

Identification of multiple mineralization phases by statistical methods

I Clark M.Sc., F.S.S.

Department of Mining, Imperial College, London

R. H. T. Garnett Ph.D., M.B.A., C.Eng., M.I.M.M.

Formerly Anglo American International (U.K.), Ltd., London (now Anglo American Corporation of South Africa, Ltd., Johannesburg, South Africa)

519.272: 622 013.34

Synopsis

Ore-reserve estimations of deposits with low mineral concentrations, such as those of copper, gold and tin, are usually made on the assumption that the grades follow a lognormal distribution. Multiple mineralization or reworking of a deposit may have modified the simple distribution expected — producing an overall distribution of grades which can be adequately described by a mixture of lognormal curves. A new method of separating components of mixed populations is described, one simulated and three practical examples of its application being presented.

Introduction

This paper has been prepared in order to describe a technique recently developed by one of the authors for the identification and quantification of component ore phases in deposits subjected to multiple mineralization. Examples to illustrate the application of the method and the necessary geological interpretation of the results have been selected from well known mineral fields.

Normal and lognormal distribution

The evaluation of an ore deposit usually requires a statistical analysis, and most methods were originally developed on the basis of the so-called normal or Gauss/an distribution. This is the familiar bell-shaped distribution, which has many advantageous features when used in analyses: most important of these is the fact that reliable estimates may be found easily for the distribution of the large unknown population from the relatively small known sample.

The normal distribution, however, is often not applicable to real sample values. For example, low-concentration mineral deposits and oil reservoirs are known to have highly skewed distributions. Also, there are limits on the values which the samples possess — a sample cannot contain a negative amount of mineral! In practice, there is a definite probability of obtaining 'erratic' high values from a deposit, as well as the relatively large probability of many low values. None of these properties exists in a normal distribution.

Although many other theoretical distributions may be used to fit such a practical curve, the first possibility to be considered seriously is that some transformation of the grades (x) might reduce the curve to a normal one. If natural logarithms of grade values are taken, it can be seen that $y = \log_e(x)$ assumes values between minus infinity and plus infinity. That part of the curve corresponding to the multitude of lower grades is 'stretched out', and the long tail into

the erratic high grades is reduced in importance. Thus, we arrive at the definition of a lognormal variable. If the logarithm of a variable can be said to have a normal distribution, then that variable is said to be lognormally distributed.^[1,2]

Much work has been completed, theoretically and practically, on the lognormal distribution. The attitude has arisen that "if a deposit's grade distribution is not normal, it must be lognormal".

Multiple mineralization

Histograms of assay data are often constructed during the exploration and development stage of a deposit. In practice, however, they often appear to be neither normal nor lognormal. Frequently, more than one peak or 'mode' is clear in the histogram. A curve expected to be simple may be too highly skewed, or may have too long a tail.

This divergence from the expected normal or lognormal curve may result from the overlapping or intermingling of two or more populations which have been recorded as one. The component populations may have the same distribution types, but possess different means and standard deviations. For example, a deposit may have been mineralized on two separate occasions. If the component grade distributions are sufficiently different, the combined distribution may tend to be highly skewed or bimodal.

The unexpected behaviour of a sample distribution should, however, not immediately be interpreted as being due to the mixing of components. There are theoretical curves which adequately describe a histogram that is too highly skewed to be lognormal. One should not attempt, therefore, to split a histogram into more than one component unless there is enough other evidence or suspicion, usually geological, to suggest that the underlying theoretical distribution is actually composed of two or more components. Each component would be assumed to represent one phase of mineralization, the distribution of which was normal, lognormal or some other simple curve.

Multiple mineralization may, in practice, comprise a repetition or modification of the first 'phase'. For example, two or more distinct periods of hydrothermal infilling of a fracture system or the superimposition of placer deposits constitute repetitious mineralization. The reworking of part of a placer deposit, the mobilization of disseminated sulphides and the oxidation of primary sulphides are examples of modification.

Statistical detection of multiple mineralization

Existing methods of estimating the components of a mixture of normal distributions are divisible into four groups:

- (1) visual methods, by means of which the user guesses from graphs or histograms the necessary parameters:
- (2) graphical methods, in which attempts are made to fit straight-line segments to a graph or figure:
- (3) mathematical methods, involving analytical formulation to give the parameters as the solution to a single equation : and

(4) numerical methods, which, by a process of repeatedly improving estimates of the parameters, provide a final solution which is the 'best approximation' of the parameters involved.

A new method promoted by the present authors is among the last of these, and is a non-linear least squares method of solution. As in all least squares methods, the solution chosen is the one which minimizes the sum of squared differences between the postulated model and the observed data. In the non-linear case the result is found by a series of intermediate approximations converging to the 'best' solution.

Since the literature on the splitting of multi-modal curves is large and widely spread, a detailed review and a description of previously used methods are beyond the scope of this paper. An excellent, recent, review is available^[3] and the following comments owe much to that manuscript.

Visual methods

One can usually make 'eyeball' estimates if the various components are adequately separated. The individual modes are sometimes clear, and the spread of each component can be used to estimate its standard deviation. This method can, however, provide misleading and inaccurate estimates, particularly if the components overlap to a great extent. It is the authors' experience that accurate visual estimates, especially of standard deviations, are difficult to achieve.

Graphical methods

Originating in the 1940s, graphical methods have been greatly developed and widely applied. They are particularly useful in the field or for quick appraisals of data distributions, since powerful calculators or computers are not required. The best known graphical method is that which employs 'probability paper',^[4,5] but other techniques have been evolved.^[6] The main drawback of the graphical approach is its inability to cope easily with a mixture of distributions whose modes are close together. It is also a very time-consuming exercise if more than two components are present.

Mathematical methods

The first successful mathematical method was provided in 1894, but since it involved the solution of a ninth-order equation, the technique was never widely used.^[7] A more recent adaptation^[8] attempted to find the solution for a particular case, equal component standard deviations being assumed. The present authors consider that such an assumption is unrealistic in the context of ore-reserve valuation, since equal component standard deviations rarely, if ever, occur.

Numerical methods

With the advent of the digital computer, approximation methods have become increasingly popular. The more recent methods include maximum-likelihood solutions^[9,10,11]. Their attempts to approximate the maximum of a very flat function, however, cause the computer to spend considerable time making very little improvement in the estimates. The non-linear least squares methods^[12,13] include extra approximations not present in the new method described below.

The proposed method

A 'model' has been developed to describe a population formed by the mixture of several component populations. Each component is assumed to possess a simple distribution, such as normal or lognormal, but all the components need not have the same sort of distribution. For instance, a mixture of a normal and a lognormal could be handled as easily as one of two normals; but the mixtures of only one type will be discussed here.

Once the type of distribution of the components has been decided upon, the model can be formulated. One must then estimate the unknown parameters present in the formulation. If a mixture of two normals is proposed, and the model for this built, it is apparent that five parameters must be determined before the population is completely described. These are the arithmetic mean and standard deviation of each component, and the proportion of the samples thought to come from the second component. For each normal (or lognormal) component added to the mixture, three further parameters are added to the model.

The proposed method comprises an iterative technique known as the Gauss—Newton,^[12,15] modified slightly to give faster convergence. In other words, it 'zeroes in' to the best estimates. It has been developed in the FORTRAN IV programming language, and is ideally suited for use with a computer terminal on a time-sharing system. One could, however, also implement it on a programmable calculator or mini-computer which had the capability of inverting small matrices. The full mathematics of the method are provided in Appendix 1 and a descriptive explanation follows. The program permits very fast estimation, the speed depending on:

- (a) the number of components in the model,
- (b) the number of groups in the histogram and
- (c) how close the user's original visual estimates are to the final estimates.

A CDC 6400 computer has been used for the analyses given later in this paper. On this machine a three-component analysis on a 25 group histogram took 5 sec, and a seven-component analysis on 67 groups required 80 sec^[*] to complete.

Since the data to be used in the histogram consist of a finite, and usually small, number of samples, they will not necessarily be typical of the whole deposit under consideration. The method does not attempt to describe the sample data but, rather, to draw conclusions about the deposit from which they were taken. Statistical and geological evidence of multiple mineralization, and the general characteristics of each phase, must be considered in order to find a valid interpretation for any analysis of the data. One could accept the statistical evidence alone for multiple mineralization without any relevant geological evidence, but the statistical interpretation naturally carries greater weight if it is corroborated by some geological data.

Description of the method

The main steps involved in the use of the method are:

- (1) a histogram is constructed from the sample data;
- (2) from (1) and from geological evidence, the number of components to be included in the

model is decided;

(3) from (1), or from a plot on probability paper, visual estimates of all the parameters are made;

(4) the Gauss-Newton method is then used to find the 'best' possible set of approximations to the parameters, given the visual estimates made; and

(5) a histogram of the model is constructed and is compared with the original data histogram, (1), by means of the chi-squared goodness of fit test.

The initial estimates can be so far from the correct values that the optimum solution cannot be found. At this stage an interactive program used on a computer terminal can be extremely useful. The histogram of the model can be used to make better preliminary estimates, and the whole process can be repeated immediately.

Illustration of the method

To illustrate a full description of the method a set of 1000 samples from a mixture of normal distributions has been simulated on the computer. A histogram constructed from these data (equivalent to step (1)) is shown in Fig. 1(a).

From a visual examination the samples obviously have not been drawn from one simple distribution. We may hypothesize that the combined population from which these samples were taken comprises two overlapping components (step (2)). Let us further assume that each component population possesses a normal distribution. Any one specimen is assumed to come from one of the component distributions, and not to be a mixture of both.† If, for example, we state that the components are in relative proportions 0.3;0.7, we mean that 30 per cent of the specimens are probably derived from population I, and that 70 per cent are derived from population II.

If we let the two components in our model have relative proportions p and $(1 - p)$, the model is expressed as follows: the probability that a sample from distribution I would lie in that group, plus $(1 - p)$ times the probability that a sample from distribution II would lie in that group. It is usually more convenient, when analysing statistical distributions, to deal with 'probabilities of being less than' some value. Thus, we choose to express the model as: the probability of being less than the upper end-point of a particular group in the histogram is equal to p times the probability of a sample from population I being less than that value, plus $(1 - p)$ times the corresponding probability for population II. To aid us in fitting this model to the data we calculate the observed proportion of the samples which lie below each end-point in the histogram. The final group in the histogram includes all values above the previous end-point.

Thus, we have a model of expected probability, and data comprising observed proportions of the samples. These allow us to estimate the five parameters necessary to describe the mixture of two normal distributions which most closely approximate the data. Assuming two components, one must first make visual or graphical estimates of these five parameters (step (3)) — the mean of each component, its standard deviation and the proportion.^[‡]

These five estimates must be fairly close to the true parameters, although by sensitivity testing it has been shown to the authors' satisfaction that any single parameter may be up to 50 per cent in error without markedly affecting the solution. If, however, more than one parameter were this far out, accuracy of the final estimates would depend completely on the

extent of separation of the component populations. The most sensitive parameter is the standard deviation, and particular care should be taken in its estimation.

There were clearly defined modes at 10.5 and 21.0. Since we are looking for normal distributions, each mode will give us a good estimate of the mean of the corresponding component. Thus, we estimate the mean of Population I to be 10.5, and that of population II to be 21.0. The two components seem to be in roughly equal proportions. 50 per cent of the histogram appears to be due to component I, and the same to II. Thus, our estimate of parameter p is 0.5, so $(1-p)$ is also 0.5.

There remains the determination of the standard deviation of each component. In Fig. 1(a) the two components overlap only slightly. Population I seems to have a range or spread of about 14 units, and II has one of about 12 units. A good rule of thumb in estimating standard deviations is that the range of values of samples taken from a normal distribution is usually about six standard deviations. So if we divide the range by six, we have an estimate of the standard deviation. Our visual estimates thus become standard deviations of 2.33 for component I, and 2.00 for II. Together with the other visual estimates these are given in Table 1 (A).

Improvement of estimates

The details of the original data histogram and the visual estimates are fed into the computer program, either via a terminal or by punched cards. The Gauss-Newton method then improves these estimates successively (step (4)). The program improves all the parameters simultaneously. Each parameter is considered in combination with all the others—not in isolation. The values for the parameters which fit the model most closely to the observed proportions of samples below each end-point in the histogram are eventually obtained. For the example described here the final estimates for the two components are given in Table 1 (B).

Reliability of improved estimates

We have found the best fit that the method can give, starting with the visual estimates. But we now have to determine how good that best is. For this purpose we use the chi-squared (χ^2) goodness of fit test (step (5)). Fig. 1 (b) shows the expected histogram constructed from the estimates for the parameters in Table 1 (B). Throughout the analysis it is assumed that the lower end-point of the first group in the histogram is zero, and that the last group contains all samples above the previous end-point. In Fig. 1(a), (b) and (c) the group shown as 2-3 actually contains all samples having a value less than 3, and the group shown as 26-27 contains all samples above 26.

In practice, in calculating a chi-squared value it is usual to merge groups in the histogram containing a 'low' number of samples. The choice of 'low' is subjective, and here the common rule has been applied that any group with a frequency of less than five is merged with its neighbour. Thus, in Fig. 1 (b) we merge the first four groups to produce one with a frequency of 7, and the last two to give a frequency of 9, but only for the duration of the test. The corresponding groups in Fig. 1(a) must, of course, also be merged for the test. Thus, we have effectively reduced the number of groups in the histogram from 25 to 21 for the purposes of the test.

To apply the test we also require the 'number of degrees of freedom' for the chi-squared value. Counting up our information, there are 21 groups in the histogram, and so we start with 21 'pieces' of information. Some information was used, however, in calculating the expected frequencies. We used the total number of samples in the histogram (one 'piece'), and estimates for five parameters (five 'pieces'). Thus, we are left with $21 - 1 - 5 = 15$ degrees of freedom for the chi-squared value, which was calculated for this example to be 85.8. Tables of 'percentage points of the chi-squared distribution' are widely available,¹⁴ and demonstrate that a chi-squared value of more than 30.6 with 15 degrees of freedom is significant at the 1 per cent level. This means that if a random sample were taken from the population described by the model, only once in a hundred times would one expect a chi-squared value larger than 30.6. Thus, it is highly unlikely that a sample giving a chi-squared value of 85.8 came from such a population. So, although we have the best fit of a two-component model to the original data, it is not a good fit. The best fit is, however, a vast improvement over a one-component model, which, with a mean of 15.2 and a standard deviation of 6.3, gives a chi-squared value of 560 with 22 degrees of freedom.

Repeated estimates

Let us inspect Fig. 1 ((a) and (b)) to try to identify any visible disagreement between the original data histogram and the model histogram. The upper parts of the two figures seem to agree fairly well, within the bounds of sampling error. So, let us look closer at the lower part - that between 2 and 17. First, the model makes the first peak much lower than the data indicate. Secondly, there is a tail in the original data histogram which is not present in the model. The first component of the model is too low, too fat and too short in the tail.

Suppose we were to hypothesize a third component, lying 'underneath', and masked by the first one. This third distribution could possess a mode equal to that of population I, but a standard deviation much larger than that of I. Two such components would give a low, fat, long distribution and a tall, thin, short-range distribution combining to appear as one component. Let us, therefore, make new visual estimates on the basis of three component distributions forming the combined original data histogram.

Since component II seems a good fit to the upper part of the histogram, we shall retain the estimates derived from the last analysis as our initial estimates for this repeat. Our three estimates for component II are then as given in Table 1 (B). Component I will be divided into two components, namely I' and III. Our previous best estimates for the mean and standard deviation of I should be kept and used for I'. We may guess that component III has a mean of 11.0 and a standard deviation of about 4.0. We still have to estimate proportions for these two components. Population I comprised 55 per cent of the previous model. Let us allocate 30 per cent of this to component I', and 25 per cent to component III.

Reevaluation of these estimated three components by the programmed method produces the final best-fit estimates listed in Table 1 (D). The histogram of the model described by these parameters is shown in Fig. 1(c). It approximates to the original data histogram far better than does the previous two-component.

The chi-squared value for this three-component model is 7.6 with 15 degrees of freedom. Although we now have eight parameters, we still have 15 degrees of freedom, since only one group in the histogram was lost because of 'low frequency'. Comparing this value with the available tables,^[14] we find that 95 per cent of the time we would expect a chi-squared value

larger than this for a random sample from the model described. The actual parameters of the three components used to generate the original data are given, for comparison, in Table 1 (E).

This example was created especially for this explanation. It is very seldom in practice that such accurate preliminary visual estimation can be achieved. A different set of samples from the same population would have given a different set of final estimates for the parameters. Very little work has been done yet to determine how accurate the estimates are, but no statistical technique will tell us the exact parameters of an underlying population. We can only approximate reality.

Log normal distributions

It remains only to describe how the method is used with the lognormal distribution. If the original untransformed sample data are available, a histogram should be constructed from the natural logarithms of each sample value. The problem is then reduced to one of separating mixtures of normal distributions. This may be impossible or be very tedious. If a histogram of the original sample values already has been constructed, it is necessary only to take logarithms of the end-points of all the groups in the histogram.

In the examples which follow, the first procedure was adopted for the Cornish tin lode, whereas the other two examples were based on previously obtained data and were modified by the second treatment. When estimates had been obtained for the parameters of the component normal distributions, the means and standard deviations of the 'parent' lognormal distributions were found as described in Appendix 2.

Applications

The identification of multiphase mineralization is important if the information can be used through geological reasoning to find further ore and to evaluate better known ore. The method may be employed upon the results of geochemical surveys, drilling campaigns and production sampling. Area searches in two or three dimensions allow the spatial and geographical distribution of different phases to be determined. The search for physical continuations and extensions of each is thereby made easier. Geological ore controls, not evident in the distribution of the combined mineralization, may sometimes be revealed. Valuation procedures can be modified to accommodate the different characteristics of two or more ore phases within a single mining area. 'Ancient' records may be examined and checked against expectations based upon updated geological analysis, or against more recent exploration, to provide an improved basis for significance testing of any discrepancies. The method proposed by the authors does not claim a monopoly of these advantages, which are possessed by any of the multiple population detection procedures, but it does possess the advantage of speed and of improved flexibility and power of detection. It therefore allows many analyses to be undertaken under circumstances in which one would previously have been deterred or prevented by one, or a combination, of reasons, such as lack of time or personnel, and complexity.

The input sample data may be in any of the following forms:

(1) grade over variable width (per cent. Or equivalent over the orebody thickness, placer depth, or over any other specified width, such as a vein or reef);

(2) grade over fixed width (per cent or equivalent over fixed drill length or uniform channel sample lengths); and/or

(3) metal content (product of the grade and the corresponding length over which each individual grade is measured).

There is no optimum form of input. Instead, the most convenient form should be used at first, whether it be metal content or grade (over fixed or variable width) and the interpretation can be completed accordingly. Spurious effects may be obtained from studies of the grade over variable width if the latter varies considerably or itself possesses a significant distribution. Subsequent analyses of the grades over fixed widths may be necessary to remove the influence of width.

Examples

Three very different examples have been selected to illustrate the application of the method. First, the study of a Central African copper deposit demonstrates the ability to detect and quantify all the ore phases within an entire mineral area. The potential benefits of area-searching a single deposit are illustrated by an examination of successive horizontal sections through a Malaysian placer tin area. Finally, a study of a Cornish tin lode demonstrates that the application may be reversed. Sections of ore may be correlated one with another by identification of the characteristics of their component mineralization phases—by the matching of their 'fingerprints'.

Central African copper deposit

A structurally complex deposit in Central Africa comprises two stratified ore horizons, which have been investigated by drilling from surface. Disseminated and massive copper sulphides occur in dolomite and shale horizons in a section of the Lower Roan Series.^[19] The primary sulphide ore, *S*, merges gradually into an over-lying zone of oxide ore, *O*, which, in places, is separated from surface by a leached zone. Sporadic low-grade mineralization occurs in the wallrocks. A typical cross-section through the deposit is provided by Fig. 2.

A total of 1133 assay results was systematically collected from a considerably larger number of samples derived from all the diamond drilling. The data were used to construct a histogram (Fig. 3). The distribution appears to be approximately lognormal. Visual methods alone do not allow the easy identification of any components representing, say, *S* and *O*. Transformation into natural logarithms, however, and graphical representation of the data on logarithmic probability paper are more rewarding. Fig. 4 shows two obvious component populations revealed by two straight sections of the curve. These are undoubtedly component, *S* with a mean grade of about 2 per cent, and *O* with one of 6-7 per cent Cu.

The above two grade estimates, together with visual estimates for the standard deviations and proportions of components *S* and *O*, were used for a two-component analysis on a computer terminal by the new method. As is shown in Table 2 (A), two components with mean grades of 1.3 and 6.0 per cent Cu were identified as *S* and *O*, respectively.

Any of the observed low-grade wall rock copper could be interpreted in two ways. It could be a simultaneous appendage to the stratiform orebodies, a part of component *S*. Alternatively, it could be a third component, *E*, an additional widespread ore phase created independently. Its existence is hinted at by the kink in the lower part of the plot in Fig. 4, but attempts to

quantify this postulated component, *E*, by graphical means have been in vain. A three-component computer analysis has, however, been partially successful (see results in Table 2 (B)). The existence of component *E* is indicated; with an average grade of 0.3 per cent it contributes only 1 per cent of the total contained copper.^[#] It may comprise both sulphide and resulting oxide ore, or sulphide ore alone (in which case a search could have been instituted for a relatively higher-grade fourth population, oxide in nature, and derived from *E*).

Other determinations confirm the above relative proportions of the sulphide and oxide ore. In their absence, however, the method could have been used to determine such proportions from total copper assays only.

Malaysian placer tin deposit

The Kinta Valley of west Malaysia is the most important, in extent and production, of all the placer tin localities in Thailand and Malaysia.^[17] The alluvium lies upon a bed-rock which, in profile, is either highly irregular limestone or more regular shales or granite. The north-south-trending valley is open in the south. In that direction the density of mining, past and present combined, by dredging and open-pit methods, decreases, whereas the average depth of alluvium increases.

The southeastern part of the valley is illustrated in Fig. 5. which shows the distribution of abandoned and existing tin mining leases. The whole valley floor is covered with alluvium, but the leases encompass only those parts which have been proved to contain sufficient cassiterite to justify working.

Near to, and against, the surrounding granite hills an eluvial-alluvial granite wash is an important tin-bearing constituent of the alluvium, and has often been called the 'old alluvium'.^[19] Elsewhere, it is relatively less frequent than younger stratified sequences of gravel, sand and clay. These horizons contain cassiterite, which generally increases in amount towards bedrock. They are overlain by later peat and barren material, and some-times by tailings from the previous operation of nearby mines.

Two major phases of younger alluvial deposition are easily identifiable in Fig. 5. The first resulted from a predominantly north-south-flowing drainage system. It was succeeded by one derived from an east-west drainage, the direction and approximate location of which are followed in part by the present-day drainage system.

Where the vertical and horizontal extent and tin grade derived from one phase alone have been sufficient, mining has proceeded. The geographical extent of the mining lease often mirrors the fossil drainage system responsible for the deposition of that phase. At greater distances from the tin source the grade of alluvium generally decreases. The physical superimposition of two or more phases is then necessary to provide the greater vertical thickness of tin-bearing ground to compensate for the lower grade.^[20]

Recognition of different tin phases is important in exploration planning in the selection of areas for banka drilling. Knowledge of their vertical and horizontal extent, and of their individual characteristics, is a vital stage in the evaluation and production planning of a property. The information allows the essential separate estimation of expected average in situ grade, digging recovery and metallurgical recovery, etc., for each phase.^[20]

An area of approximately one square mile within the limits of Fig. 5 has been banka-dnilled. and the new method has been used in an attempt to identify and quantify any distinct phases of alluvium which may exist. The results of 273 banka drill holes were available in the form of grades (in kati/yd³) of successive 5-ft vertical lengths from surface to bedrock. The deepest bedrock elevation attained was nearly 220 ft below surface. The individual 5-ft section grades were accumulated to produce frequency histograms for 20-ft^[x] vertical sections from surface to 220 ft. The first six sections from zero to 120 ft are shown in Fig. 6. They illustrate the increasing grade with depth and the tendency towards a lognormal distribution. It is uncertain, however, from either the histograms or the geographical situation, whether the drilled area contains cassiterite derived from one or more depositional phases. Nor does the graphical approach illustrated in Fig. 7 provide any assistance.

Use of the proposed method, however, allowed the identification of four phases, which one is barren — the results obtained are detailed in Table 3. The phases are illustrated diagrammatically in Fig. 8 within the context of an idealized section through the area.^[†] Phases I and II, probably related to the different fossil drainage directions, now can be searched for by reference to old banka-drilling records of adjacent and nearby areas to suggest further unbored extensions.

The recognition of such phases both prior to and during production is hampered by the fact that only in the working open-pit properties, unlike dredging operations, is the ground exposed and visible. While boring is in progress sufficient geological and mineralogical information may be systematically collected, but its interpretation is time-consuming and requires relevant experience. In the majority of cases, therefore, no more than old banka drilling records may be available — which state only the grades without any geological information. But such paucity of data need no longer prohibit an essential geological interpretation.

Cornish tin lode

Previous underground and mineralogical studies at Geevor mine, Cornwall, have revealed several phases of mineralization within the lodes.^[21] The economically important phases are those which have yielded cassiterite, usually in combination with a gangue mineral, such as chlorite, tourmaline, quartz or hematite. The geological observations were substantiated by graphical statistical methods.^[29] A distortion of the approximately lognormal^[23] metal content^[‡] distribution of a lode was interpreted to result from a combination of two phases — a high-grade tin-hematite phase super-imposed upon a more widespread, lower-grade, tin-chlorite-tourmaline phase. The reddish-brown hematite contrasted underground with the blue-green tourmaline and chlorite, allowing the distribution of the two phases to be mapped in detail, unless the hematite became too dominant and masked the entire lode with its coloration.

Simms lode, discovered in the early 1960s, is typical in that it contains more than one phase of tin mineralization. Like the other productive veins of the mine, it reveals a distribution, based either upon grade over a variable width (lb/ton over the lode width) or upon metal content, which approximates to lognormal.

The lode, now in the late stages of its development, is illustrated in Fig. 9. Considerable thought must now be given to the quantity and grade of ore which might be exposed by opening up additional, perhaps marginal, levels and by development at the present extremities.

The analyses were based upon 4550 development samples and 3010 stope samples. The lode is cut by two quartz-filled faults, known locally as 'guides' or 'cross-courses'. To the west it divides, most development having proceeded along the better of the two branches. This is the 'right-hand', indicated as (A). Simms lode is only slightly displaced by the western guide and is easily traced from (B) to (C) through the structure. The eastern guide, however, seriously dislocates and displaces the lode (between C and D). It has been suggested, in consequence, that the development to the east (D) and west (A, B, C) of the guide is upon two different lodes. The two lengths of lode admittedly differ somewhat in appearance under-ground. Traces of other ore have already been found by diamond drilling — a lode containing two visible mineralization phases, over an unknown extent, and with which section D could be synonymous. Therefore, any additional evidence for two different structures could justify an exploration programme designed to seek the implied undeveloped extensions.

The sampling data from each of the four previously mentioned sections were examined in turn. The most easterly, (D), was the suspect portion, which required examination to determine whether it was a simple faulted extension of the remainder or whether it showed different characteristics indicating it to be another lode. Such differences could be in the number of mineralization phases, the average grade of each, and the extent to which each contributed proportionally to the number of samples taken, and, thus, to the overall grade.

The results obtained are presented in Table 4. Within each section of the lode three comparable mineralization phases have been detected (I, II and III). Their dimensions and extent of representation may be appreciated visually in Fig. 10, which presents histograms of grade (over the variable lode width) in lb/ton plotted on a natural logarithmic base. The extent of conformity of mineralization indicates that section D is a faulted extension of the remainder rather than another lode. It is likely, however, from the resemblance between the distributions of sections A and D that the lode has branched in the vicinity of the eastern guide, and that section D, like section A, is the better of two or more branches of the lode.

Such an approach as described above has an immediate application in Cornwall by providing an additional method of possible identification of several lodes cut and separated by a major cross-course. Lodes could be identified either as veins limited in extent and lying on one side only of the cross-course or as faulted extensions of one already established on the far side, with any loss or gain of ore phases associated with the cross-course.^[24]

Acknowledgement

Permission to utilize and publish data involved in the three examples contained in this paper has been kindly provided by Geevor Tin Mines Ltd., and by various member companies of the Anglo American group. Computer facilities were made available by the Imperial College Computer Centre, London.

References

1. Aitchison J. and Brown J. A. C. Lognormal distribution (London : Cambridge University Press, 1957), 176 p.
2. Sichel H. S. The estimation of means and associated confidence limits for small samples from lognormal populations. In *Symp. Mathematical statistics computer applications in ore valuation* (Johannesburg: S. Afr. Inst. Min. Metall., 1966), 106-23.

3. Clark M. W. Methods for the analysis of complex distributions. Unpubl. manuscript, 1973, Kings College (Department of Geography), University of London.
4. Hald A. *Statistical theory with engineering applications* (New York: Wiley, 1952). 783 p.
5. Lepeltier C. A simplified statistical treatment of geochemical data by graphical representation. *Econ. Geol.* 64. 1 969, 538-50.
6. Bhattacharya C. G. A simple method of resolution of a distribution into Gaussian components. *Biometrics*. 23, 1967. 115-35.
7. Pearson K. Contributions to the mathematical theory of evolution. *Phil. Trans. R. Soc.* A185, 1894, 71-110.
8. Rao C. R. The utilization of multiple measurements in problems of biological classification. *Jl R. statist. Soc.* series B, 1948, 159-203.
9. Day J. E. Estimators for the parameters of a mixture of two Normal distributions. M.Sc. dissertation. University of Exeter, 1966.
10. Hasselblad V. Estimation of parameters for a mixture of normal distributions. *Technometrics*. 8.1966, 431-44.
11. Jones T. A. and James W. R. MAXLIKE: FORTRAN IV program for maximum likelihood estimation. *Geocom Programs* no. 5 in *Geocom Bull.* 5, 1 972, 186-202.
12. McCammon R. B. FORTRAN IV program for non-linear estimation. *Kansas Geol. Surv. Computer Contribution* 34, 1969, 20 p.
13. Mundry E. On the resolution of mixed frequency distributions into normal components. *Math Geol* 4, 1972. 55-60.
14. Lindley D. V. and Miller J. C. P *Cambridge elementary statistical tables* (London: Cambridge University Press, 1953), 36 p.
15. Draper N. R. and Smith H. *Applied regression analysis* (New York: Wiley, 1966), 407 p.
16. Mendelsohn F. ed. *The geology of the Northern Rhodesian Copperbelt* (London ; Macdonald & Co.. 1961), 523p
17. Ingham F.T. and Bradford E. F. Geology and mineral resources of the Kinta Valley, Perak. *District Mem. Geol. Surv. Malaya* 9, 1960, 347 p.
18. Harrison H. L. H. *Valuation of alluvial deposits* (London: Mining Publications, 1954), 308 p.
19. Newell R. A. Characteristics of the stanniferous alluvium in the Southern Kinta Valley, West Malaysia, *Bull. Geol. Soc. Malaysia* no. 4. 1971 15-37.
20. Garnett R. H. T. Unpublished Company reports. Associated Mines (M). Ltd., 1962-65.

21. Garnett R. H. T. Local mineral zoning in Geevor tin mine, Cornwall. *in* Symp. Problems postmagmatic ore deposition, Prague 1963. vol. 1, 91-6.
22. Garnett R. H. T. Distribution of cassiterite in vein tin deposits Trans. Instn Mm. Metall. (Sect. B: Appl. earth sci.). 75, 1966. B245-73.
- 23 Poyntz C. D. Use of statistical sampling techniques in mining Unpubl. M.Sc. thesis, Cranfield College of Technology, 1969.
24. Garnett R. H. T. Structural control of mineralization in South-West England. *Min. Mag.*. Lond.. 105, Dec. 1961, 329-37

FOOTNOTES FOR PAPER

[*]These figures refer to computing times: the actual working time, or 'real' time, is about 15-20 min.

[†]If the specimen size is too large, individual specimens will consist of ore from more than one component phase, thus preventing recognition. Analysis must then proceed with smaller-size specimens to permit the isolation of the components.

[‡]The example provided by Fig. 1(a) has been chosen by the authors to permit easy visual estimation for those unfamiliar with these techniques.

[#]Complete analysis of all extracted drill core exhibiting even trace values only would probably demonstrate the more widespread existence of the third phase, and the average grade would decrease

[§]The grade is measured locally in terms of katis per cubic yard. 18 0.1 kati/yd³ is equivalent to 28.5 g/m³.

[¥]any other section interval, provided that it was a multiple of 5ft, could have been used. Subsequent work could comprise narrowing the section down 15, 10 and eventually 5ft, in combination with an area search.

[†]The actual bedrock is considerably more complex.

[‡]The width of a lode is expressed in terms of inches, W, and the grade is determined after channel sampling over the same width with a vaning shovel assay. It is quoted as A lb SnO₂ per ton, or as A lb/ton. The total tin metal content is expressed as A lb/ton over W in, or as WA/12 ft lb/ton.

FIGURES FROM PAPER

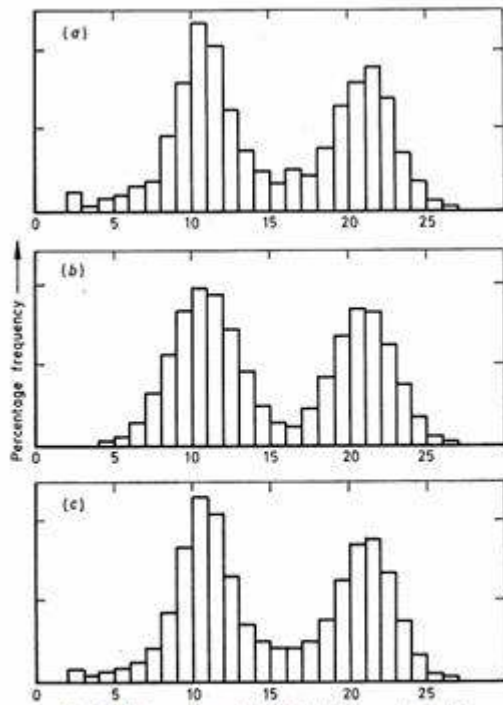


Fig. 1 Histograms of simulated data and models:
(a) original population;
(b) best-fit estimates of two-component analysis;
(c) best-fit estimates of three-component analysis

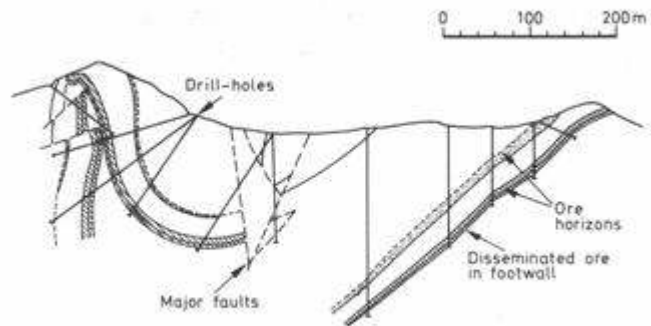


Fig. 2 Typical section through Central African copper deposit

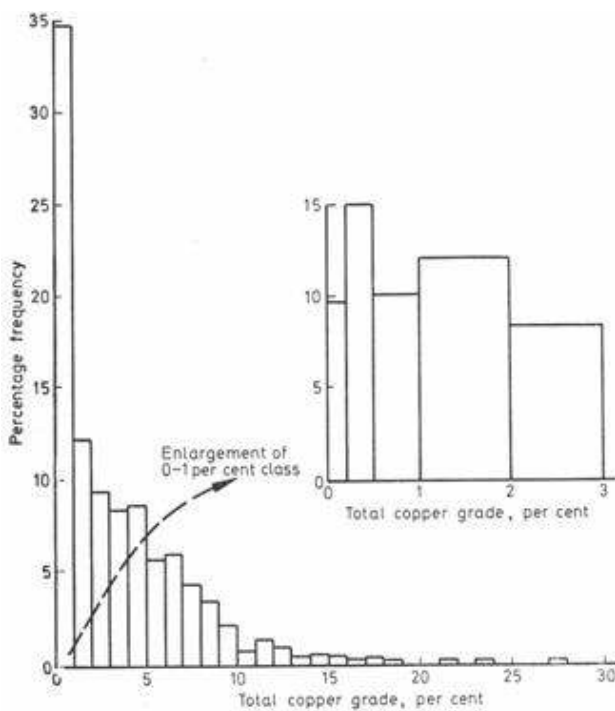


Fig. 3 Central African copper deposit: histogram of grades

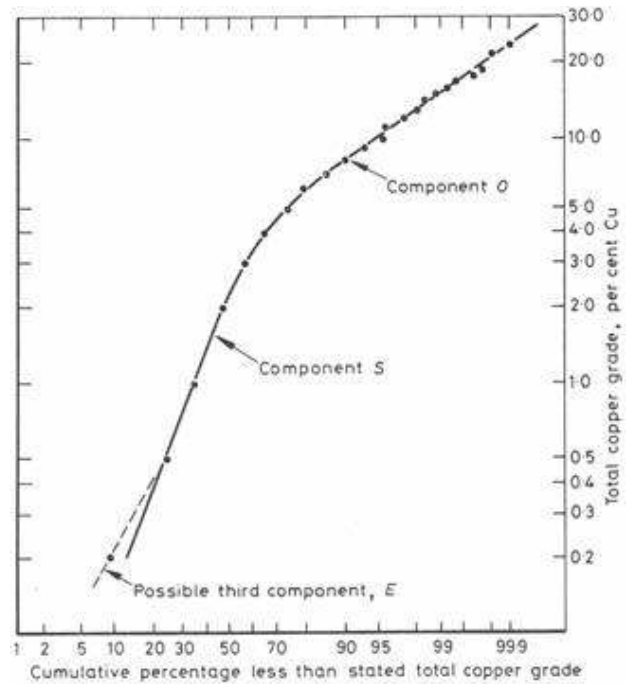


Fig. 4 Central African copper deposit: grade distribution plotted on logarithmic probability paper

butes only 1 per cent of the total contained copper.* It may comprise both sulphide and resulting oxide ore, or

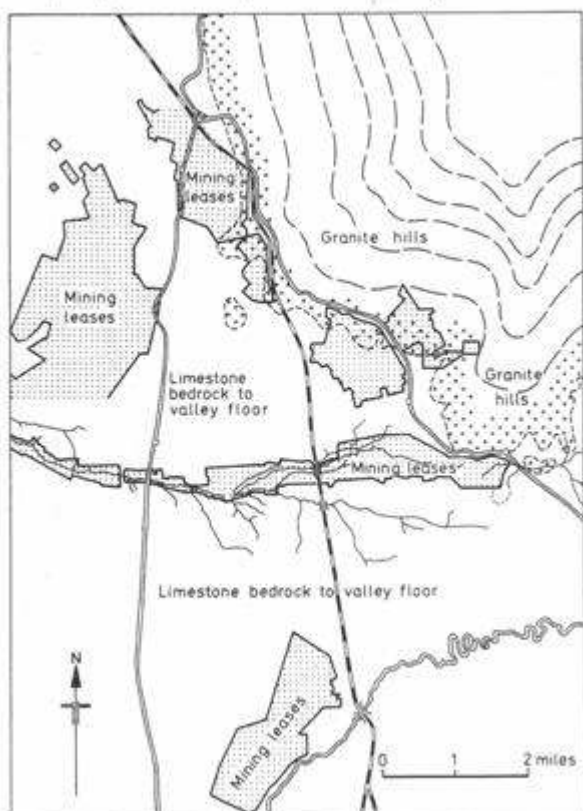


Fig. 5 Malaysian placer tin deposit, southeastern part of Kinta Valley

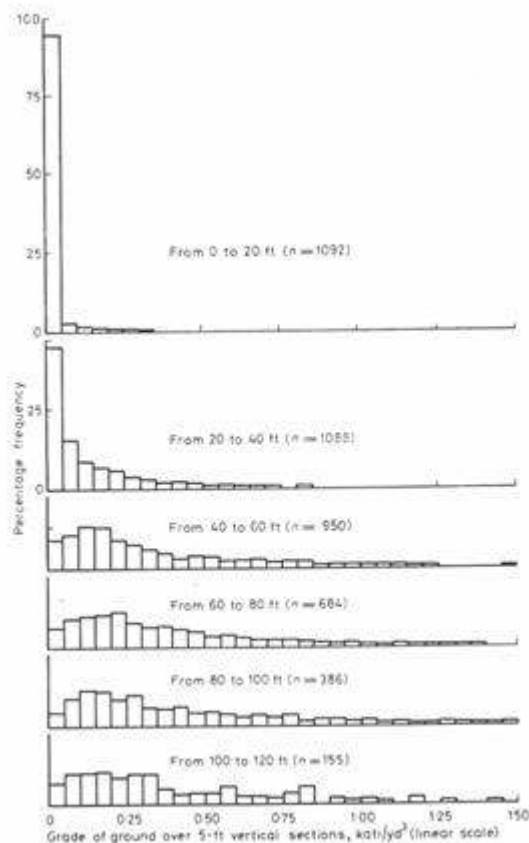


Fig. 6 Malaysian placer tin deposit: histogram of grades from increasing depths

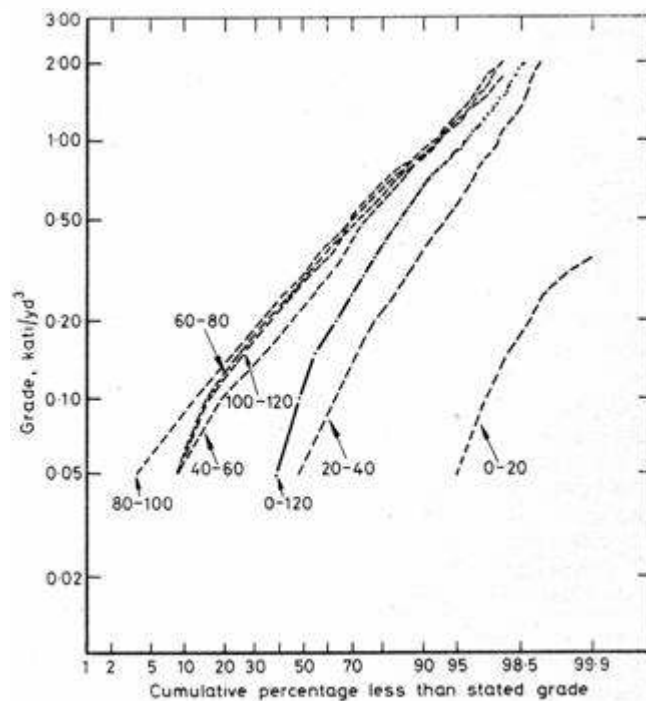


Fig. 7 Malaysian placer tin deposit: grade distributions plotted on logarithmic probability paper

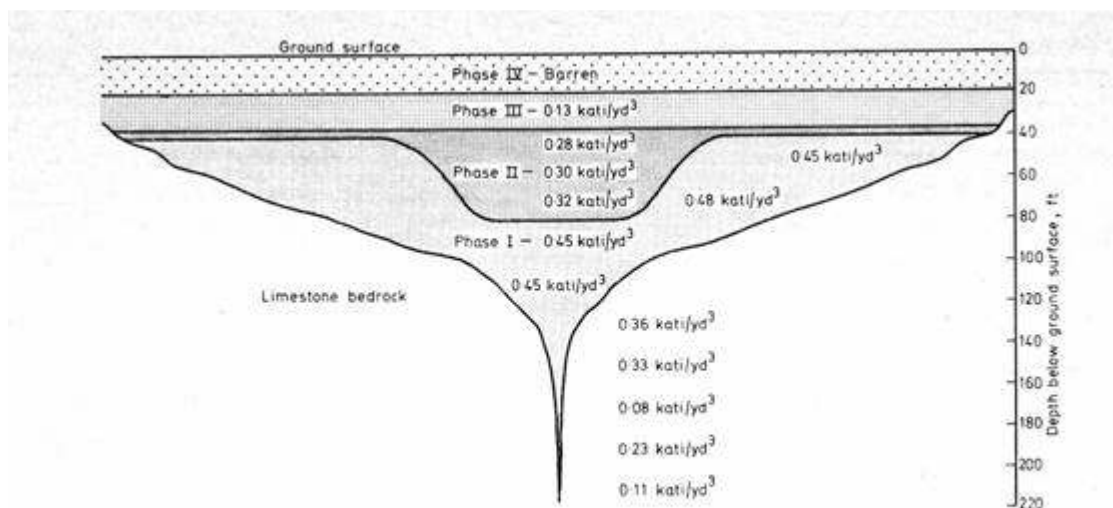


Fig. 8 Malaysian placer tin deposit: idealized vertical section

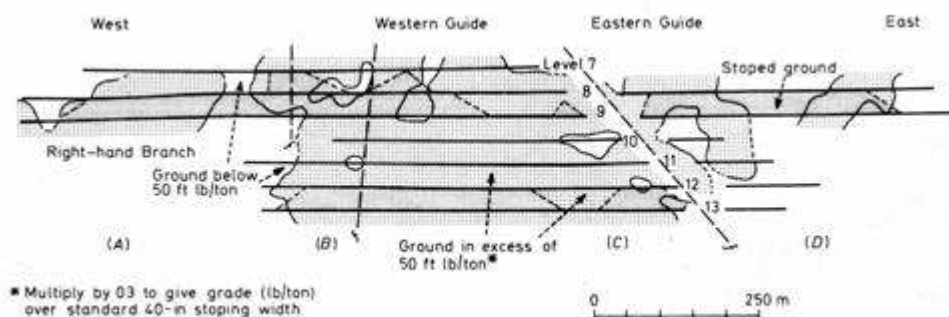


Fig. 9 Cornish tin lode: longitudinal section of Simms lode showing developed and stoped extent

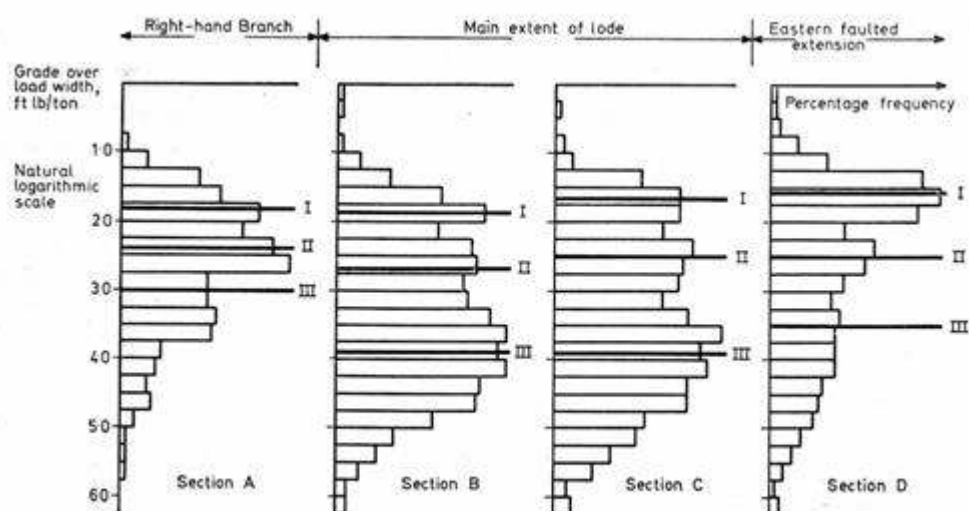


Fig. 10 Cornish tin lode: model grade histograms

TABLES

Table 1: Two- and three-component analysis of simulated data

Type of analysis	Type of estimate	Component	Mean	Standard deviation	Percentage	Chi-squared value	Degrees of freedom	Illustrated in Fig.
Two-component	(A) Initial	I	10.5	2.33	50	Not applicable	15	1(b)
	visual	II	21.0	2.00	50			
	(B)	I	10.81	2.17	55	85.8		
	Best fit	II	20.91	2.00	45			
Three-component	(C)	I	10.81	2.17	30	Not applicable	15	1(c)
	Initial	III	11.00	4.00	25			
	visual	II	20.91	2.00	45			
	(D)	I	10.73	1.34	34	7.6		
	Best	III	12.91	5.20	29			
	fit	II	21.20	1.72	37			
Original	(E)	I	11.0	1.5	35	Not applicable	15	1(a)
population		III	13.0	5.0	25			
		II	12.0	1.5	40			

Table 2 Two- and three-component analyses of Central African copper deposit

(A) *Two-component mineralization* (Results of an attempt to identify two components with a combined grade of 3.42 per cent Cu and standard deviation of 3.54 per cent)

	Sulphide ore,	Extra phase,	Oxide ore,
Ore phase	S	E	O
Average grade, per cent Cu	1.3	Not	6.0
Standard deviation, percent Cu	2.5	search-	3.3
Proportion of samples, per cent	55	ed for	45
Proportion of contained copper, percent	21		79
Degrees of freedom		13	
Chi-squared value		21.4	

(B) *Three-component mineralization* (Results of an attempt to identify three components with a combined grade of 3.42 per cent Cu and standard deviation of 3.54 per cent)

	Sulphide ore,	Extra phase,	Oxide ore,
Ore phase	S	E	O
Average grade, per cent Cu	2.0	0.3	6.1
Standard deviation, per cent Cu	2.9	0.4	3.3
Proportion of samples, per cent	41	16	43
Proportion of contained copper, per cent	23½	1	75½
Degrees of freedom		10	
Chi-squared value		26.1	

Table 3 Malaysian placer tin deposit

	Number of assays, 5-ft sections	Average grade, kati/yd3	Phase detected	Phase	Mean grade, kati/yd3	Standard deviation, kati/yd3	Proportion of samples, per cent	Proportion of combined grade, per cent	Degrees of freedom	Chi- squared value
0-20	1092	0.01	2	IV III II I	0 0.13	0.08	92 8	0 100	3	3.8
20-40	1088	0.16	2	IV III II I	0.13 0.28	0.46 0.22	80 20	65 35	14	7.4
40-60	950	0.40	2	IV III II I	0.30 0.45	0.50 0.50	33 67	25 75	22	40.3
60-80	684	0.43	2	IV III II I	0.32 0.48	0.56 0.42	34.00 66.00	25.00 75.00	21.00	16.2
80-100	386	0.45	1	IV III II I	0.45	0.49	100	100	21	21.2
100-120	155	0.45	1	IV III II I	0.45	0.55	100	100	12	12.7
120-140	46	0.36	}	insufficient samples to justify an analysis						
140-160	15	0.33	}							
160-180	4	0.08	}							
180-200	4	0.23	}							
200-220	3	0.11	}							

Table 4 Cornish tin lode

Section (see Fig. 9)	A			B			C			D		
Average grade (over lode width), lb/ton	23.75 3			52.33 3			59.95 3			41.89 3		
Mineralization phases revealed												
Phase	1	11	III	I	II	III	I	II	III	I	II	III
Average grade, lb/ton	7	12	32	7	15	72	6	13	81	5	12	68
Standard deviation, lb/ton	2	1	39	4	4	75	2	3	99	2	1	123
Proportion of samples, per cent	25	10	65	25	6	69	19	10	71	37	5	58
Proportion of combined grade, per cent	7	5	88	3	2	95	2	2	96	5	1	94
Number of samples used	1200			1227			2833			1585		
Degrees of freedom	11			13			13			15		
Chi-squared value	26.1			18.0			40.8			28.5		

MATHEMATICAL APPENDIX

Appendix 1

Brief explanation of method

Consider the model of a population made up of a mixture of two Gaussian (normal) components. The probability density function would be given by

$$P(x; m_1, s_1, \rho, m_2, s_2) = \rho \cdot f(x; m_1, s_1) + (1 - \rho) \cdot f(x; m_2, s_2) \quad (1)$$

where $f(x; m_i, s_i)$ is the probability density function of a normal distribution with mean m and standard deviation s . That is

$$f(x; m_i, s_i) = \frac{1}{s_i \sqrt{2\pi}} \exp \left\{ -\frac{(x - m_i)^2}{2s_i^2} \right\} \quad (2)$$

The form of the model used in the analysis is the cumulative probability distribution function F , where

$$F(z; m_1, s_1, \rho, m_2, s_2) = \int_{-\infty}^z P(x; m_1, s_1, \rho, m_2, s_2) dx \quad (3)$$

That is, $F(z; m_1, s_1, \rho, m_2, s_2)$ is the probability that a sample drawn from this population will be less than z . For simplicity in the following equations let us write t_1 for w_1 , t_2 for s_1 , t_3 for p , t_4 for m_2 , and t_5 for s_2 . For the complete list of parameters we shall write $\mathbf{t}$. Then equation 3 becomes

$$F(z; \mathbf{t}) = \int_{-\infty}^z P(x; \mathbf{t}) dx \quad (3a)$$

Let there be n groups in the histograms. Let the upper and lower bounds of the histogram be $+\infty$ and $-\infty$, respectively. Let the intervening end-points of groups in the histogram be denoted by $z_1, z_2, z_3, \dots, z_{n-1}$. Let the proportion of samples observed to lie below end-point z_i be denoted by y_i . We wish to find values of $\mathbf{t}$ which show the closest agreement between the model F and the observations y . We must, therefore, minimize the sum of squares S , where

$$S = \sum_{i=1}^{n-1} (y_i - F(z_i; \mathbf{t}))^2 \quad (4)$$

If the model were a linear function of the values in $\mathbf{t}$, we could solve equation 4 by ordinary least squares methods. Since it is not, we must use the iterative non-linear least squares. Suppose that we make a sufficiently accurate guess at the values of $\mathbf{t}$, say $\mathbf{t}_0$. We can now improve this estimate by an amount, say $\Delta \mathbf{t}$ by the Gauss-Newton method. Let us denote the partial differential of function F at point z_i with respect to parameter t by $dF(z_i; \mathbf{t})/dt_j$. Let us define a vector $\mathbf{g}$ whose five elements are denoted

$$g_j = \sum_{i=1}^{n-1} \frac{\partial F(z_i; t)}{\partial t_j} (y_i - F(z_i; t)) \quad (5)$$

Let us also define a five by five matrix **D**, which has as its elements

$$d_{jk} = \sum_{i=1}^{n-1} \frac{\partial F(z_i; t)}{\partial t_j} \frac{\partial F(z_i; t)}{\partial t_k} \quad (6)$$

Then Δt , the improvement in the parameters, is found by solving the system of five simultaneous equations denoted by

$$D \Delta t = g \quad (7)$$

This process is then repeated with the new values of t until no further improvement in S can be found in that region. The present authors have used a modified version of the Gauss-Newton method for faster convergence. This involves taking different multiples of Δt and choosing that nearest the minimum of S . The multiples used here were 1, 1/3, 1/9, 1/27 and 1/81.

Appendix 2

Calculation of lognormal parameters

If x has a lognormal distribution with mean λ and standard deviation ω , $y = \log(x)$ has a normal distribution with mean m and standard deviation s , where

$$\lambda = \exp \left\{ m + \frac{1}{2}s^2 \right\}$$

$$\omega = \lambda \sqrt{(\exp(s^2) - 1)}$$

If component 1 is found to have mean m_1 , and standard deviation s_1 , when analysing for the 'underlying' normal distribution, component 1 actually has mean

$$\lambda_1 = \exp \left\{ m_1 + \frac{1}{2}s_1^2 \right\}$$

and standard deviation

$$\omega_1 = \lambda_1 \sqrt{e^{s_1^2} - 1}$$

Discussions and contributions

Identification of multiple mineralization phases by statistical methods

I. Clark M.Sc., F.S.S.

R. H. T. Garnett Ph.D., M.B.A., C.Eng. M.I.M.M.

Report of discussion at October, 1974, general meeting (Chairman: K. C. G. Heath, President) and contributed remarks. Papers published in Transactions/Section A (Mining industry), vol. 83, 1974, pp. A43-52, A53-62 and A79-84, respectively

Mrs. I. Clark, in introducing her joint paper, said that its main purpose was to outline an approach by which it was possible to use histograms of ore grades to detect and identify multiple phases of mineralization within an ore deposit. It was often desirable to gain an overall picture of a deposit from the sample information gathered during exploration or development, and the simplest way to do that was by construction of a frequency diagram, or histogram, of the ore grades. Although that type of summary of the data ignored the spatial distribution of the samples, it could provide valuable information about the frequency distribution of grades — information which was necessary for the application of a wide range of statistical techniques, classical or otherwise.

Most statistical approaches, old and new, to the evaluation of ore deposits had been based on the assumption that the frequency distribution of ore grades could be described by a simple probability curve, such as a normal or lognormal. It was the authors' experience that where a deposit had been formed by a single phase of mineralization, the histogram of sample grades could usually be assumed to come from such a simple distribution. Where a deposit — or part of a deposit — had undergone multiple mineralization, or considerable reworking, however, the histogram of sample values was modified to such an extent that it could no longer be adequately characterized by a single, simple curve.

Figs. 1-3 showed how two normal curves might combine to produce a complex distribution. In Fig. 1 a mixture of two component distributions whose means were a considerable distance apart combined to form a complex and obviously bimodal curve. Fig. 2 gave the same two distributions, in the same proportions, but with the means considerably closer together producing a complex curve in which the lower peak was no longer obvious. The shape of the curve was, however, still indicative of a mixture, and a plot of such a distribution on normal probability paper would clearly indicate the presence of two components. Fig. 3 gave the same two distributions, in the same proportions, but the means were so close together that the combined curve appeared to be unimodal and highly skewed. A plot on probability paper would indicate that the curve was decidedly non-normal, but would not necessarily indicate two components.

Suppose that there was geological evidence of multiple mineralization, or of reworking of a deposit, and that the histogram of the ore grades did not conform to a simple unimodal distribution. The problem that faced the analyst was twofold : to identify how many components to expect in the overall distribution and then to characterize each of the component distributions by a single curve. The first of those, the number of components to search for, must be based mainly on the geological knowledge of the deposit. It was possible to analyse for different numbers of components, and then to decide, on the basis of, say, a chi-squared test, which mixture described the deposit best. Probability plots were also very

useful in choosing the number of components, especially if there were sufficient separation of the means.

The task of splitting the complex distribution curve into its components had been considered by many authors in the past, and a review of that work would constitute a paper in itself. The present paper was introduced to offer a new method which had advantages of speed and accuracy over previously available computer-orientated methods. Being computer-orientated, the method did not claim to replace existing graphical techniques of estimation — indeed, since it was necessary to provide initial estimates of the parameters involved, it was perhaps desirable to use a graphical method to provide the first approximation for the computer. Although the method required a computer for implementation, that need not be a large computer. A FORTRAN IV program had been developed to run on a 24K mini-computer at the Royal School of Mines, London, and would solve for four component populations. It would also be possible to implement the method on any desk computer, or programmable calculator, which was capable of solving relatively small sets of simultaneous equations. For example, a three-mode analysis required the solution of a set of only eight equations.

The potential advantages of the new method, therefore, were that it was very simple to use, even for someone unfamiliar with the underlying processes of the analysis, that the technique was fairly stable, usually reaching an optimum solution even when the initial estimates for the parameters were grossly inaccurate, and that that method was fast and economical in computer time.

Those were, however, also the potential disadvantages of the technique. It might be too simple to use and too easy for the user to search for more components than were strictly necessary. To a certain extent, the method itself safeguarded against that in that a search for more components than were actually justified by the data usually led to an unstable condition, or to one in which the fit to the data was worse than for fewer components.

To summarize, the main uses of the technique seemed to be threefold : to identify and characterize various phases of mineralization within a deposit, and perhaps enable partitioning of the deposit into more homogeneous sub-areas; to enable the application of conventional (classical) significance tests and confidence limits to such a complex distribution by splitting it into more manageable component distributions; and to aid in the simulation of ore deposits by both classical and geostatistical methods, since all such simulations must be based on the underlying frequency distribution of ore-grade values.

Dr. T. L. Thomas said that he was extremely interested in the paper by Garnett and Clark, since Mrs. Clark was a colleague at Imperial College, and he had worked on associated problems with Dr. Garnett over a number of years.

The present paper was of importance since it brought together a practical geologist and a practical statistician. Dr. Garnett had known for many years that irregularities in the distribution curves, particularly of tin samples, were significant and indicated phases in the mineralization. When fitting a lognormal curve to such distributions, the chi-squared test had shown that the values differed significantly from the simple lognormal model. Mrs. Clark's system of using the method of least squares to fit a multiple lognormal model brought significant improvements. The value of her new method was in its simplicity. It converged rapidly and the equations could be solved on a mini-computer, or even on a desk-top computer, since large numbers of linear equations were not involved.

He hoped that many mining engineers would read the paper, since it had been written in a manner which could be understood by the non-mathematician, and the mathematical principles involved were placed at the end.

He would like to congratulate the authors of the second paper on producing some practical results of a geostatistical investigation : it was interesting to note that the semi-variograms obtained were by no means straightforward, thus limiting the range of a solution which assumed a simple spherical scheme. The authors suggested, but did not apply, a multiple spherical model, and he would like to ask Mrs. Clark whether her method could be modified to obtain the best fit of such a model to a practical semi-variogram.

Dr. M. Guarascio said that the method proposed by Clark and Garnett was based on the assumption that the simple specimens came from one of the component distributions and were not a mixture of both. If the geological evidence had permitted a clear separation between two or more sorts of ore (e.g. oxide and sulphide), there were no reasons for putting the two different sets together with the aim of separating them again numerically. If, instead, it had not been possible to differentiate the specimens by means of geological observation (the zones from which the specimens came), it was likely that each specimen was a mixture of several components. In such a case the proportion given by the method might not be representative of the in-situ situation.

With regard to the use of the sample histogram in ore-reserve estimation problems, there was a risk of possible biased conclusions because the sample histogram was not a completely satisfactory representation of the in-situ spatial distribution of the ore in that it did not take into account the spatial location of the different types of ore. which was the main factor in ore selection.

A. G. Royle said that there had been one point which had interested him very much in the Clark-Garnett paper: that was with regard to the tin deposit where a succession of assays from a large number of boreholes had been grouped into 20-ft vertical lengths, and from those the deposits had been reconstituted. He had made an interesting study in which a deposit had been sliced into horizontal slices, and he had taken the variograms in each slice. There had been some extremely strong anisotropies. There was a deep layer with a range of 80 units in the east-west direction, 74 in the north-south direction and one vertically. In the middle section the range was 80 units east-west, but only 16 in the north-south direction and again one vertically. The data set in that case was excellent, and it was possible to determine the ranges in different directions.

P. G. Linzell said that alluvial deposits were very erratic in their nature, and even more so when one was handling diamonds. It was a fact that where grain sizes were predominantly less than 1 mm, diamonds were never found. Those were apparently present only in the coarser material. That applied to one part of the world where he had had experience of alluvial deposits containing diamonds.

If people started to apply that statistical analysis to such deposits, he felt that they would need to do so on the basis of grain size. He would be interested to know what effect that had had on the work done so far.

Professor E. Cohen emphasized that the subjects under discussion were models which could be only as good as the geological information on which they were based. How closely they represented the orebodies often remained conjectural until confirmation by mining experience

had been obtained. Equally, extrapolation from one deposit to another was not always relevant. For example, reference had been made in the discussion to the absence of -1 mm diamonds from alluvial deposits. Yet the original kimberlite from which the alluvials were derived usually contained much greater numbers of small diamonds than larger ones, in sizes well below 1 mm. One might conclude that the processes of formation of alluvial deposits had caused the destruction of small diamonds, which was most unlikely. More plausibly, the small stones were deposited elsewhere, due to the size classification effects of alluvial transport. Appropriate factors would thus have to be built into any model of alluvial diamond gravels, account being taken of various combinations of depositional conditions. With different geographical circumstances those factors would suffer changes in significance and, of course, they would be quite irrelevant in assessing kimberlites or residual deposits close to kimberlite where little transport had occurred.

Thus, one had to depend on the complete build-up of geological and mineralogical information, and there was great risk of constructing a statistical model on an inadequate basis. He did not wish in any way to detract from the importance of the work that had been presented by the authors but to emphasize that the development of such methods placed additional responsibility on the geologist.

Professor R. N. Pryor commented that the Clark-Garnett paper had been presented in an admirable way, details of the research and its application being given. The statistical part had been put over in quite simple terms and the theory had been put to the test in three examples. That was all reported with clear diagrams and he felt that the paper constituted a very valuable contribution to the Transactions.

He had a question: how should people set about using that technique? When the technique had been discussed at the Royal School of Mines, he had thought that it might be useful for mining engineers, but now he was more inclined to think of it as a tool for the geologist to learn more about the geology and to assist in predictions of genesis, etc.

The Chairman said that when they were making use of mathematical models to try to deduce the nature or behaviour of a mineral deposit (or for that matter a mining or metallurgical operation, or a cash flow) it was essential that they kept firmly in mind what was the reality and what was the simulation. They should not, for example, in their enthusiasm for the elegance of the model reject what did not fit comfortably. It was a little disturbing to read in one of the papers that some trend was caused by a change in the characteristics of a semi-variogram, and therefore of the deposit, in some direction. He would have thought that it was caused by a change in the characteristic of the deposit affecting the semi-variogram.

The three examples given by Clark and Garnett did not seem to him to be on all 'fours'. To take the last first, what they might perhaps call the 'case of the Cornish tin lode' seemed to be a piece of pure detective work — simple enough when they knew the answer. They might suspect that there were geological as well as statistical clues to be found. In the second example, the Malaysian tin deposit, geological information was expressly excluded on the grounds that old bore records might be the only source of information. In the first example, the Central African copper deposit, it would seem that there must be geological (or, rather, mineralogical) information that would enable the sample population to be split into two, predominantly sulphide and predominantly oxide, from the start. The procedure described could then be applied with advantage to each part separately.

He was sure that all would be grateful to the authors for describing so clearly the ingenious logic used to obtain their results. They had added to the value of the paper by their description of practical applications.

Dr. P. S. B. Stewart said that he was interested in the sort of problem posed in the first paper. One tried to separate information which had been boiled up together. Looking at Table 1 (p. A46), he had been a little disappointed with the precision with which the original population had been recovered by the procedure. He was not sure whether that was due to the original population of assays being represented just as a histogram or whether the authors had taken a sample from the original population. It seemed to him that if the trial data represented the whole of the original population, the mathematical procedure should have recovered original parameters precisely.

If the histogram actually represented a simulated sample of the population of assays, it was a rather ideal sample, and he wondered how the stability and precision of the mathematical procedure would have been affected if a smaller or less ideal sample had been used or a simulated error applied to the histogram before attempting to recover the original distributions from it.

Contributed remarks

Dr. A. J. Sinclair: Clark and Garnett are to be congratulated on the clarity with which they have presented their new approach to extracting component populations from a data set. Elsewhere, such procedures have been called partitioning (e.g. Harding^[2]). Functionally, the new method appears sound. Their aims in applying it to partitioning density distributions of assay data are identical to the aims of others who thus far have used different methods.

It should be emphasized that the new method, in common with other partitioning procedures in general use, relies on assumptions concerning the form of density distributions of component populations. There is an abundant literature indicating that numerous minor elements commonly follow a lognormal law fairly closely (e.g. Shaw^[5]). It would be ludicrous, however, to suggest that any particular group of data is exactly lognormally distributed! Herein lies a problem with any precise, statistically based, partitioning method. Does a data set comprised of approximate lognormal distributions necessarily warrant analysis by a procedure that extracts ideal lognormal distributions in a highly sophisticated and exact manner? After all, the partitioned populations are, at best, approximations of the real distributions. The question is an important one, to which there is not always a clearcut answer. It is apparent that the desirability of a precise technique rests with the quantity and quality of the data to be analysed. Quality is important because low reproducibility in sampling and/or chemical methods leads to a smoothing that in some cases can mask completely the presence of two or more modes in a complex distribution. Quantity is important because data must be adequately representative of the populations under study.

The major advantage of the new partitioning procedure is its speed and reproducibility. By the same token, its principal disadvantage lies in the necessity of partitioning being done on a computer, thus restricting its use. The new method has an obvious advantage in the analysis of abundant high-quality data for which the nature of the component distributions is known (i.e. lognormal, normal, etc.), particularly if three or more component populations are present. As with all techniques, the application to real problems involves a high degree of subjectivity,

both in specifying the number of populations present and in the ultimate interpretation of their relevance.

The writer is a supporter of the use of probability plots as a preliminary approach to partitioning problems and has found this graphical method adequate for a wide variety of geochemical data,^[6] including assay information. The specific variable to be studied in the case of assay data can be either metal per cent (or some comparable measure of proportion) or metal accumulation (grade x distance), as the case demands. The importance of recognizing which type of variable to use in a particular study is obviously recognized by the authors, but it is not emphasized enough for those unfamiliar with the theory of regionalized variables (see, for example, Matheron^[4] and Journel^[3]).

I must disagree with some statements made by Clark and Garnett regarding their Fig. 4 (p. A47). The two straight-line segments of their curve are not what reveals the possible presence of two populations. Instead, it is the inflection point located approximately at the 40 percentile that shows the probable existence of two populations. On the assumption that these populations are lognormal, they can be estimated graphically by use of the partitioning procedures implied by Harding^[2] and outlined by Sinclair.^[6] There is absolutely no indication of the suggested possible third component, E, in Fig. 4. Their statements appear to be based on a misconception as to the significance of patterns on probability graphs, a good introduction to which was given by Belviken.^[1] Fig. 5 shows the probability graph of Clark and Garnett partitioned according to Harding's procedure. To attempt to extract more information from the lower end of the curve without changing the grouping intervals (bar intervals) is to overinterpret! The plot does show, however, the possible existence of one or more populations in the upper 2 per cent of the data, although they would be difficult to partition on the basis of the data as presented.

An ideal recombination of the two partitioned populations in the proportion 60 per cent 0 and 40 per cent S produces almost exact coincidence with the real data curve. The only departure of the ideal from the real situation occurs close to the 50 percentile and is illustrated by the positions of triangles (points on ideal curve) relative to the original data curve. This difference between ideal and real distributions could result from a number of causes. The real distribution might not be exactly lognormal, for example, or the upper population, as partitioned, might be too crude an approximation of data that appear to consist of several component populations. Furthermore, some of these sub-populations that make up the 0 population might be normal rather than lognormal — a not uncommon situation in the percentage range being considered.

One aspect of the interpretation of populations as representative of stages of mineralization deserves detailed consideration: that is, the common implicit assumption, not always met in nature, that each value contributing to a histogram represents one and only one population. Consider a hypothetical example: a mineral deposit is formed by two stages of mineralization affecting two zones that overlap partially in space, each zone characterized by its own level of metal abundance (Fig. 6 (a)). Assuming no trends, a random sampling of the mineralized zones will give rise to a histogram of assay values such as that shown in Fig. 6 (b).

It is apparent that each value in the zone of overlap will be the sum of contributions from populations I and II and that these combined values represent a third population, III. In practice, the amount of overlap, representativeness of sampling and differences in metal abundances arising from each mineralization stage determine the number of populations that will actually be recognized. Fig. 7 shows diagrammatically how two periods of

mineralization can lead to the existence of one, two or three recognizably separate populations of metal abundance. Complications of this sort could lead to added complexity in dealing with correlation problems of the type represented by the Geevor example of Clark and Garnett.

In summary, the main advantages of the new partitioning technique lie in its statistical base. In applications to real problems involving assay information the technique suffers the same difficulties of subjectivity in interpretation as do other partitioning methods.

References

1. Bolviken B. A statistical approach to the problem of interpretation in geochemical prospecting. In *Geochemical exploration* (Montreal Canadian Institute of Mining and Metallurgy. 1971). 564-7. (CIM spec. vol. 11)
2. Harding J. P. The use of probability paper for the nryph»ca analysis of polymodal frequency distributions. *J. marine BioL Assoc.* 28, 1949, 141—53.
3. Journel A. Geostatistics and sequential exploration. *Min. Engng*, N.Y., 25 Oct. 1973,44-8.
4. Matheron G. The theory of regionalized variables and its applications. *Cah. Centre morphoL Math. Fontainebleau* no. 5, 1971. 211 p.
5. Shaw D. Element distribution laws in geochemistry. *Geochim cosmochim. Acta.* 23, 1961, 116-34.
6. Sinclair A. J. Selection of threshold values in geochemical data using probability graphs. *J. Geochem. Expl.*,3, 1974 129-49.

B. W. Hester Assays are of interest to the economic geologist on two counts. Foremost, and of immediate concern, is the basis they provide for evaluation of a mineral deposit. Second is the understanding they provide of the distribution of the valuable mineral within the deposit. All too often, the engineering problems of the evaluation eclipse proper consideration of the contributions the assays can make to solving the geological problems.

Clark and Garnett are to be complimented on devising an elegant, readily usable method to aid comprehension of the metal distributions represented by arrays of assay results under this latter heading. It is surprising, however, that they do not seek to substantiate the conclusions they reach for the three examples by using all the geological information available or supplementing their interpretations by applying widely used statistical techniques which are independent of their proposed method.

Just because a population of assays does not conform to the normal distribution law does not necessarily imply it must be lognormal. In any event, when dealing with very low metal contents, such as hard-rock gold deposits and placers of all kinds, the regression effect, as described by Krige,^[1] can be of sufficient size to distort any underlying law of distribution. Evidence for the presence of this in placer tin deposits of Malaysia was given by Broadhurst and Batzer^[2] and in placer gold deposits by Hester.^[3] The authors seem not to have

investigated the influence that this effect might have had on the results of their analysis, but it would surely be possible to correct any data as a first stage in employing their method.

Very often, frequency plots of assay data give the visual appearance of being distributed lognormally. When the same data are plotted in cumulative form, the result is similar to the result shown in Fig. 4 of their paper. From this graph the 'location constant' can be computed, which, when added to the assay before taking the logarithm, reduces the plot of the data to an approximation to a straight line. In the case of Fig. 4, inspection suggests an approximate value for this constant of 1-5. This artifice is a standard tool in the graphical treatment of suspected lognormal distributions of the simpler types, and I wonder if the authors chose to apply it first before using their method and, if so, with what result. Inspection of Fig. 4 suggests that the addition of the appropriate constant might reduce the graph to a single straight line, or, at least, a smooth curve. Should this be so, it could change the authors' conclusion that two phases of mineralization are present.

Evidence casting doubt on two such phases in partly oxidized copper deposits comes from the Mons Cupri area of the Pilbara district of Western Australia, where a body of low-grade copper sulphide ore has been sampled by closely spaced diamond drill holes. The primary mineralization consists essentially of disseminated chalcopyrite with minor pyrite in a suite of acid volcanic rocks. Oxidized ore is characterized by the presence of copper carbonates associated with voids. When the cumulative frequencies of logarithms of assays from both oxidized and primary ore are plotted together, the resulting graph has two straight-line components connected with a curve just like that of Fig. 4. By following the procedure of adding a constant to the assay first, as discussed above, a single straight line results. Curiously, a similarly shaped curve results from plotting separately assays of each type of ore. There is no other reason for supposing any second component of mineralization process to be present, and there is every indication of the assays being distributed in a random manner. If this conclusion is correct, might not the same be true for the authors' example? It should be a simple matter in the case of their example to plot the assays of oxidized ore separately from the primary to test which of the possibilities theses is correct.

Jones and Beaven^[1] presented the frequency distribution of assay data from a placer tin deposit in Thailand, but were unable to define with certainty any underlying law governing the distribution. Clearly, an understanding of the distribution of tin within deposits of this type presents problems of unusual difficulty, and the authors do well to attempt this task with their example.

Fig. 5 (p. A48) shows no geographical name—neither is the site of the example identified. Both of the authors' references 17 and 19 include very similar maps, which show this illustration to be of the Kampar area. These maps also show extensive gravel pump mining outside the mining leases shown in Fig. 5. From this, it would seem that a qualification is needed to the authors' statement (p. A48) that 'the leases encompass only those parts which have been proved to contain sufficient cassiterite to justify working'.

Reference 19 of their paper contains only an allusion to the term 'old alluvium'. Contrary to the impression given in the paper neither this term nor that of 'young alluvium' is introduced or proposed by Newell: rather, a more detailed division of the alluvial sequence was proposed. Newell's main conclusion, based on a careful field and statistical examination of the alluvial section, was that the sequence is divisible into three tin-bearing and one barren unit. This substantiates the interpretation of the present authors and it is surprising they do not invoke this very tangible support for their conclusion.

Evidence for the two directions of drainage systems to which the authors ascribe the two major phases of younger alluvial deposition is not at all clear in Fig. 5. Newell made no note of these directions. Gobbett^[5] noted alluvium-filled valleys draining north-west and southwest in the Kampar area. He related these directions to faults in the underlying bedrock. Isopachs of thickness of alluvium were presented in support of his interpretation. Evidence for the authors' bare statement on the drainage directions in the area covered by Fig. 5 is clearly needed, as it conflicts with these grievously published observations.

In what must be the complex distributions of metal values both examples of tin mineralization chosen by the authors a perfect fit of data to any mathematical model can hardly be expected. A 'goodness of fit' at the 95 per cent level of confidence as indicated by chi-squared values is customarily taken as an adequate threshold figure in most circumstances. Chi-squared values in Table 3 (p. A49) show the magnitude of variations from the model for the four sedimentary units recognized in the sequence. Some appear to fall well below the 95 per cent level, but others are much higher. The validity of the authors' interpretation is well founded by the independent field work referred to above, but an explanation of this variation would be of interest.

Similar wide variations in the chi-squared values occur in Table 4 (p. A51) in connexion with the example of the Cornish tin lode. The worst fit with the model is obtained with the assays from section C. This is surely a surprising result in that this section contains more assays than a combination of any two of the others.

The object in trying to fit an array of assays to a mathematical model is to discover any underlying law governing the distribution of valuable material within a deposit. In tabular deposits this is best achieved by considering the accumulate (thickness multiplied by assay) as the variable — rather than the simple assay. Use of this approach is logically and pragmatically preferable in that it produces results with physical meaning and reduces the dimensions of the distribution from three to two. It has been used exclusively by the various authors in discussions on the gold deposits of South Africa (Sichel, Krige and others) whose works in this field are well known. Collectively, these works contain a wealth of information on the lognormal distribution of accumulates.

There is no comparative collection of work on the distribution of simple assays — in fact, all consideration of using these figures was rejected in favour of accumulates at an early stage. Quite apart from their use in statistical studies, accumulates find wide application in mining geology — for example, Connolly's well known contouring method. Use of accumulates may be dispensed with caution in statistical studies of large, disseminated deposits throughout which the values are distributed randomly and where the sample length and attitude is constant (as in the example at Mons Cupri above), provided that the specific gravity does not change.

We know from the published plan of some workings on Simm's lode at Geevor mine^[6] that the width of mineralization varies greatly, so the situation does not comply with the above restraints. Garnett^[6] used accumulates almost exclusively for his illustrations, which include one from the Simm's lode. It is surprising to see their use abandoned in the example chosen here.

Garnett presented histograms showing, separately, the frequency of the distribution of assays in cumulate form from sampling of stopes and development drives. Sample sites of the former widely and erratically spaced throughout the plane of the lode. Those of the latter are

closely spaced along a series of parallel straight lines along the plane of the lode. The author explained a substantial difference between the resulting two histograms as due to the development sampling including more samples in waste than did the slopes. To an extent this is doubtless so, but a contributing factor of importance is the difference in design of the two sampling programmes. This alone leads to two distinct populations being sampled, which, when they are combined as the authors appear to have done in their example, would surely produce the sort of result presented in the paper. It would be interesting to learn whether the authors considered these particular problems, and, if so, how the results given in the paper compare with the results from using, say, only stope samples in accumulate form. Comparison of this distribution with one with only development samples could be very illuminating.

Visual comparison of the histograms in Fig. 10 (p. A51) is hindered by the inconsistencies of scale of 'percentage frequency'. Sections A and D are obviously drawn to a smaller scale, but is it the same? Could the authors reproduce the data summarized by these four histograms in lognormal-cumulative frequency form so that readers may judge for themselves the similarities distinguished by the authors' method?

References

1. Krige D. G. A statistical approach to some basic mine valuation problems on the Witwatersrand. J. chen. metall. Min. Soc. S. Afr.. 52, Dec. 1951, 119-39.
2. Broadhurst J. K. and Batzer D. J. Valuation of alluvial tin deposits In Malaya with special reference to exploitation by dredging. In Opencast mining, quarrying and alluvial mining (London: IMM, 1965). 97-113.
3. Hester B. W. Contribution to discussion (of Williamson D. R. and Thomas T, L. Trend analysis of alluvial deposits by use of rolling mean techniques). Trans. Instn Mm. Metall. (Sect. A: Min. industry). 82, 1973, A28-9.
4. Jones M. P. and Beaven C. H. J. Sampling of non-Gaussian mineralogical distributions. Trans. Instn Min. Metall. (Sect. B: Appl. earth sci.), 80, 1971. B316-23.
5. Gobbett D. J. Joint pattern and faulting in Kinta. West Malaysia. Bull. Geol. Soc. Malays/a no. 4. 1971. 39-48.
6. Garnett R. H. T. Distribution of cassiterite in vein tin deposits. Trans. Instn Min. Metall. (Sect. B: Appl. earth sci.). 75, 1966, B245-73.

FIGURES FROM DISCUSSION

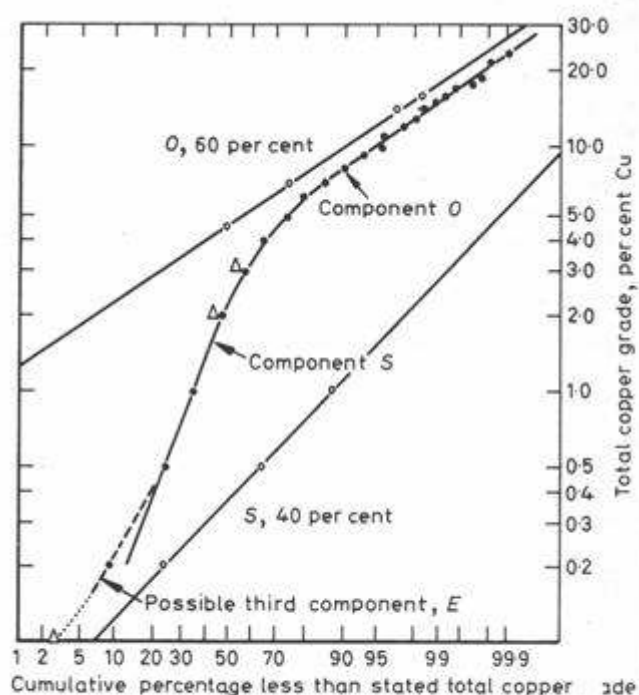
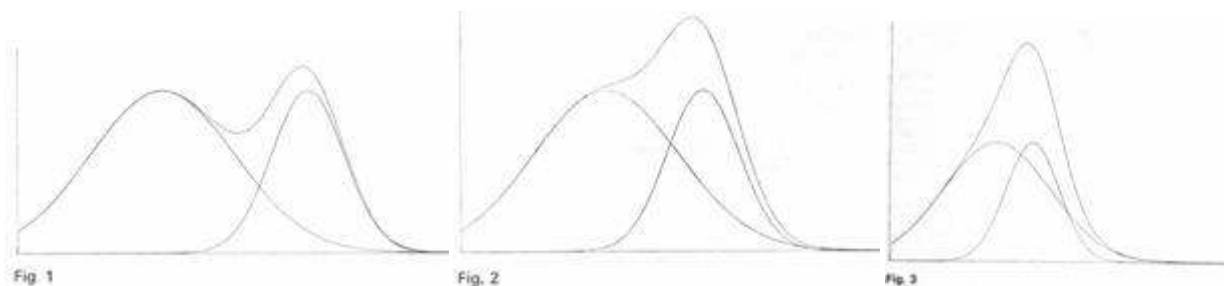


Fig. 5 Partitioned probability plot of 1133 Cu assays from Central African copper deposit. Small open circles are construction points used to define partitioned populations O and S. Open triangles show departures of an ideal mixture of O and S from real data curve (Fig. 4 of Clark and Garnett, with additions)

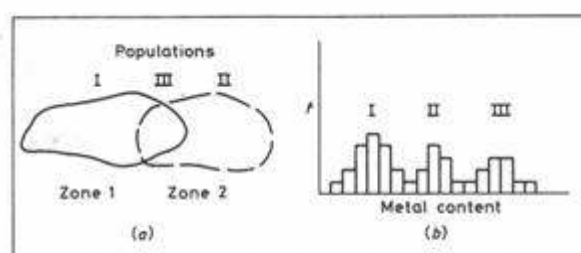


Fig. 6 Generalized representation of distributions of populations resulting from spatial overlap of two stages of mineralization, each characterized by deposition of different metal abundances

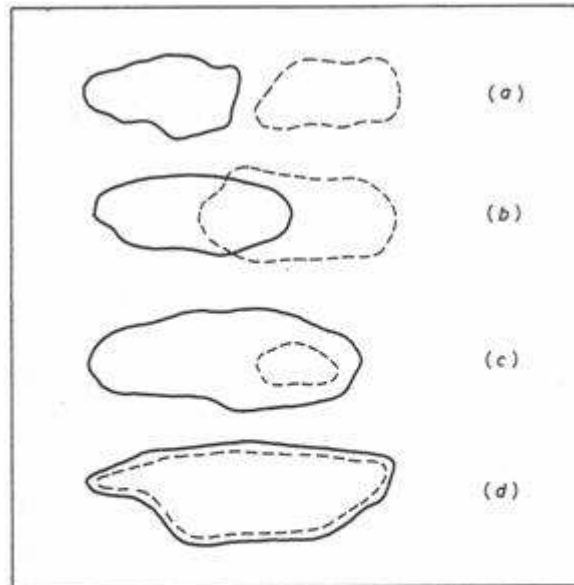


Fig 7: Diagrammatic representation of spatial interrelationship of two stages of mineralisation (a) Spatial distinction results in recognition of two populations I and II. (b) Partial overlap allows recognition of populations I, II and III. (c) Complete enclosure of one within another permits recognition of populations I and III or II and III. (d) Coincidence (superimposition) of two zones allows recognition only of population III. (Refer to Fig. 6 for definition of populations)