormal model to the pattern indicated by the remaining values and substituting a value which would agree with the upper tail of a perfect lognormal distribution.

Firstly, provided the so-called 'outlier' value is not seen as of doubtful origin due to a mistake, bad assaying, etc., the 'outlier' must be accepted as a real member of the complete population. Such values will recur from time to time to provide eventually in a large accumulation of values, the correct relative occurrence of such grades. In these circumstances there would be no justification for doing any cutting. It would introduce a global negative bias in the estimates and will thus result in a better Mine Call Factor. This might appear to be an advantage and might please the production department, but *such an improvement will, however, be a paper one only and not real.*

The problem of the presence of an 'outlier' in a small set of values, particularly in a set of borehole values, should be approached realistically. On assumption of lognormality, the 't' estimator, using either the 2- or 3- parameter model, could provide some relief. However, the log variance used will also be affected by the 'outlier' and the answer could still be unrealistically high. A Bayesian approach, where the available experience for similar occurrences provides a realistic range for log variance levels, can then be used together with the t" estimator (Krige,1952). The following example of the Free State Geduld mine illustrates this point clearly:

On flotation the property had 5 boreholes, with grades ranging from 988 to the famous hole of over 100000 cm.g/t; the arithmetic mean was 23774 cm.g/t. Sichel's 't' estimate was slightly more realistic at 18486 cm.g/t but had the mine been viewed in the perspective of the surrounding mines and a macro spatial structure, the t'' macro-kriged estimate would have been 5399 cm.g/t compared to the average of the first 20000m of underground development at 5351 cm.g/t.

Where kriging estimates are made, e.g. for ore blocks or stope faces, and enough data is accessed for the valuation, the effect of such an 'outlier' will similarly be minimised.

I have encountered in practice, a case where thousands of values were available for a massive 3D orebody, but all grades above a certain limit were adjusted downwards to result in values which fit the upper tail of a lognormal model. However, ample evidence was available, as shown on figure 2, to prove that the straight lognormal was not the correct model, that the data had a much longer tail and that the data could have been fitted very well with a compound lognormal. *The use here of what appeared to be a very sophisticated cutting procedure was obviously not justified.*

GEOLOGY, GEOSTATISTICS AND ORE BODY SUBDIVISIONS

All the geostatistical models referred to above require that the relevant populations are reasonably homogeneous, i.e. that the data do not cover a mixture of populations with significantly different characteristics. This is often evident from the frequency distribution, e.g., a display of bimodality.

To split the data into two or more homogeneous populations purely on geostatistics is, generally, not practical and often impossible. *A good geological model for the mineralisation of the orebody can, however, provide invaluable information and lead to a logical in situ split of the orebody into logical subdivisions.* On our gold mines, for example, these could define the individual facies of deposition.

Of equal or often greater importance can be *the variations in spatial structures within an orebody. Again, geological input is essential to overcome this problem.* Such variations could cover levels of variability, locally and globally, continuity, directions of anisotropy, etc.

A good example was that of the large Chuquicamata copper mine in Chile (Krige,1995). Before a good model of mineralisation was available, the orebody was subdivided only on the evidence of rock types, i.e. a limited geological input. Following the postulation of a detailed model of mineralisation of the ore body, a number of further subdivisions were introduced to cater for changes in the directions of the networks of fractures which formed the main channels of mineralisation. The incorporation of these geological subdivisions in the geostatistical analysis showed, for each subdivision, *an excellent agreement between the main directions of mineralisation as defined by the geological model and the directions of maximum continuity as indicated by the geostatistical models fitted to the data.* The final ore reserve estimates showed a very significant improvement in the standard of the block valuations as measured by the correlation levels between the block estimates and the follow-up grades of the subsequent internal blastholes. Following the introduction of the geological model of mineralisation and the use of blasthole data in addition to drillhole data, the correlation coefficient between block estimates and the follow-up values improved from 76% to over 88%.

A good geological model is, therefore, invaluable and if the geostatistical model does not agree with the geological one, there are grounds for serious concern. Either or both models should be critically re-examined so as to establish the essential agreement.

A very dangerous practice is that of subdividing the orebody, not on geological grounds, but directly on grade only. This can lead to serious biases, particularly where the data in one or more subdivisions are insufficient to allow of proper geostatistical analyses. A test

for validity of such a practice is that of doing a relative, or logarithmic, variogram analysis for the data with and without such a subdivision. If the two variograms are very similar, the subdivision could very well be artificial.

THE ROLE OF SUPPORTS

'Support' is the geostatistical term for the physical ore parcel which supports the ore grade being studied as a variable. This concept is also very basic to geostatistics and was first covered by Ross in 1950 (see Krige,1951), who showed how, for the same area covered, the spread, or variance, of the frequency distribution of grade reduced consistently as the size of the support was increased from that of a single chipped channel sample to the average of a string of these samples along a development end or stope face. Within the

same mine section these distributions all had the same mean but a wide range of variances.

This perspective of visualising an ore body as made up of an infinite possible collection of populations, each with its own distinctive support level, is a useful approach. It is instructive to study the effects of different support sizes not only on the relevant population variances but also on the shapes of the distribution models, as well as on the spatial patterns of continuity in critical directions, etc. In practice this can serve to indicate whether the lognormal model or another model should be used for tonnage-grade estimates.

ORE RESERVES, SELECTIVE MINING AND CONDITIONAL BIASES

Most orebodies are mined selectively to a cut-off grade, on the basis of ore units selected to be sent to the plant or to be discarded as waste or uneconomic material. The sizes and

shapes of *these selective mining units (SMU's)* are determined by the nature of the mining operation. The population of SMU's within the orebody is thus the target of all resource/reserve analyses. With this perspective of the orebody and, particularly at *the exploration stage*, when the data is very limited, the logical approach is to estimate:

(i) the global mean grade – where the data is extensive and a single kriged estimate is impractical, any significant clustering of data should be handled by an *appropriate declustering technigue* before a global mean is estimated,

(ii) the spread (variance) of the distribution of the grade estimates of the SMU's,

(iii) the frequency distribution model of the SMU estimates, and then

(iv) the tonnage-grade curves which provide the required estimates of the resources or reserves in situ above any cutoff.

The first exercise on this basis was done in 1952 (Krige,1952) for the main part of the Orange Free State Goldfield and in retrospect was shown to have been a realistic and good estimate. The 1952 estimate showed a mill tonnage of 617 million at a recovery grade of 16.8 g/t compared to the production figures up to the end of 1980 for the 10 mines in this sector of 365 million tons at 14.5 g/t. Bearing in mind that the estimate was confined to the Basal reef and the production figures include production from other lower grade reefs and that much additional production took place since 1980, the 1952 estimate can be accepted as realistic. *Orthodox methods could not have made such a prediction*.

When the available data is on a relatively close grid, e.g. for an operating mine, the requirements for a resource/reserve estimate cover not only a tonnage-grade curve but that this be related to grade estimates for individual SMU blocks. Mine planning can then be

carried out using a 2D or 3D data base of the individual SMU grade estimates. The following support problem arises at this stage:

(i) Some of the SMU estimates will have data supports on the basis of the final data grid for selective mining, e.g. closely drilled blast holes in a 3D ore body being mined,

(ii) Some SMU's will have a data support of only widely spaced data, e.g. the final grid of exploration drill holes, and

(iii) Some SMU's will have a data support consisting of a mixture of (i) and (ii) above.

In all three cases the basic original geostatistical concept will apply. i.e. *the need to do block estimates with the lowest error variance and which are globally and conditionally*

unbiased. These were originally called BLUE estimates, i.e. Best Linear Unbiased Estimates. All BLUE estimates (today not necessarily linear) will meet the original objective of the 1951 Elementary Kriged estimates, i.e. of being conditionally unbiased. This requirement cannot be overstressed. *I have repeatedly encountered cases where geostatisticians, who should be fully aware of this 50 year old basic concept, still do block estimates which have suboptimum error variances and are badly conditionally biased. Furthermore, no proper validation analyses are done to check for such biases. and for the general standard of the estimates. There is no excuse for such a practice, particularly for final estimates in category (i) above; in effect this would take ore valuation back to the orthodox period before the birth of geostatistics.*

Final selective decisions are taken on the basis of category (i) estimates and any remaining conditional biases will result in significant tonnages of marginal ore going to waste, the average grade of ore above cut-off being overvalued and the resultant net profit being overestimated and, in fact, being realised at a level below the optimum(Krige,1966).

The common reason for the presence of such conditional biases in ordinary kriged block estimates, particularly at the exploration stage, is the use of an over-restricted search routine in accessing the data to be used in a block valuation. The motivation for such a practice is to avoid the so-called smoothing of the block estimates so as to produce a set of SMU estimates which will have a spread close to that expected at the final mining stage when closely spaced data will be available. *The problem is that such individual block valuations will, individually, not have the lowest error variance, will be conditionally biased*

and cannot be used safely in any mine planning exercise and cannot be used for any reserve declarations. This practice must be strongly condemned.

VALIDATION EXERCISES

Bearing in mind the critical importance of resource/reserve estimates, and the data and facilities available today, there is no excuse for not doing proper validation studies in the case of existing mines and studies to anticipate the standard of the estimates *in all cases*. In existing mines validation of the procedures should be done on the closely spaced data

in mined out parts of the orebody. This should be done by simulating normal block valuations using the data normally available, but excluding of the follow-up data inside the block, or using previous block valuations. These block estimates are then correlated with the corresponding closely spaced follow-up data now available inside these blocks. The resultant correlation graphs and statistics will provide a *check, not only on the validity of the geostatistical techniques used, but also on the proper use of these techniques, e.g. of the search routine and other parameters which are left to the discretion of the user.*

At the feasibility, or any other stage, the critical results and the overall standard of the valuation can, in any case, be anticipated given all the models and parameters used(Krige,1996). In this way the likely results of a later validation exercise can be anticipated. If these are not satisfactory, the approach and/or the parameters used can be modified until the results are acceptable.

Validation is, therefore, an absolutely essential component of any valuation exercise.

CONCLUSIONS

Attention has been focused on the need, in the field of ore valuation, of never losing sight of the basic historical concepts, as relevant today as ever, regardless of the type and sophistication of the techniques used. These concepts are as follows:

Geological input, particularly a model of the mineralisation.

The proper frequency distribution model(s) for sample and block values.

Detailed spatial perspectives and models.

Block valuations(SMU,s) without conditional biases and with lowest error variances.

Realistic confidence limits.

No cutting of high grades unless fully justified.

Logical ore body subdivisions based essentially on geological input.

The effects of changes in support sizes and types.

The essential checking of the standard of the valuations via validation exercises and/or analyses which can anticipate this standard.

The author's experience in recent years, locally and overseas, and covering a wide spread of project types, has shown that there is indeed a real need to stress the dangers of ignoring any of these basic concepts.

Looking to the future, new and more sophisticated and efficient techniques will, of course, continue to be developed but should remain within a framework which does not violate these basic concepts.

REFERENCES

De Wijs,H.J.(1951/3). Statistical ore distribution. *Geologie en Mijnbouw, Nov.1951, Jan.1953.*

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. (1966). Two- dimensional weighted moving average trend surfaces for ore valuations. *Proc. Symp. on Math. Stats. and Comp. Appls. In Ore Valn., 13-38, S.Afr. Inst. Min. & Metall.*

Krige, D.G. and P.G. Dunn (1995). Some practical aspects of ore reserve estimation at Chuquicamata Copper Mine, Chile. *25th APCOM, Brisbane, Australia, July, 1955.*

Krige, D.G. (1996). A practical analysis of the effects of spatial structure and data available and used, on conditional biases in ordinary kriging, *5th Int. Geostatistics Congress, Wollongong, Australia, 1996.*

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

Trustcott, S. J. (1929). The computation of the probable value of ore reserves from assay results. *Trans. Inst. Min. Metall., Lond., 39., 482-496.*

Watermeyer,G.A.(1919). Application of the theory of probability in the determination of ore reserves. *J.Chem.Metall.Min.Soc.S.Afr.,19, 97-107.*

CAPTIONS FOR FIGURES:

Figure 1 showing the variance-size-of-area graph used 1n 1952 in an estimate of the expected mill tons and recovery grade for the main sector of the OFS goldfield, based on the Basal reef results of 91 boreholes.

Figure 2 showing the plot on log-probability paper of the distribution of drill hole values from a massive 3D orebody, the fitted Compound Lognormal model and the lognormal model which was used for the cutting of high grade values.

Figure 2

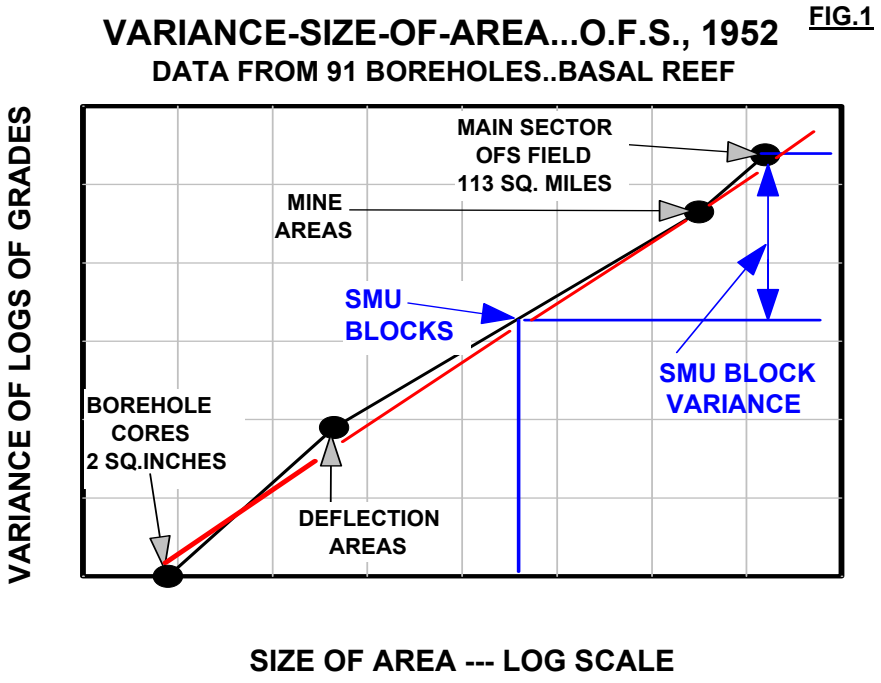
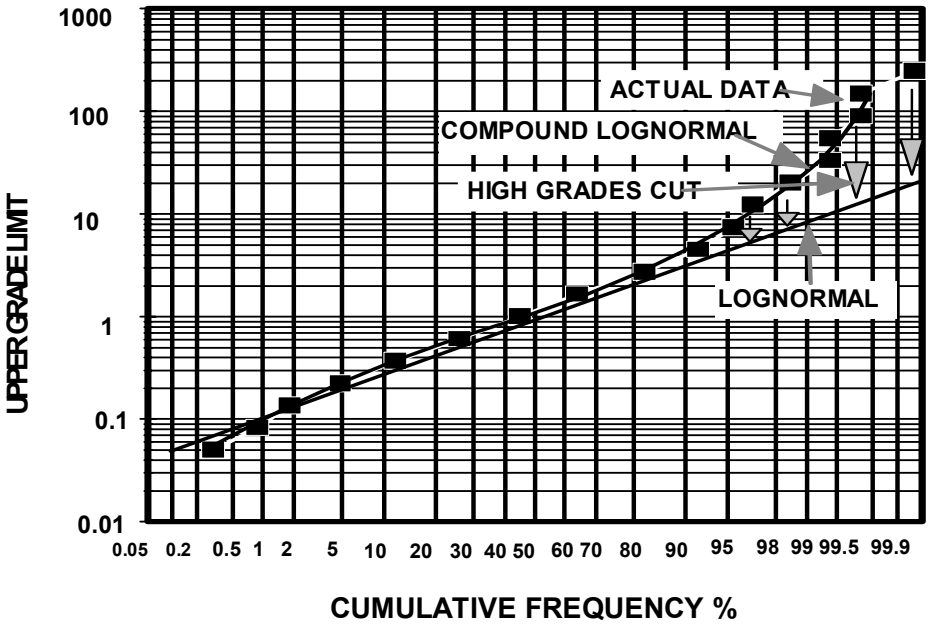

VARIANCE-SIZE-OF-AREA...O.F.S., 1952
FIG.1
DATA FROM 91 BOREHOLES..BASAL REEF
VARIANCE OF LOGS OF GRADES
MAIN SECTOR
OFS FIELD
113 SQ. MILES
MINE
AREAS
SMU
BLOCKS
SMU BLOCK
VARIANCE
BOREHOLE
CORES
2 SQ.INCHES
DEFLECTION
AREAS
SIZE OF AREA --- LOG SCALE

Figure 2:


1000
100
10
1
0.1
0.01
UPPER GRADE LIMIT
ACTUAL DATA
COMPOUND LOGNORMAL
HIGH GRADES CUT
LOGNORMAL
0.05 0.2 0.5 1 2 5 10 20 30 40 50 60 70 80 90 95 98 99 99.5 99.9
CUMULATIVE FREQUENCY %