SMYTH, C.P. Lithogeochemical prospecting: signals processing applied to segmentation of geochemical borehole profiles. APCOM 87. Proceedings of the Twentieth International Symposium on the Application of Computers and Mathematics in the Mineral Industries. Volume 3: *Geostatistics*. Johannesburg, SAIMM, 1987. pp. 327 – 343.

Lithogeochemical Prospecting: Signals Processing Applied to Segmentation of Geochemical Borehole Profiles

C.P. SMYTH

Anglo American Corporation of South Africa, Johannesburg

Large volumes of quantitative geochemical data, presenting a formidable interpretation task to geologists, are generated in the course of exploration drilling for new ore deposits. The incorporation of geological expertise into computer programs developed for the interpretation of this data would improve the speed and accuracy with which it is interpreted.

A detailed analysis of the geochemical log interpretation problem, together with a survey of relevant research in Artificial Intelligence and petroleum well-logging techniques, confirms that artificial intelligence, particularly as developed in the fields of expert systems and pattern recognition, provides techniques for implementing such data interpretation programs. In addition, it provides data analysis techniques significantly more powerful than those currently applied to geochemical data, and very appropriate to the 'hypothesise and test' character of mineral exploration problem-solving.

The conversion of raw borehole log data to meaningful higher level geochemical entities, in terms of which geological reasoning may take place, is a critical step in applying these techniques to log data analysis. A method of achieving this conversion as an exercise in signal-to-symbol transformation is presented. [1]

Introduction

The interpretation of geochemical borehole profiles is an exercise in signal understanding, and hardly any signal understanding task can be performed using the raw numerical signal values directly: some description of the signal must first be obtained (Nii[2]). Such descriptions should be as compact as possible, and their elements should correspond as closely as possible to meaningful objects or events in the signal-forming entity.

A discussion of the abstract aspects to computing these descriptions and their importance to interpretation of geochemical profiles is presented below. Andrews P. Witkin [3] has described the practical problem of computing these descriptions as follows:

'A great deal of effort has been expended to obtain this kind of primitive qualitative description, and the problem has proved extremely difficult. The problem of <u>scale</u> has emerged consistently as a fundamental source of difficulty, because the events we perceive and find meaningful vary enormously in size and extent. The problem is not so much to eliminate fine-scale noise, as to separate events at different scales arising from distinct physical processes. It is possible to introduce a <u>parameter of scale</u> by smoothing the signal with a mask of variable size, but with the introduction of scale-dependence comes ambiguity: every setting of the scale parameter yields a different description; new external points may appear, and existing ones may move or disappear. How can we decide which, if any, of this continuum of descriptions is 'right'?

There is rarely a sound basis for setting the scale parameter. In fact, it has become apparant that for many tasks, no one scale of description is categorically correct: the physical processes that generate signals (such as images) act at a variety of scales, none intrinsically more interesting or important than another. Thus the ambiguity introduced by scale is inherent and inescapable, so the goal of scale-dependent description cannot be to eliminate this ambiguity, but rather to manage it effectively, and reduce it where possible.

This line of thinking has led to considerable interest in multi-scale descriptions. However, merely computing descriptions at multiple scales does not solve the problem; if anything, it exacerbates it by increasing the volume of data. Some means must be found to organise or simplify the description, by relating one scale to another.'

Every point raised by Witkin is wholely relevant to geochemical 'signals': his problem of scale is isomorphic to the classic geochemical background/anomaly distinction problem.

This.paper presents one solution to these problems, specifically developed for signals resulting from the chemical analysis of borehole samples.

Formal and abstract aspects of borehole profiles

In the domain of geological inference and reasoning (by geologist or computer) one can distinquish between physical objects (minerals, boreholes), information objects (borehole logs, data files) and executive objects (the computer, the geologist). Executive objects are active, while physical and information objects are passive. They are active because they can manipulate either, or both, of physical objects (sample them, name them, destroy them) and information objects (create them, define attributes for them).

It is useful, in the discussion of data, to be explicit about the status of any particular attribute value, with respect to the object it qualifies, and the relevant operative executive object, because of the bearing this status has on whether or not the value must be stored.

The status of an attribute value is primary with respect to its qualified object and a specific executive object, if, once the object has been identified, the value cannot be computed by the executive object from other primary attribute values of the identified object (ie: it must be stored somewhere).

The status of an attribute value is secondary with respect to its qualified object and a specific executive object, if, once the object's primary attributes have been located, the executive object is capable of computing that value from the primary attributes, or fetching it from the attribute values of constituent subobjects or super-objects of the qualified object. Whether attribute values are primary or secondary, then, depends both on how objects are identified (eg: fourth sample from top, or, sample ABC-24), and what knowledge is available to the executive object regarding generation of attribute values.

A borehole segment is an information object representing any sampled section of a borehole.

For the geologist, provided the name of the borehole from which the segment originates is known, and provided the borehole data base is available, the most obvious primary attributes of a segment are its starting depth, ending depth and name, if it has one. Any other attributes are secondary, as they can be computed or fetched from the data base.

Establishing the status of segment attributes with respect to the computer, and deciding whether secondary attribute values should be stored explicitely are both subjects of major importance to signal-to-symbol transformation, because borehole segments are the symbol structures to essential pattern recognition in borehole profiles.

A borehole interval is an information object representing any continuous sequence of segments or intervals on a borehole. (The reason for this recursive definition of an interval will become clear below.) As such, the relationships between intervals (successions of segments) and their segments, and segments (successions of samples) and their samples, are very similar.

The similarity goes further than their both being sequences: with significant exceptions, for any set of primary attributes measured (or calculated) for any particular interval of borehole segments, those attributes are measured for all the segments of the interval (in the most general case, the interval of segments is the entire borehole). This means that, as for storage or representation of primary sample attribute values, an appropriate construct for primary attribute values of segments is also the Third Normal Form of the Relational Data Base model.

In fact, on close inspection, it becomes clear that a sample may be regarded as a degenerate (single-sample) segment, about which certain information, such as slope and correlation co-efficient, as a result of its degeneracy, is missing.

Exploitation of this similarity contributes to standardisation of data structures and data structure manipulation primitives.

While these relationships between intervals and their components and segments and their components are similar structurally (syntactically), their semantics are quite different. An interval has meaning only in terms of its individual component segments, and the relations, such as discontinuities, between them.

In contrast, a segment has meaning with respect to the attribute values used to derive it, only on the explicit assumption that no discontinuity, at the segment's defined scale, exists between its component samples.

Consequently, intervals and segments have entirely different representations. Intervals are represented by lists of their component segments, while segments are represented by 14-tuples holding segment attribute values.

In general, there are two methods of referencing an attribute value of an information object - by its name or by its value. (In this context, we understand 'name' to be name or address - which is acceptable in the context of our distinction between primary and secondary status of objects or attributes.) If it is to be referenced by name, with the ultimate objective of using its value, that name must hold sufficient information to allow the active executive object to locate the value, if it is stored, or perhaps to recompute it, if it is a secondary attribute value.

Ensuring that newly generated information objects and their attributes are given names that hold this information is the responsibility of the executive objects which create them. At the same time, however, and particularly when large numbers of information objects are involved, this responsibility can be burdensome and error-prone for the geologist. It should, therefore, when possible, be the responsibility of the system to generate these names, capitalising, in so doing on its strengths in inheritance administration and search control.

The system defined below assumes responsibility for the naming of segments and intervals only. Segments are represented by the same constructs in memory and storage, and retain the same names in both contexts. Intervals, however, have different representations in memory and in storage, and require more sophisticated naming conventions.

Quantitative geochemical attribute values as signals: Consequences for segmentation

Attributes of geochemical borehole profile signals

Providing the entire sample is analysed, and ignoring analytical error, a 'whole rock' chemical analysis value of a sample which is a continuous section of drill-core, or drill-chips, represents the mean composition of that sample along the section.

Mathematically we may think of such a value as being

$$\left[\int_a^b f(x)\,dx\right] \;/\; (b-a)$$

where:

- x = depth
- a = depth at which sample starts
- b = depth at which sample ends
- f(x) = chemical attribute value at depth x

We shall call such a signal value an 'integrated signal value'. Signal values of this type are in a different class from signals measured discontinuously, at a succession of regularly (or irregularly) spaced points, which we shall call 'spot signal values'. Most digitally acquired borehole magnetometer logs are examples of the latter category. The difference may, or may not, be significant, and its significance is a function of the nature and scale of the phenomena sought in the signals.

Importantly, the two categories differ in the relationship existing for each between adjacent signal values. For integrated signal values, the adjacency relationship may be conjunction or disjunction, while for spot signal values, the relationship is always one of disjunction.

Sequences of conjunctionally related integrated signal values carry more information about signal discontinuities than do spot signal values - information which may be used both to assist segmentation, and during pattern definition, particularly if discontinuities are important pattern components, as they are in geochemical borehole logs.

How this information is used during segmentation is detailed below, and revolves primarily around administering segment-edge signal values.

The role of signal discontinuities as pattern components makes it important to be able to refer to them explicitly during pattern definition (ie: during interval parsing). It proves expedient, however, to express them only implicitly (as relationships between signal values or segments) in segmented representations of raw signal data, as is made clear below.

The above characterisation of borehole signal values helps develop methods of transforming and describing the signals in a manner most sensitive to meaningful entities they may reflect, by highlighting signal attributes which may assist segmentation.

Use of signal attributes for segmentation

One interesting attribute, in terms of

the mathematical representation of the signal value given above, is the signal value numerator (viz: $\int f(x)\,dx$), which may be calculated by multiplying the signal value by its sample length.

Since geologists, and particularly mining geologists evaluating ore reserves, think naturally, if unconsciously, of the 'area under the curve' when interpreting borehole profiles, this seemed a potentially useful metric to involve in the segmentation procedure. It seemed even more attractive in view of the problems surrounding the application of conventional segmentation techniques (eg: piecewise linear approximation) to signal values integrated over different interval lengths (ie: different sample lengths) - and in view of the recognition that segments themselves could be regarded as conjunctional integrated signal values.

There is another significant characteristic of geochemical attribute values as signals, namely that, in the majority of cases, when there is a change in the direction of change of the value, it is in response to a real change in the sampled world, rather than being 'introduced noise'. Since establishing whether or not such real changes are significant can reasonably be left to higher levels of the signal-understanding task, this too may be incorporated into the segmentation activity.

An important qualifier to this application of 'first derivative slope change' as a segmentation criterion is that, in the context of borehole signals, its effect must be independent of the direction of evaluation along the profile. This is necessary to ensure the same representation of features penetrated in different directions by a borehole. Its primary effect is that, in certain situations one sample must be included in two segments. It has more subtle effects on the administration of missing signals.

Segmentation overview

The Segmentation Method presented below provides a variable-scale signal-to-symbol transformation technique which is very sensitive to the information content of its input signals, together with symbol representation constructs which efficiently manage multi-scale signal descriptions.

Signal-to-symbol (ie: signal-to-description) transformation is implemented by three scale-independent preliminary phase operations, designed to provide the most primitive level of signal description. These are followed by 'fourth phase transformation', which is scale dependent, and may involve a series of transformations, producing successively more generalised descriptions of the original signals.

Signal descriptions are represented by nested lists of segments, with finer scale descriptions nested within more general representations. Different nesting structures are used during preliminary phase and fourth phase transformations, largely as a result of the predominance of splitting in the former, and merging in the latter. It is likely that with streamlining of the preliminary transformation phases, use of the 'splitting' structure may be discontinued.

Segments are the primitives used for signal description, and constitute the elements of the nested lists which describe signals. They are represented by 14-tuples, which hold the following segment attribute values, which are primary attribute values in this context:

(1) The segment name (identifier).

(2) The name of the first sample in the

segment.

(3) The name of the last sample in the segment.

(4) The depth of the top of the first sample.

(5) The depth of the bottom of the last sample.

(6) The estimated value of the quantitative attribute (X) used to define the segment at the top of the segment.

(7) The estimated value of X at the bottom of the segment.

(8) The maximum value of X in the segment.

(9) The minimum value of X in the segment.

(10) The (weighted) mean value of X in the segment.

(11) The slope of the linear regression of X on depth.

(12) The intercept of the linear regression of X on depth.

(13) The X: depth correlation coefficient.

(14) The number of samples in the segment.

For the geologist, only the first three items above are truly primary segment attributes.

Segments are named uniquely within each signal description. A signal description (in fact an interval) has only one name pertaining to the multiple-scale nested list which describes it. Different levels of that description are accessed procedurally, not by name.

Segmentation: Phase one

During Phase One of its operation, the Segmentation Process applies the knowledge that, in the majority of cases, along a sequence of borehole samples, when there is a change in the direction of change of an attribute value, that change reflects a change in the sampled world, and may be useful as a segment boundary. For simplicity of application, therefore, it sets segment boundaries at all changes in direction of signal value change.

Specifically, Phase One Segmentation operates under the following conditions:

(i) Knowing that the objects upon the attributes of which segmentation is based are samples, and that the attribute values are conjunctional integrated values.

(ii) Requiring that the segmentation be direction independent.

(iii) Observing that three significantly different patterns may surround changes in the direction of change of an attribute value.

(iv) Knowing that segmentation responses to these different patterns should be sensitive to the information signal values will carry (being conjunctional integrated signal values), regarding the location of discontinuities in the sampled object. Specifically, this information carried by a signal value marginal to a segment edge may:

a: include information related to signal values on either side of the marginal signal value (ie: the discontinuity occurs within the sample). This cannot arise for spot signal values.

b: include information related only to one of its adjacent signal values (ie: a single discontinuity occurs at one margin of the sample).

c: include information related to neither of its adjacent signal values (ie: there are discontinuities on both margins of the sample).

In practice, application of this knowledge is distributed between Phase One and Phase Two segmentation, because it can require consideration of relative signal value magnitudes (considered in Phase Two), rather than simply their directions of change.

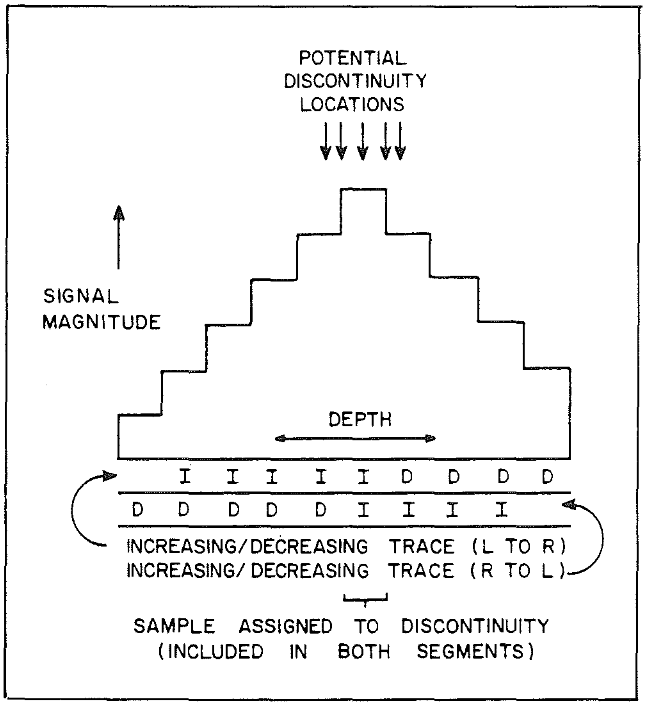


FIGURE 1(a). Type (1) direction-change pattern. Finite automaton searches for XXYY pattern in L to R direction, where X = I and Y = D, or X = D and Y = I

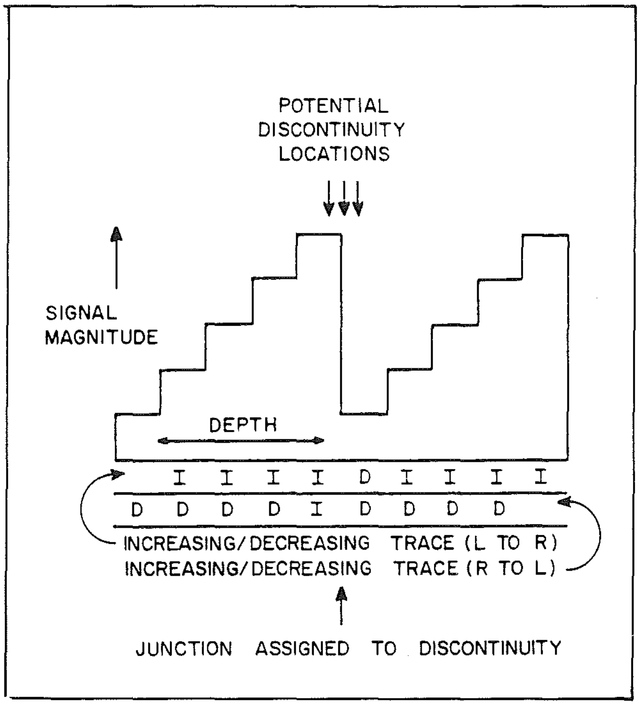


FIGURE 1(b). Type (2) direction-change pattern. Finite automaton searches for XXYXX pattern in L to R direction, where X = I and Y = D, or X = D and Y = I.

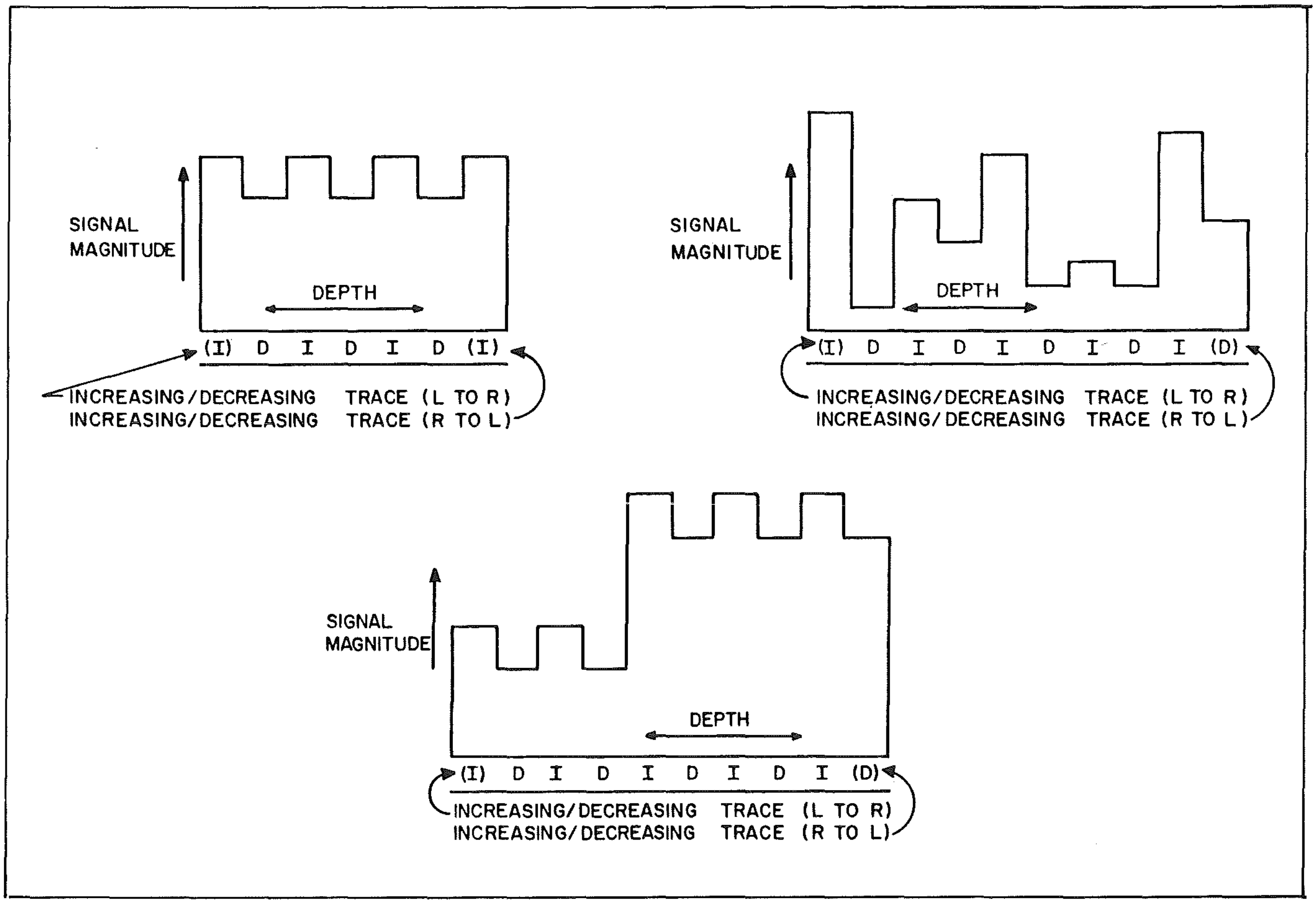


FIGURE 1(c). Type (3) direction-change pattern

The three different patterns surrounding direction changes, are presented and annotated in Figures 1(a), 1(b) and 1(c), and described below.

Pattern (1): In patterns of Type (1) in Figure 1, on the basis of direction of signal value change alóne, there is no evidence for the exact location of a discontinuity, if, indeed, one exists, at all. Further, depending on the direction in which the value sequence is evaluated, different values become the first values of a new sequence. Consequently, because conditions (ii) and (iv)a above apply, the marginal signal value is included in segments to either side of it - as the first in one, and the last in the other.

Pattern (2): In patterns of Type (2) in Figure 1, on the basis of direction of signal value change alone, there is evidence for a discontinuity, and for its unequivocal location - although this becomes evident only after the direction of change of the second sample after the discontinuity has been checked. The pattern recognition is not direction-sensitive, and condition (iv)a or (iv)b above applies, allowing segmentmarginal samples to be included in only one segment each. (The discontinuity itself may occur in either of the two samples, or exactly between them.)

Pattern (3): Patterns of Type (3) may be regarded as special cases of Type (1) or Type (2), and may be managed as either, depending on whether condition (iv)a, (iv)b or (iv)c is thought to apply. If condition (iv)c applies, management as for Type (2) suffices. A fourth alternative is to regard the entire extent of such sequences as one segment. This policy was adopted initially, because such patterns arise commonly in geochemistry, when the only difference between successive values is signal noise, or because of sampling effects resulting from under-sized samples. It has the effect of minimising the generation of many single sample, or two sample segments. However, resolution of signal structure can be severely compromised, and this policy was dropped in favour of the Type (2) approach. This causes Type (3) patterns to produce a succession of single sample segments. Clearly, during data interpretation, condition (iv)a, (iv)b and (iv)c may need to be considered, and management of this task at higher levels in the system is facilitated by single sample segments, rather than by the storing of two-sample segments, which would result from the Type (1) approach.

The above theoretically developed approach to segmentation is implemented in three steps:

(i) A dichotomized trace (in one direction) of the relevant borehole profile is produced, reflecting each signal value's decreasing or increasing relationship to the sample before it.

(ii) This trace serves as input to a finite automaton subroutine, which uses it to recognise and output Phase One segments (as defined by their first and last samples).

(iii) Segment attributes are calculated, and stored in segment 14-tuples, and accumulated into a list of segments which constitutes the interval that represents the entire length of the borehole.

In practice, all these functions are carried out by a Pascal program (PROLSEG4), which is called from a Core program. The only operations for which the Core program itself is responsible, are the following:

(i) Receipt of the user's command to initiate the activity, and his selection of attribute (element) and data set (borehole).

(ii) Initiation of the PROLSEG4 program.

(iii) Receipt from the PROLSEG4 program of all of the output segments, their structuring into a nested list, and naming of the resulting interval.

Segments resulting from execution of Phase One Segmentation of the Barium profile from Borehole 46gr are shown in Figure 2, alongside a plot of the original data.

Figure 3 explains the structure of the nested lists used to represent intervals during this and the following stage of

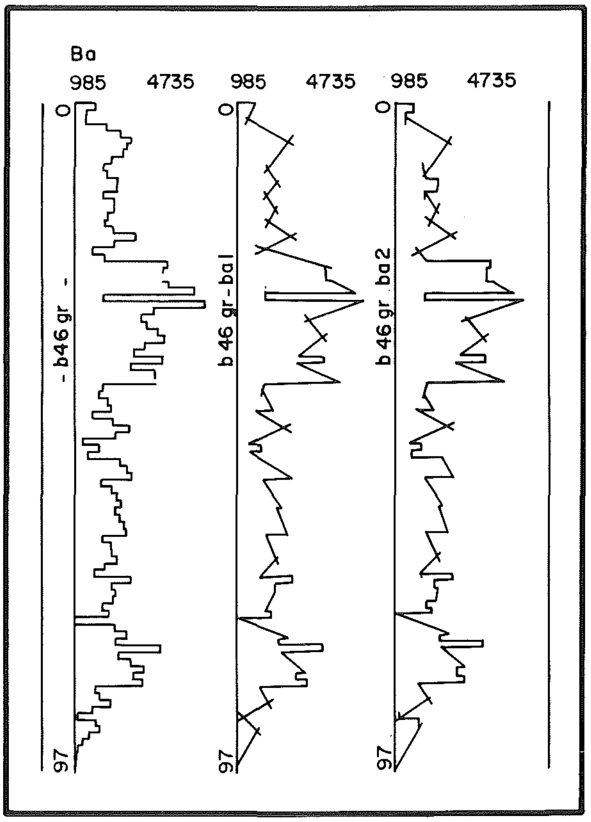


FIGURE 2. Barium profiles for borehole b46gr: L.H.S.: Raw Ba data, with two missing samples; middle: Profile output by phase one segmentation; R.H.S.: Profile output by phase two segmentation

segmentation. This representation was originally designed to manage both the splitting (Phases One and Two) and merging (Phases Three and Four) phases of segmentation. It has since proven sub-optimal for the representation of merged segments, and might eventually be discontinued altogether, should the first three phases of the Segmentation system be combined into one process. It should be noted that discontinuities have no explicit expression in these structures, but have to be derived from the marginal values of abutting segments.

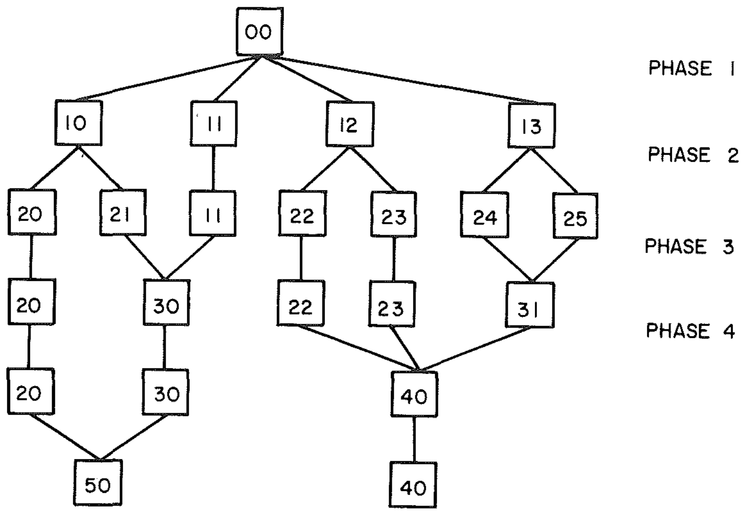
Segmentation: Phase two

Phase Two segmentation detects signal features which are potentially significant with respect to characteristics of the sampled rock, but which are not detected by Phase One Segmentation because they are not accompanied by a change in direction of change. It looks for these features by examining, one segment at a time, the signal values from samples within a segment - looking for an 'abnormally' large signal value change between adjacent signal values. If it finds such a change, it splits the segment into two segments, across the abnormal change. Examples of signal value sequences showing such features are presented in Figure 4. Currently, this activity looks only for one feature around which to split within each segment. A more sophisticated 'splitter' would be sensitive to a number of such features, particularly if segments are constrained to linear representation, since uni-directional curvilinear changes (which do occur in borehole profiles) could then be approximated by a number of relatively short linear segments.

Phase Two Segmentation is implemented in the Core program in PROLOG, with calls made to Data Base Read and Statistics Facilities, which are implemented in Pascal, when necessary.

Though implemented in PROLOG, the program is procedural in character. It processes the signal values grouped with each Phase One segment by evaluating a splitting metric for each position (between signal values) at which a segment may be split, and then checking the highest valued metric for the segment against a pre-set threshold. If the metric is higher than the threshold, the segment is split, otherwise attention moves on to the next segment. Evaluation of the metric is explained in Figure 5. Output from conducting Phase Two Segmentation on Phase One segments from the Barium profile of Borehole 46gr is shown in Figure 2.

DIAGRAMATIC GRAPH REPRESENTATION :



Nested List Representation of the Top of the above Tree:

```
[&lt;segment 00 attributes&gt; (empty) ([<seg 10 ats> (00) (LIST A)] ,
                                  [<seg 11 ats> (00) (empty) ] ,
                                  [ etc                      ]) ]

where LIST A = [ ( < seg 20 ats > (10) ( < seg 50 ats>(20,30) (empty))),
                ( < seg 21 ats > (10) ( < seg 30 ats>(21,11) (empty)))]
```

General Structure:

```
interval := [< tuple of seg atts > (list of names of parent interval
              or intervals) (list of subsegs or a single enclosing
                                                          segment) ]
```

ie:

```
     interval := list of A's
     where A  := [ tuple X, list Y, List Z ] ;
     tuple X  := a tuple of segment attributes;
      list Y  := either: name of parent seg (if seg arose from split)
                     or: names of parent segs (if seg arose from merge)
      list Z  := either: list of subsegments into which current
                         level segment is split
                     or: super-segment into which current level
                         seg has become the 'left-marginal' merged
                         seg;
                         (a super-seg is represented only within
                         its left-marginal sub-segment)
```

FIGURE 3. Nested list structure designed to represent both split and merge components of an interval. (It has major problems in representing overlap.)

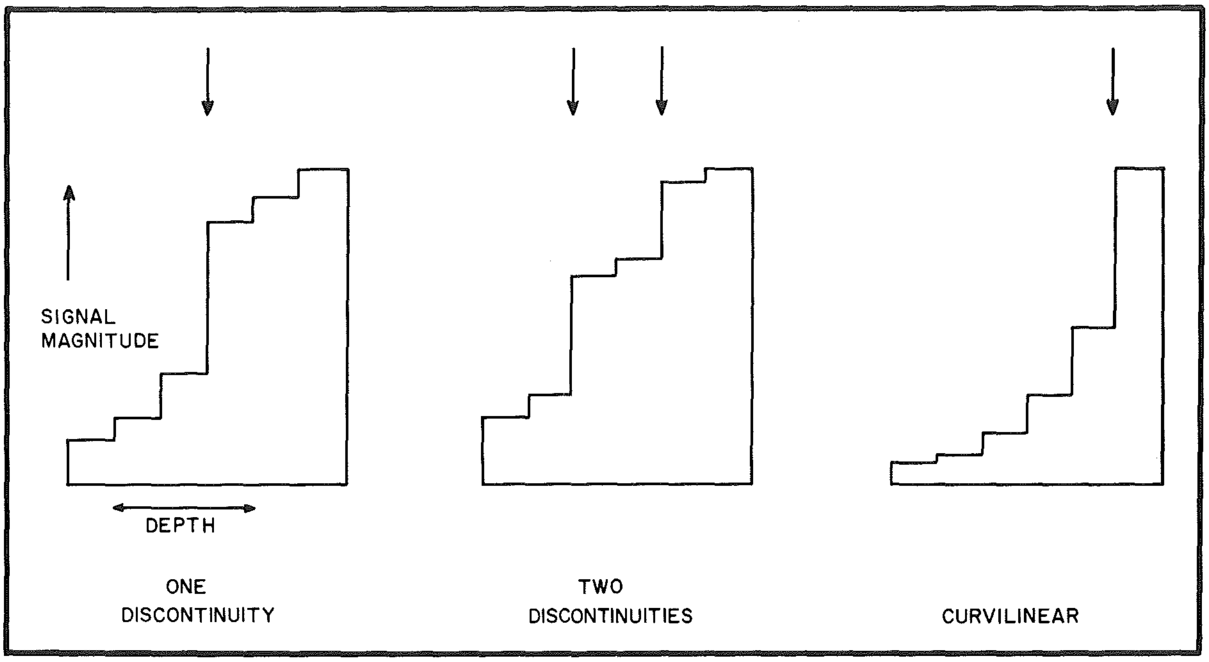


FIGURE 4. Signal value sequences showing significant discontinuous change

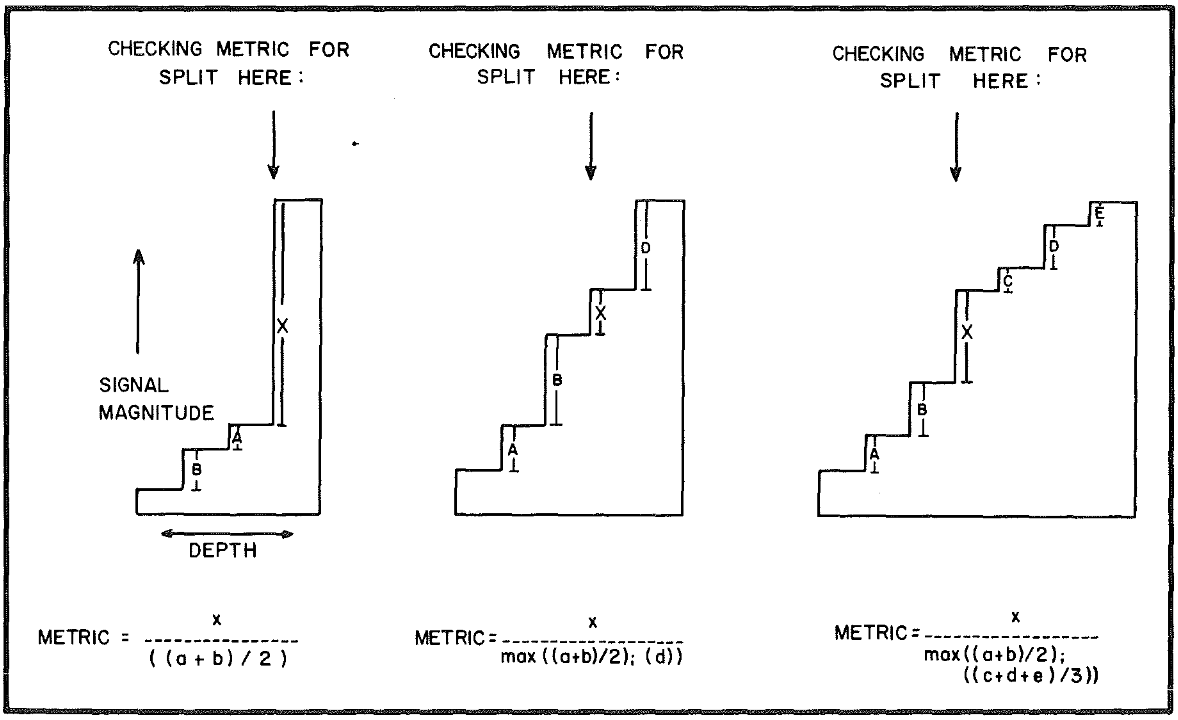


FIGURE 5. Evaluation of the 'segment splitting metric'

Segmentation: Phase three

It is the principle of the first two phases of segmentation to give expression to as much structure in the original signal as possible, although this may incur the cost of generating and storing more segments than are necessary. The task of 'overlooking' these unnessary segments is left to higher levels of signal transformation. Nevertheless, the overall objective of segmentation is to express the signal in as concise a form as possible.

The sequential combination of Phase One and Phase Two Segmentation can generate erroneously redundant single-sample segments, which are completely overlapped by non-degenerate segments accounting for the same subsection of a profile. How such a situation can develop is illustrated in Figure 6.

Further, the heuristics employed in the first two phases of segmentation may generate segments whose presence may be entirely accounted for by noise in the original signal. (Some of the important knowledge a geochemist uses in interpreting a profile is his awareness of the (highly variable) noise level inherent in a geochemical signal value. The noise level is a function of the attribute measured, the method used to measure it, its measured magnitude, and sometimes of the values of other sample attributes.)

Phase Three Segmentation has, as its primary objectives, the elimination of redundant degenerate segments, and degenerate segments whose isolated existance is most likely to result from signal noise. At the same time, it transforms the internal representation of an interval from that designed for splitting operations, to that designed for the merging phases of segmentation. The structure of this representation is explained in Figure 7. As in Figure 3, it should be noted that discontinuities between segments are not explicitly represented.

Phase Three Segmentation is implemented in the Core program in PROLOG, with calls to a Statistics Facility for computation of new segment attribute values.

Segmentation: Phase four

The first three phases of the Segmentation Facility effect signal-to-symbol transformation at the finest practicable scale, given the nature of the signal, and the nature of the features that are to be recognised in the signals. However, as explained in the introduction, multi-scale

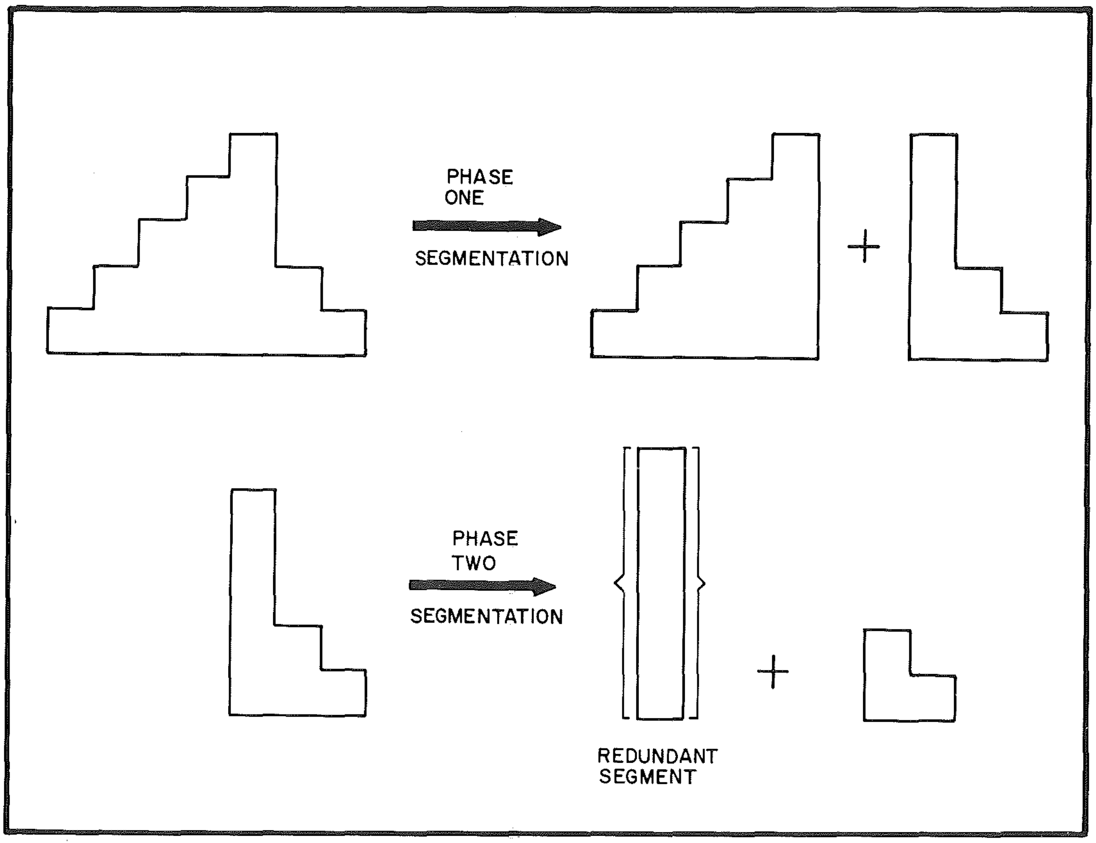
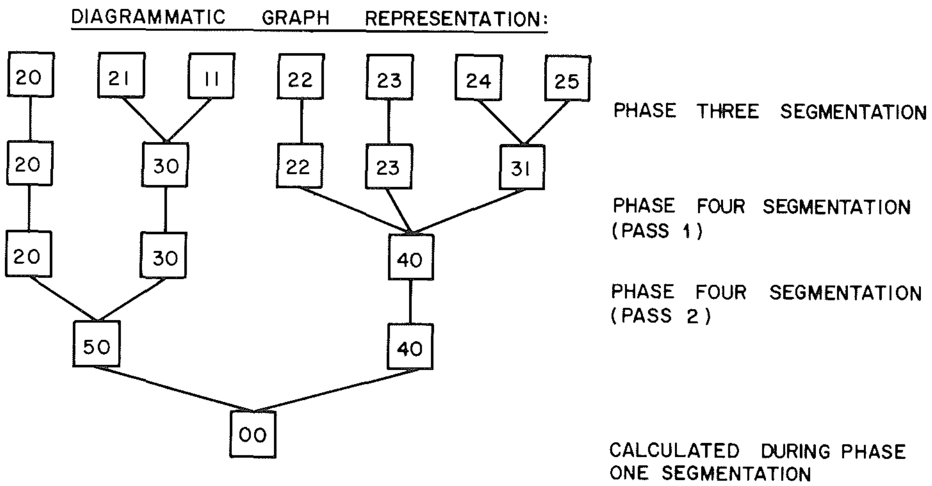


FIGURE 6. The development of redundant degenerate segments, by a combination of phase one and phase two segmentation



General Interval Representation:

```
[< tuple of 'super-segment' attributes > (list of its sub-intervals)]
```

Nested List Representation of the above interval:

before Phase Three Segmentation:

```
[< seg 00 ats   (( < seg 20 ats > (empty)),
                 (  < seg 21 ats > (empty)),
                 (  < seg 11 ats > (empty)),
                 (  < seg 22 ats > (empty)), etc)]
```

after Phase Three Segmentation:

```
[< seg 00 ats   (( < seg 20 ats > (empty)),
                 (  < seg 30 ats > ((seg 21 ats > (empty)),( < seg 11 ats > ( empty))),
                 (  < seg 22 ats > (empty)), etc)]
```

FIGURE 7. Nested list structure designed to represent successively merged components of an interval during segmentation

representations of the original signal are needed for optimal feature recognition.

Phase Four Segmentation transforms relatively finer scale symbolic representations of geochemical profile signals to relatively coarser scale representations. As an activity, it may be executed repeatedly (with different parameters) on the same interval, to give ever more coarse-scale representations of the same signal. Coarser scale representations carry with them all their finer scale detail, which may readily be accessed when necessary.

Transformation of finer scale signal representations to coarser scale representations is essentially the controlled merging of segments.

Phase Four Segmentation uses an 'Adjacent Segment Area Difference' metric (ASAD) to control segment merging. If this metric, as evaluated for both directions across a boundary between two

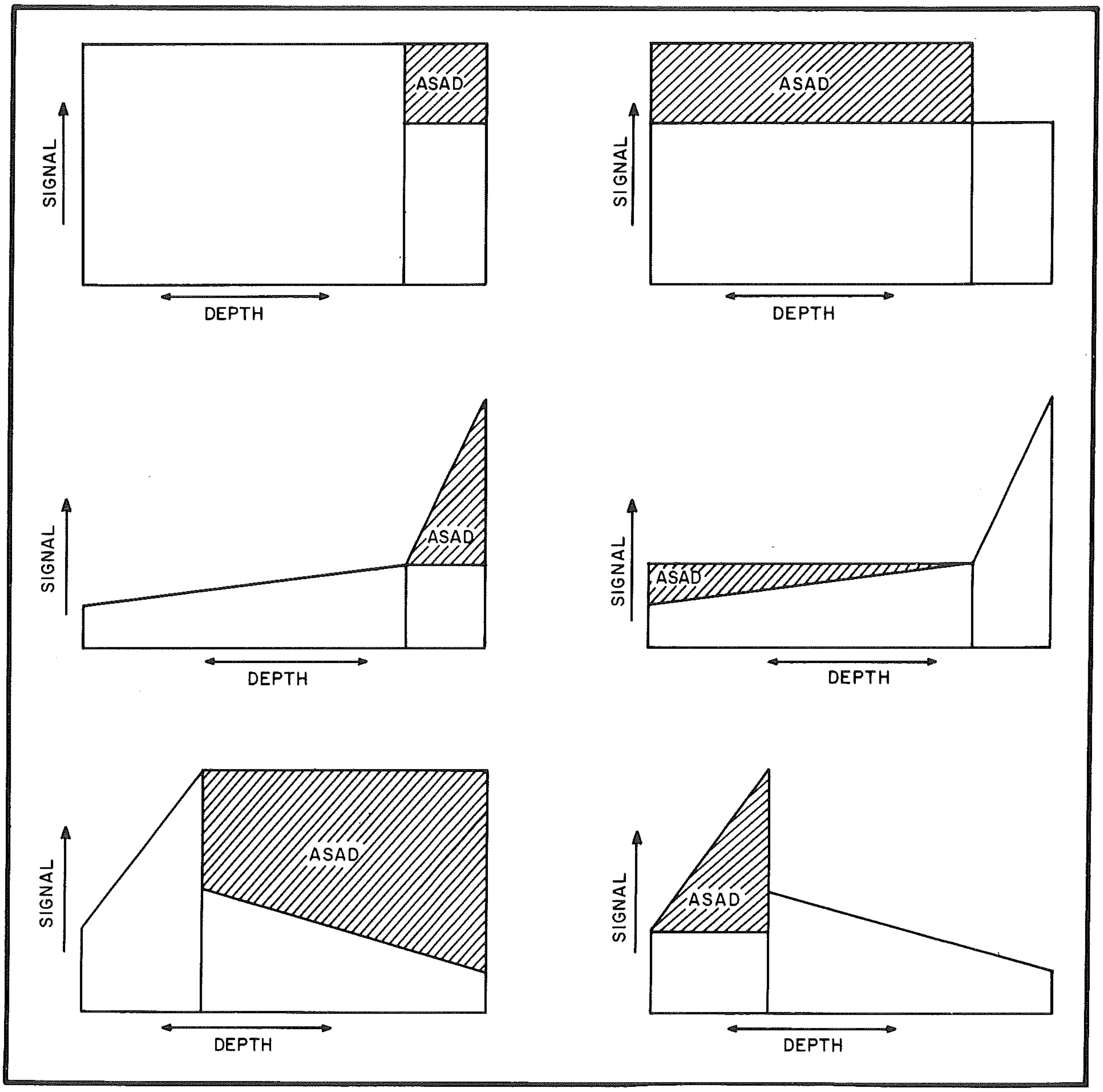


FIGURE 8. Graphical illustration of the evalution of the 'adjacent segment area difference' metric, emphasising its direction-dependency

segments, remains below a user-selectable threshold, then those two segments may be merged, provided that a second, discontinuity-related, constraint is also satisfied. Evaluation of the ASAD metric is illustrated in Figure 8.

The user-selectable discontinuity-preserving constraint on segmentation is provided because of the importance 'short-term' (and therefore with relatively small ASAD's) discontinuities assume in certain geological contexts, which requires their explicit representation at coarser scales than exclusively ASAD-controlled segment-merging would permit.

Examples of the operation of ASAD/DP - controlled segment merging are shown in Figure 9, together with an example of merging controlled by the ASAD metric alone.

Phase Four Segmentation is implemented in the Core program in PROLOG, with calls to a Pascal Facility for the calculation of the ASAD metric, and to the Statistics Facility for the computation of new segment attribute values.

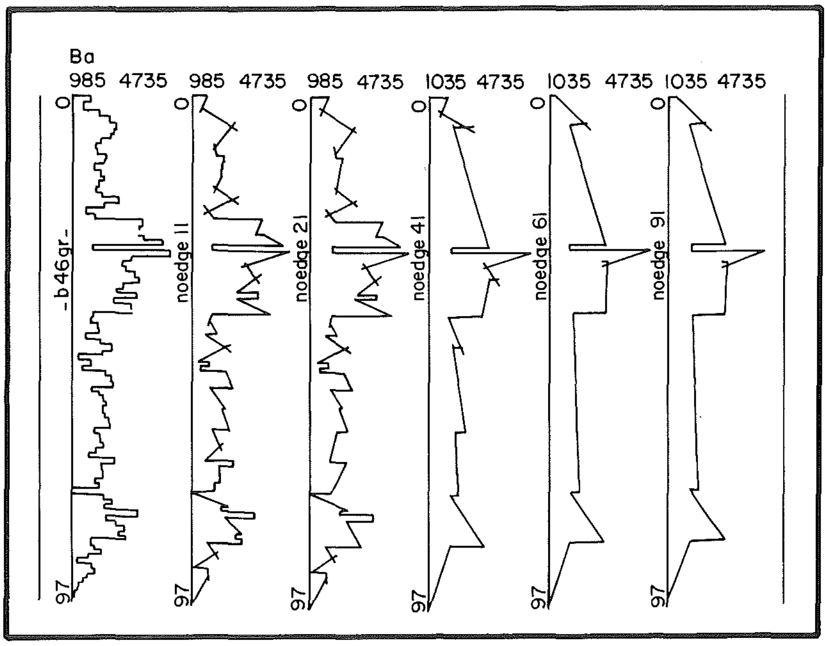


FIGURE 9(a). Phase four segmentation without the discontinuity preserving (DP) metric (ASAD thresholds for 9(b))

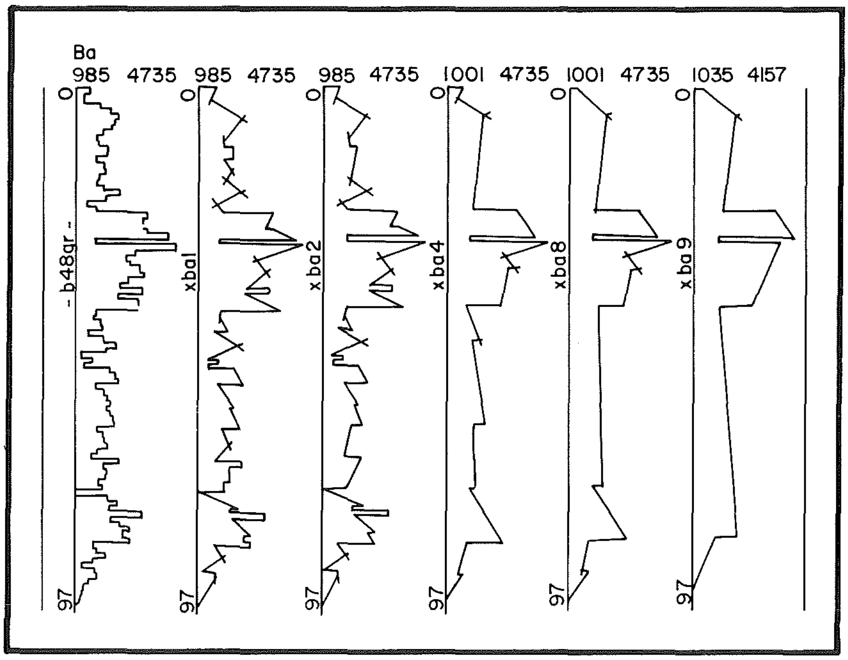


FIGURE 9(b). Phase four segmentation with the discontinuity preserving (DP) metric set to 0,35. Profile xbal is derived from prifile b46gr ba 2 of Figure 2 with an ASAD threshold of 1000, xba2 from xbal with a threshold of 2000; likewise with thresholds of 4000, 6000, and 900 up to profile xbz9

Segmentation: Variable sample lengths

In its current implementation, the system has to provide especially for samples of variable length only during grouping of samples in Phase One and Phase Three Segmentation, and whenever calculating the attributes of segments including variable length samples.

The general principle applied during grouping of samples into a segment is that, unless samples have achieved full 'segment' status (ie: not until the first three phases of segmentation are complete), they may not be grouped together

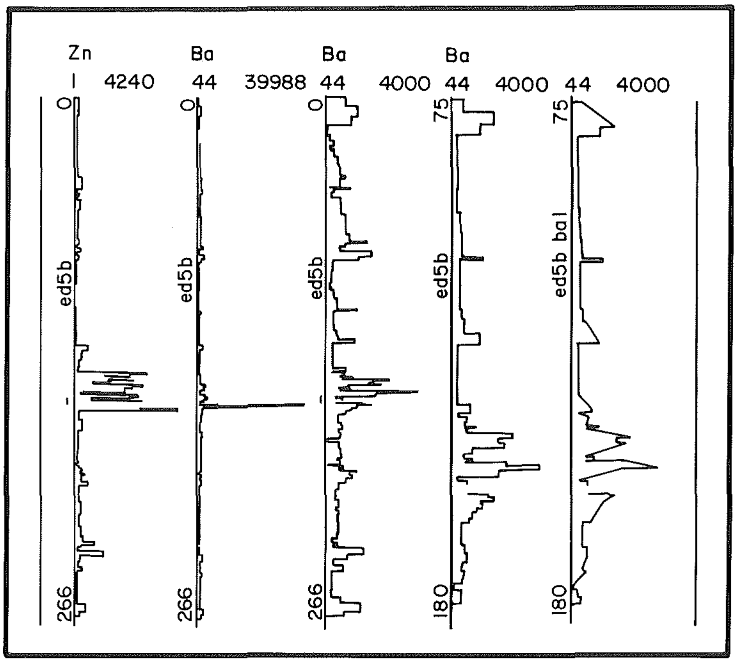


FIGURE 10. Segmentation of data from samples of variable length.
Unsegmented Zn and Ba profiles are presented to illustrate the dynamic range over which segmentation algorithms have to perform, and to illustrate the importance of scale in data interpretation. Profiles (3) and (4) illustrate the utility of varying horizontal and vertical scales respectively, to better appreciate the structure of the data. Profile (5) illustrates output from phase one segmentation on the data of profile (4), with obvious examples of sample length considerations overruling the direction-change segmentation heuristic (e.g.: the isolated, short, high-valued sample one third the way down the profile)

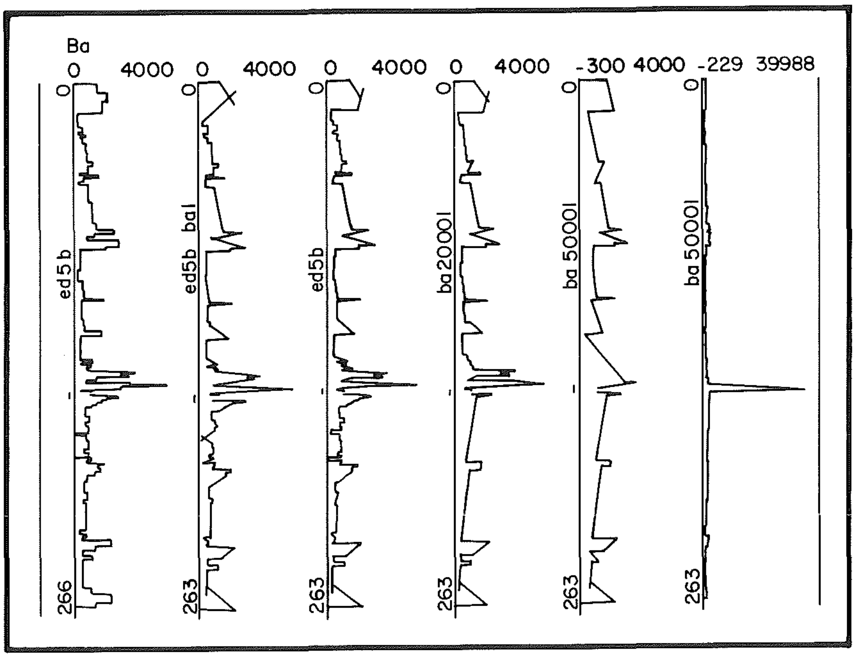


FIGURE 11. Segmentation phase one, two and four as applied to samples of variable length over the full extent of a borehole. Profile (2) is the output from phase one segmentation, profile (3) the output from phase two segmentation (the effect of which is well illustrated in the top three samples of the profile). Profile (4) and (5) result from phase four segmentation at ASAD thresholds of 2000 and 5000 respectively (DP metric was held constant at 0,35). Although the linear regression approximation to the long segment near the centre of profile (5) is obviously inappropriate (and produces a negative result at one extremity), the technique provides a good approximation to most of the profile. The profile plotted in position (6) is the same as that plotted in position (5), but is plotted at a different scale. It should be compared with the second profile of Figure 10

with other samples of substantially different lengths. Factors delimiting the maximum acceptable length-differences between samples of one segment during low level segmentation are user-adjustable, by variation of 'current-maxlength' and 'current-minlength' factors.

Currently a rather unsophisticated, but functional, method has been adopted for calculating the attributes of segments including variable sample lengths. It requires the substitution of each signal value within a segment by the closest integral number of equivalent integrated signal values equal in length to half the length of the shortest signal value in the segment, at appropriately spaced locations along the sample length. Segment attributes are then calculated from all such 'derived' integrated signal values of equal length for the segment.

It is recognised that a more sophisticated technique is required for this purpose. Implementation of such a technique was delayed for consideration together with consideration of more sophisticated segment represention (ie: curvilinear segments).

Examples of segmentation output from signal profiles with variable sample lengths are presented in Figures 10 and 11.

Segmentation: Refinements and conclusions

Among the many refinements that should be added to the Segmentation Facility, as presented, the following are some of the most obvious:

(i) Discrimination of more subtle discontinuities, and more than one of these per segment, during Phase Two Segmentation.

(ii) Incorporation of both segment slopes into ASAD metric calculation (by use of extrapolated segment lines to demarkate area margins).

(iii) Representation of segments as curves.

(iv) Better statistical manipulation of variable length samples.

(v) Automatic Phase Four thresholding, based on assessment of ASAD metrics obtaining after Phase Three Segmentation, and after each cycle of Phase Four Segmentation.

Nevertheless, the Segmentation Method described above is fully functional. It transforms geochemical profiles into multiscale representations, appropriate for input to parsing routines designed to recognise geologically significant features of the original profiles. It thus provides a technique for direct interfacing between large sets of drill data, and Expert Systems designed to interpret such data.

Acknowledgements

ACKNOWLEDGEMENTS: This research was undertaken at the Imperial College of Science and Technology in London, with financial assistance from the Anglo American Corporation of South Africa.

References

1. SMYTH, C.P. Expert System assisted interpretation of geochemical borehole logs. MSc Thesis, Imperial College (London), 1985

2. NII, H.P. and FEIGENBAUM, E.A. Signal to symbol transformation: HASP/SIAP Case Study. The AI Magazine vol. 3, pt 2, 1983 pp. 23-35.

3. WITKIN, A.P. Scale-Space filtering. Proc. 8th Int. J. Conf. on AI 1983 pp. 1019-1022.