

## Sequence slotting for stratigraphic correlation between cores: theory and practice\*

R. Thompson<sup>1</sup> & R. M. Clark<sup>2</sup>

<sup>1</sup>*Dept. Geophysics, Univ. Edinburgh, Mayfield Rd. Scotland EH9 3JZ;* <sup>2</sup>*R. M. Clark, Dept. Mathematics, Univ. Monash, Australia*

Accepted 2 February 1989

**Key words:** sequence slotting, magnetic susceptibility, isotopes, palaeomagnetism

### Abstract

Sequence slotting is an objective numerical method which allows stratigraphic records to be compared and matched. Quantitative core correlation can be easily performed using sequence slotting on many types of paleolimnological and geological data. Dynamic programming algorithms have greatly enhanced the speed with which sequence slotting can be carried out.

We have further modified the sequence slotting method to limit or even to prevent the formation of long blocks in the slotted sequences. Such blocking or clumping has previously restricted the application of sequence slotting in many practical situations. We have applied the modified dynamic sequence slotting technique, based on common path length summation of Euclidean distances, to magnetic susceptibility data, to isotopic measurements and to palaeomagnetic directions on the sphere.

### Introduction

The sequence slotting method permits matching and comparison of records which lack precise dating information but which are internally ordered. Such ordered records, in which the data sequence is known unambiguously, are obtained from many stratigraphic successions such as marine, lake and cave sediments.

When geological marker horizons, for example ash beds, are available, stratigraphic correlations can be made between successions without undue difficulty. In most circumstances, however, such clear geological information is scarce or unavaila-

ble. Stratigraphic comparison between records then tends to be made either (i) 'by eye' by looking for features which appear similar – the technique of 'bump matching', or else (ii) by resorting to time series analysis, such as the calculation of lagged cross correlation coefficients, even though the data on which the calculations are performed are rarely independently dated or equally spaced, or (iii) by erecting a simple model of the depositional relationship between the sequences e.g. a linear stretching. In this paper we investigate a range of stratigraphic situations in which a fourth approach, sequence slotting, may be of value.

A convenient method for visualizing these four

\* This is the tenth of a series of papers to be published by this journal that was presented in the paleolimnology sessions organized by R.B. Davis for the XIIth Congress of the International Union for Quaternary Research (INQUA), which took place in Ottawa, Canada in August 1987. Dr Davis is serving as guest editor of this series.

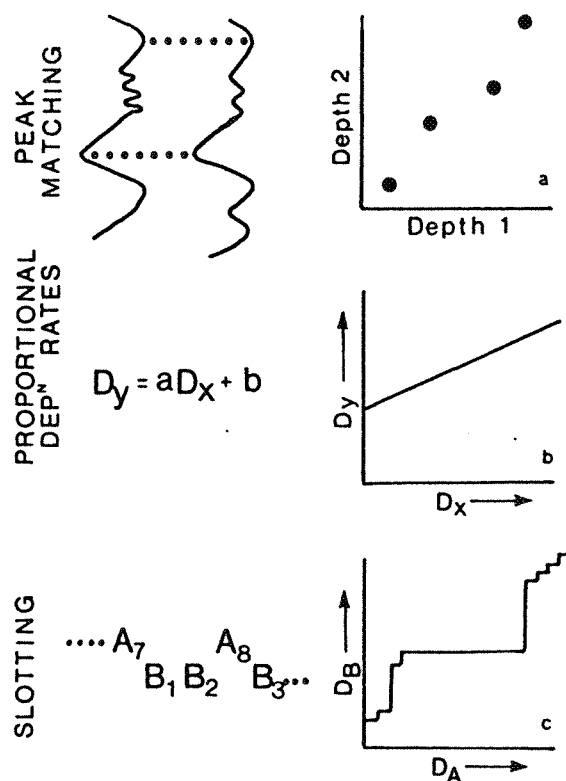


Fig. 1. Core correlation methods. Depth-depth plots at right hand side

- a. Correlation by eye – 'bump matching'.
- b. Simple linear deposition model e.g. time series cross correlation.
- c. Sequence slotting.

approaches is in terms of depth-depth plots (Fig. 1). Visual matching of maxima and minima fixes isolated points on a depth-depth plot (Fig. 1a), whereas the linear relationship of a simple depositional model appears in Figure 1b. Cross correlation using time series analysis also leads to a linear relationship as depicted in Fig. 1b. Sequence slotting produces a step-like plot for the depositional relationship (Fig. 1c).

All four approaches involve a 'search' for a 'good' fit. Such searches necessarily yield appealing fits or high correlation coefficients, because poor fits and low coefficients are rejected in the search procedures. High correlation coefficients discovered in such analyses may be of little significance on account of the excessive searching and data manipulation that may have gone into producing the optimal fits.

We have investigated a variety of approaches to the method of sequence slotting and considered ways in which earlier versions of the technique may be improved or modified for them to handle geological and geophysical data. The most important of these modifications concerns removing the characteristic disjoint-blocking pattern produced by sequence slotting. This annoying effect has previously hindered practical application of the technique. We have applied our modified slotting method to three types of data – isotopic, susceptibility and palaeomagnetic directions – and then discussed the results of these analyses. Finally we examined the quality of match of stratigraphic records caused by search procedures. In particular, we examined the quality of fit of palaeomagnetic records by making use of a slotting statistic and a simulation study.

### Sequence slotting

The sequence slotting procedure involves combining two sequences to form a single joint sequence such that similar objects are placed together from the two sequences, while the original ordering within each sequence is preserved (Gordon, 1973). Dynamic programming methods allow all possible slottings to be assessed in an efficient manner, so that the dissimilarity between the sequences is minimized subject, of course, to the constraints of the original orderings (Delgoigne & Hansen, 1975).

Consider two sequences A and B, of length  $m$  and  $n$  respectively, of ordered observations or objects at stratigraphic horizons denoted by  $A_1, A_2 \dots A_m$  and  $B_1, B_2 \dots B_n$ . Each observation can consist of measurements on  $p$  variables  $X_1, X_2 \dots X_p$ . A dissimilarity measure is then set up to assess the resemblance of the observations to one another. Call the dissimilarity between the  $j$ th object in sequence A and the  $k$ th object in sequence B  $d(A_j, B_k)$ .

One dissimilarity measure (Gordon, 1973) is then

$$d(A_j, B_k) = \sum_{i=1}^p w_i |X_{ij}(A) - X_{ik}(B)|$$

where  $w_i$  are weights on each variable. For data on the unit sphere the arc length between directions can be used as a dissimilarity measure (Embleton *et al.*, 1983) which yields a formula of

$$d(A_j, B_k) = \sum_{i=1}^p w_i (X_{ij}(A) \cdot X_{ik}(B))$$

An alternative measure for points in  $p$  dimensional space is the weighted Euclidean distance between them (Clark, 1985). In this case

$$d(A_j, B_k) = \left( \sum_{i=1}^p w_i (X_{ij}(A) - X_{ik}(B))^2 \right)^{1/2}$$

These different dissimilarity measures may themselves be combined in a variety of ways to estimate the concordance between the two sequences. One assessment of total concordance is to use the sum of local dissimilarity measures (Fig. 2b), namely the dissimilarities between a given point and the immediately preceding and following points in the other sequence (Gordon, 1973; Gordon & Reymont, 1979). Another calculation yields a total concordance measure equivalent to the combined path length (CPL) of Figure 2a. Holmquist (1989) has suggested that there are advantages in summing the products of the local dissimilarities about a given point when forming the total dissimilarity measure, particularly when the correlation between adjacent objects within the sequences is low. Gordon (1980) has also taken the minima of the local dis-

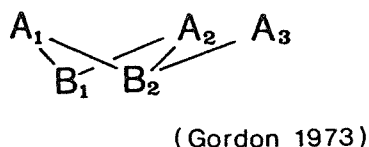
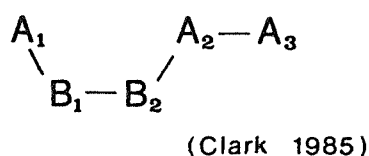


Fig. 2. a. Common path along joint slotting. b. Local dissimilarities between two sequences.

similarities. We have followed Embleton *et al.*, (1983) and Clark (1985), and minimized the total combined path length (CPL). Our path lengths were then based on the Euclidean distances between adjacent horizons in the combined sequence.

The introduction to the sequence slotting method of the dynamic programming technique, by making use of optimality principles, has made sequence slotting extremely efficient. Optimal slotting of two sets of several hundred objects requires only a few seconds on a mainframe computer whichever dissimilarity or concordance measure is adopted. The key to the dynamic programming approach is that the total dissimilarity minimization problem can be broken down into a series of extremely simple subproblems that can be solved in a well structured cascade. The dynamic programming method thus provides a fast, exact algorithm to solve the total discordance minimization problem.

### Standardization

Data standardization is one extremely important practical aspect of using any mathematical approach to match sequences. In almost every practical application of sequence slotting, some form of standardization or data scaling needs to be carried out. Scaling methods developed for multivariate analyses can be directly applied to sequence slotting data. Rummel (1970, p 289-296) for example has reviewed the conventional approaches to scaling in factor analyses. Zhou *et al.* (1983) summarize some of the effects of scaling in factor analysis using both artificial and real data and point out its crucial importance in many practical situations.

We have found no universally satisfactory standardization procedure. Nevertheless, one approach we find relatively reliable for many multivariate data sets is that of (i) transforming the data spread to roughly that of a normal distribution e.g. by using a logarithmic or square root transformation, (ii) subtracting the mean of each sequence individually, (iii) individually setting the

variance of each sequence to unity, and finally (iv) giving each variable equal weight. Steps (ii) to (iv) of this approach correspond to Gordon's (1980) second standardization procedure. On the sphere we standardize by rotating the data to make the mean directions equal. Compositional data need especially careful scaling (see for example, Clark *et al.* (1986)) before sequence slotting is applied.

One must be careful to guard against excessive manipulations of the data during standardization as extreme ordination attempts could lead to spurious data matchings. Nevertheless an iterative approach to standardization can prove useful in some circumstances. For example, it can be used to standardize only over the range of overlap of the slotted sequences. Alternatively we have used iterative scaling to produce significantly better concordance measures through the use of search techniques based on linear programming such as simplex optimization.

Additional approaches to data manipulation that we have experimented with, prior to slotting, include robust smoothing and the calculation of first and second differences or gradients. Unfortunately neither of these manipulations was able to help appreciably with a major difficulty in sequence slotting – namely the problem of clumping or blocking. This blocking problem forms the theme of the next section.

### Unblocking

The heart of the practical problem of adapting available sequence slotting programmes to cope with 'messy geological data' lies with the necessity of removing unrealistically long blocks or groups of horizons from the optimal solutions produced by sequence-slotting. This recurring problem of unnaturally long blocks can arise in two distinct situations. One of these situations is when the parameters being slotted vary little down the core. A slight discrepancy between cores in these circumstances can have a profound influence on the slotting. It can produce long blocks and have the effect of making it appear that the sequences grossly mismatch, when in fact they hardly differ. The other situation in which geologically un-

justifiable blocks are produced is when one variable differs markedly, over a short section, between cores while the other variables exhibit similar fluctuations in the cores. Optimal, unconstrained sequence slotting invariably creates long blocks in this type of situation.

An example of the first of these two mismatch situations, i.e. that associated with low variability within the core, arises in connection with the magnetic susceptibility data considered in the section on multicore slotting. A good example of the second type of blocking mismatch is found in the study of Anderson (1986). He describes how some unusually high concentrations of diatoms in one of his cores have an exaggerated effect on their slot sequencing.

Some form of unblocking technique is needed to get around these two common difficulties in slotting data. Ideally if data standardization could be carried out perfectly, there would be no need to include such unblocking procedures in the sequence slotting algorithms. Without such exact scaling methods, we have to present sequence slotting with more flexibility, by including a parameter which allows the optimal slotting to vary between that of a smooth monotonic depth/depth plot as in Fig. 1b and the unconstrained slotting of Fig. 1c. We have used two approaches to achieve this modification to optimal slotting. The first is to set explicitly the maximum number of consecutive horizons permitted from either sequence. The second is to encourage horizons from the two sequences to interleave with one another by downweighting the between sequence dissimilarities.

An example of the effect of the first explicit block constraint on sequence slotting is shown in Fig. 3. Standardized oxygen isotope ratios are plotted against the position in the joint slotted sequence for two oceanic sediment cores. The uppermost plot shows the unconstrained joint slotting with no restrictions on block length, while the lowermost plot shows the joint slotting with a maximum length for interior blocks, in either core, of three.

Such oxygen isotope records back to around stage 23 are generally regarded to be of high

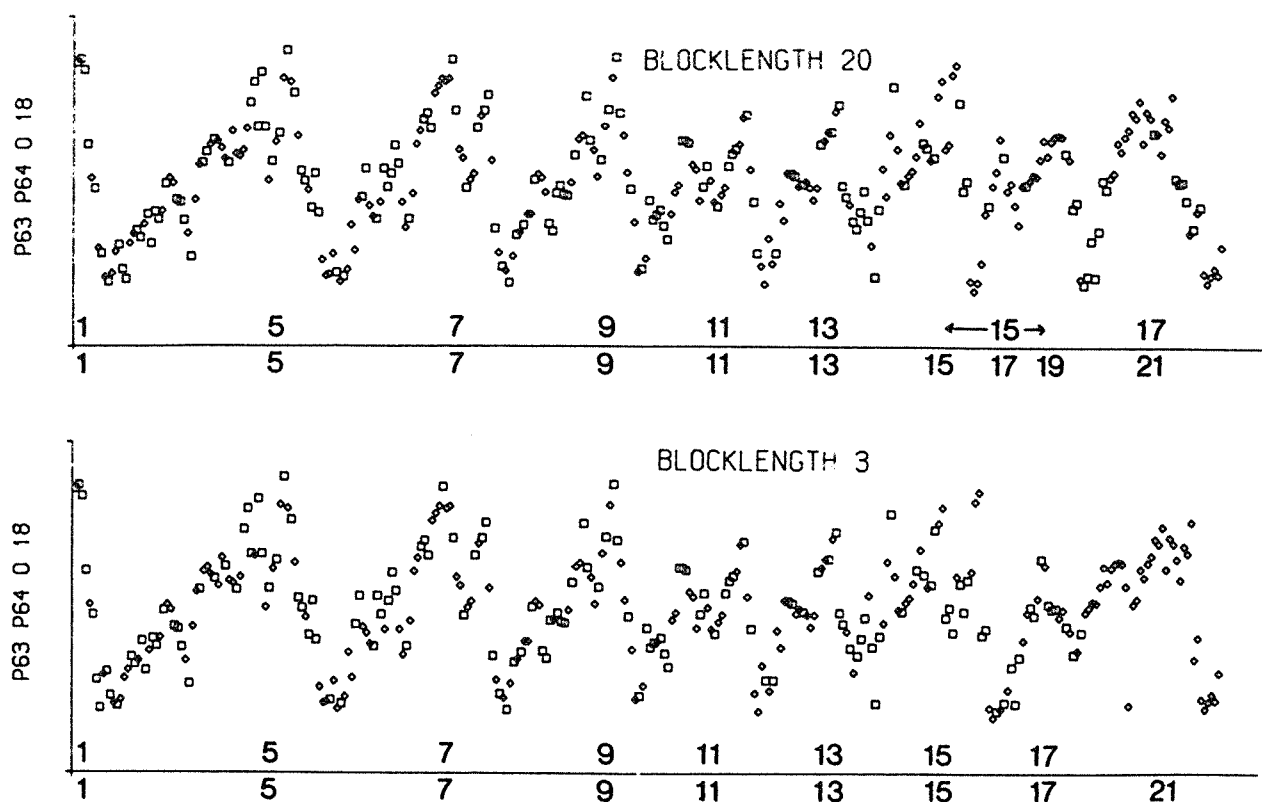


Fig. 3. Standardized oxygen isotope slotting of deep sea cores P6304-9 (squares) and P6408-9 (diamonds). (Data from Emiliani 1966 and 1978).

- a. Optimal unconstrained slotting
- b. Slotting constrained by maximum internal block length of three in either core. Isotopic  $^{18}\text{O}$  maxima labelled along horizontal, common path position axes. Isotopic features match correctly in unblocked analyses of b. See text for discussion of goodness of match statistic delta.

quality (Lowe & Walker, 1984), relatively easy to match by eye and to contain several distinctive features. Twenty-one of these features span approximately the last one million years. Nine successive interglacial maxima are labelled by odd numbers (Fig. 3). The remaining (even numbered) glacial isotopic minima make up the rest of the stage features.

Unconstrained sequence slotting of cores P6304-9 and P6408-9 manages to match the isotopic features correctly between stages 1 and 13 but then to mismatch the features 15 to 21 (Fig. 3a). We can observe in the region of mismatch how long blocks have been created by the slotting algorithm in order for it to obtain a good mathematical fit. Explicitly setting the maximum

internal block length to four creates a very slightly poorer fit, but yields the geologically correct result.

A depth-depth plot (Fig. 4) summarizes how the two isotope records fit together. Maximum internal block lengths of two, three or four all produce closely similar fits compared with the incorrect unconstrained (optimal) slotting (Fig. 4). We have yet to devise a routine method of determining the most appropriate maximum length for the blocks. However, for many earth science data sets, we have found maximum internal block lengths of around three or four give good results.

An example of the use of unblocking through downweighting of between-core dissimilarities,

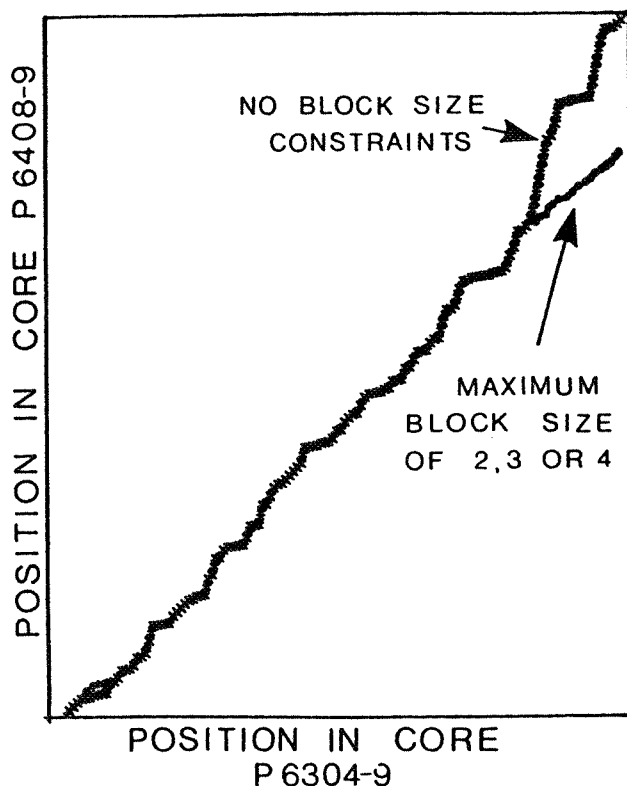


Fig. 4. Depth-depth plot for slottings of cores P6304-9 and P6408-9 as presented in Fig. 3. The slottings with maximum internal block lengths of 2,3 and 4 are all very similar.

compared with within-core dissimilarities, is included in the next section which deals with multicore slotting.

### Multicore slotting

The problem of matching several sequences, rather than just a pair of sequences, can be approached in several ways. The most common approach is to reduce the multicore problem into one involving matching only pairs of cores at a time. This is a natural approach to take with cores collected along a transect e.g. from shallow to deep water. The first sequence is matched with the second sequence (Fig. 5). Then the second sequence is matched to the third, without reference to the first slotting, and so on for all cores along the whole transect (Fig. 5a).

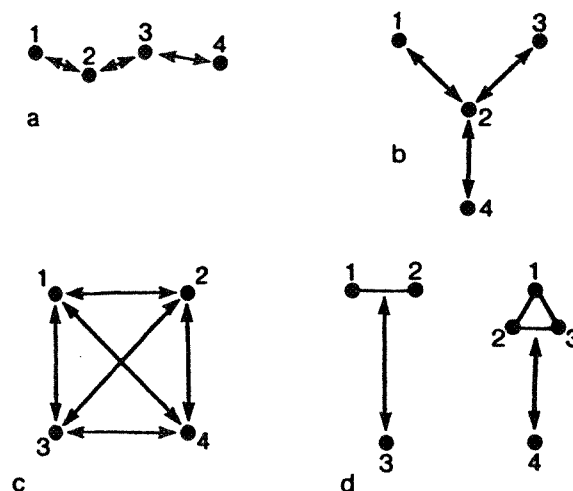


Fig. 5. Multicore slotting strategies

- a. Transect
- b. Master core
- c. True multicore
- d. Optimized iterative

A similar approach is that of selecting a master core (Fig. 5b). Cores are matched individually to the master core. Once again only pairs of cores are dealt with at any one time. A similar drawback to that of core correlation along a transect is that the information recorded by some cores, for example cores 3 and 4 in Fig. 5b, is not utilized in the correlation of the remaining cores (e.g. 1 and 2 in Fig. 5b).

Ideally in multicore slotting, all cores should be matched simultaneously (see Fig. 5c). While such an approach is theoretically possible it has not so far proved tractable in the sequence slotting method. This intractability is because in the FORTRAN computer coding developed to date, the main array sizes increase according to the power of the number of cores to be slotted. Consequently storage space, even on a large mainframe computer, is totally inadequate for practical situations involving four or more cores.

An intermediate approach that we have adopted involves an approximate iterative optimization. As an example we have investigated three cores from Lake Kinneret (Fig. 6). Our multicore slotting method begins by taking two cores, standardizing the data and slotting the cores together by minimizing their combined path

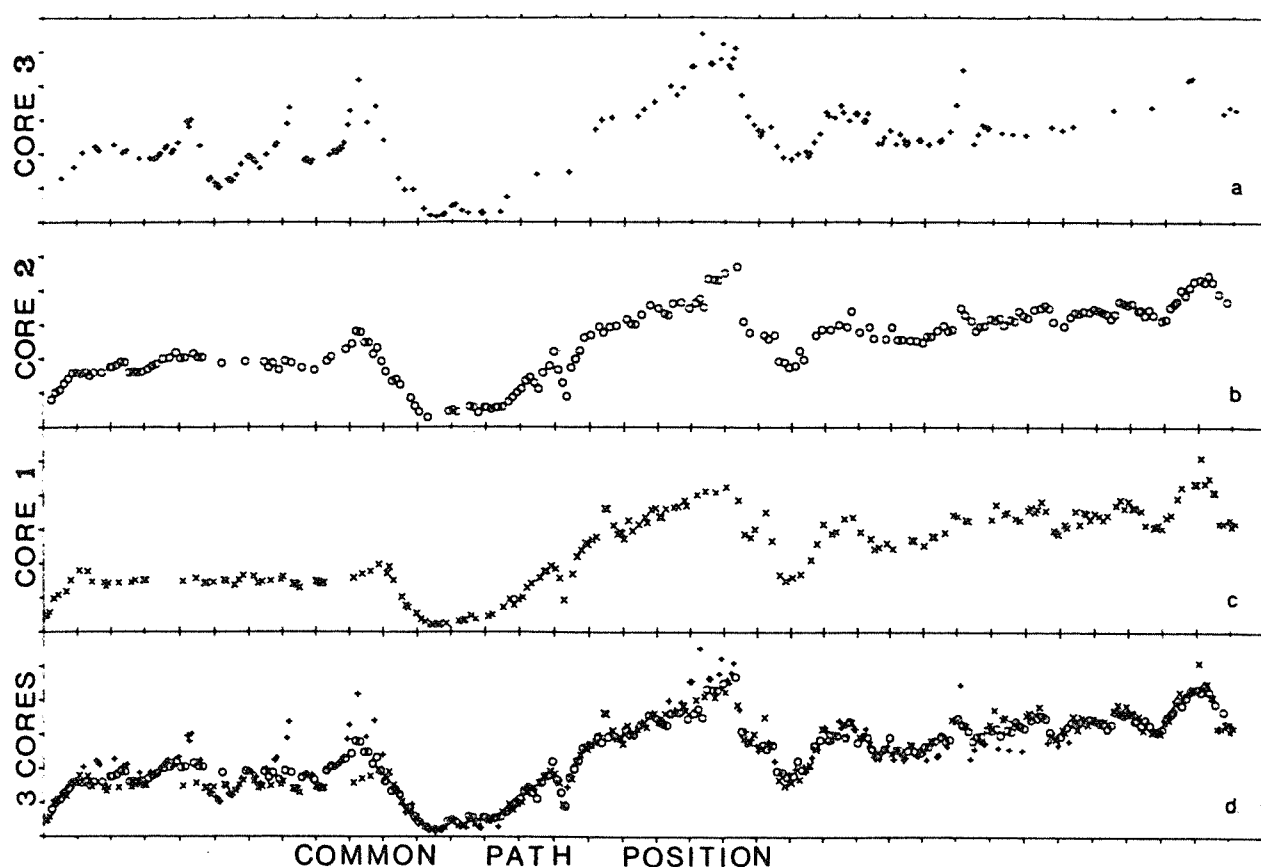


Fig. 6. Standardized Lake Kinneret magnetic susceptibility data from three coring sites. The susceptibility data are plotted against common path position of the multicore slot.

a. b. and c. show the data from individual cores

d. shows the three core slot. Between core dissimilarities have been down-weighted by a factor of 0.5 compared to within core dissimilarities as part of the unblocking procedure.

length exactly as in our sequence slotting procedure for two cores. This matched pair is then treated as one sequence and a third, standardized core slotted to it to produce a combined, joint sequence of three cores. We then begin the iterative optimization procedure by removing the first core but then immediately reslotting it back into the combined sequence of core two plus core three. During this iteration we retain the ordering of core two with core three which remained after removal of the core one horizons from the joint slotting of the three cores. Now, because of the order in which the cores were combined, a different joint slotting will almost certainly have been produced. We iterate a second time. In this step we remove core two from the combined sequence

and as before immediately reslot it. This second iteration step is followed by the same removal and replacement procedure but now using core three. We continue performing these three steps iteratively until the joint slotting of the three cores stabilizes. A fourth standardized core can now be added, and a further set of removal and replacement iterations performed. Our programme is written in such a way that the coding is able to handle any number of cores.

The iterative approach is relatively easy to program. The main modifications needed to the two core program concern (i) keeping track of which horizons in the joint slotting are from which core and (ii) increasing the array sizes. The key point here is that since slotting is always per-

formed pairwise, the array sizes are only marginally larger than that needed for slotting just two sequences.

Although slightly different solutions are arrived at according to the order in which the cores are slotted, the differences in the final multicore slot are rather small. Furthermore a perturbation analysis of the Kinneret susceptibility data (Fig. 6), performed by A. Gordon (Pers. Comm.), using a simulated annealing algorithm, could find no better slotting solution than the best obtained by our iterative procedure. These two results give us confidence that our approximate iterative method produces slottings that are close to the optimal slotting. Indeed in practical terms, especially in multicore work, such approximately optimal slottings are quite satisfactory.

The Kinneret susceptibility data of Fig. 6 are plotted against position in the best combined

slotting. The upper three plots in Fig. 6 show susceptibility for each of the three cores individually, while the lowermost plot shows the combined slotting. Unblocking was again found necessary to achieve a realistic slotting. In this example, unblocking was obtained by down-weighting the between-core dissimilarities by a factor of one half compared to the within core dissimilarities. Without any unblocking, clear mismatches could be seen between the cores, particularly around common path position 400. This mismatching, in the unconstrained slotting, was a consequence of relatively low variability within the cores compared with between-core differences in this part of the sediment succession. However, with the inclusion of downweighting we see how all the major features in the three cores align well (Fig. 6).

Table 1. Quality of match statistic for ordered sequences on the unit sphere

Type of sequence	Data source	Delta	R
Identical data		0.00	1.00
Earth's rotation axis	VLBI, LASER	0.29	0.99
Palaeomagnetic demag. of sediment core	NRM/15mT	0.30	0.90
Geomagnetic Observatory Annual means	Eskdalemuir/Lerwick	0.51	0.98
Etna volcano secular variation	Historical/Lava Flows	0.56	0.92
Palaeomagnetic reversals	Bessasta/Borgarfj	0.60	0.86
Within lake secular var.	Lomond 1/Lomond 3	0.63	0.81
Paleomagnetic secular variation within lake	Windermere 1/Windermere 3	0.65	0.79
ARIMA	101, 100	0.72	0.67
Between lake sec. var.	Windermere/Lomond	0.73	0.70
Sec. var. between lake and stalactite	Akiyoshi/Kinki	0.77	0.65
No overlapping sections		> 1.00	0.00



### Quality of fit

One measure of the quality of fit is the Psi statistic of Gordon (1973) and Gordon and Birks (1974) which uses a standardized measure of the total local dissimilarities. Gordon (1982) suggests a statistic, delta, for use with the total path length approach to sequence slotting. The statistic, delta, is based on the difference between the length of the common path (CPL) and the lengths of the individual paths of core A (APL) and core B (BPL). If Euclidean distance is used as the dissimilarity measure, then delta measures the adequacy of piecewise linear interpolation within blocks in the combined sequence. More specifically,

$$\text{delta} = (2(\text{CPL})/(\text{APL} + \text{BPL})) - 1$$

As an example of the use of such quality of fit statistics, we have compared slottings of data sets on the sphere. Our particular interest was to assess the quality of match of palaeomagnetic secular variation records between lakes (Table 1). Based on this measure, we find poor agreement of palaeomagnetic records between lakes compared with the fits obtained for other types of geophysical data. Furthermore we find no better agreement of palaeomagnetic records between nearby lakes compared with lakes thousands of kilometres apart, as judged using the statistic delta. Such a result would not be anticipated geophysically if lake sediments are good recorders of both geomagnetic declination and inclination.

We also investigated a simulation procedure of assessing quality of fit. After some trials we found that we could generate *random* autoregressive or integrated moving average (ARIMA) series which successfully mimicked the main characteristics of lake sediment palaeomagnetic secular variation data sets (see Fig. 7). An indication of the success of the simulation is that in informal questioning a number of experienced investigators have found some difficulty in distinguishing such ARIMA data from their own 'real' palaeomagnetic measurements. On slotting our ARIMA series with each other, we obtained an average delta

value of 0.67, which is the same delta value that we obtained for our between lake slottings. Our conclusion, once again, is that the palaeomagnetic data from different lakes are of poor quality, indeed of poorer quality than generally judged subjectively.

A further method of investigating quality of match is through an interpolated correlation coefficient, calculated following slotting. The right hand column of Table 1, tabulates such correlation coefficients for the data on the sphere that we have investigated. Low correlation coefficients, below about 0.7 (as found for randomly generated ARIMA data), reflect poor matches while the closer the coefficient to 1.0 the better the overall fit.

The above analyses do not make use of any dating information, only the stratigraphic ordering. This points to the importance of accurate, independent dating controls in geophysically assessing palaeomagnetic data. The analyses demonstrate how difficult it is to match lake sediment records using only the shapes of the palaeomagnetic secular variation curves. With such stratigraphic data, the human eye appears to be extremely good at 'picking out' correlations when they really exist, but to be rather poor at determining when data sets actually differ. The statistic delta may thus be particularly useful in quantifying quality of match of 'poor' data sets.

### Discussion and comparison of alignment methods

Any method for cross-matching two sequences or series will produce an allegedly 'optimal' matching, no matter what data are used. A difficulty common to all such alignment methods is how to assess the practical significance of the final 'optimized correlation' or match.

In the case of sequence slotting the randomization test given by Clark (1988) provides a simple statistical test for assessing how well any given pair of sequences are matched, based on the value of delta. In this test, each sequence is split *at random* into two sub-sequences, each of which is then slotted against every other subsequence.

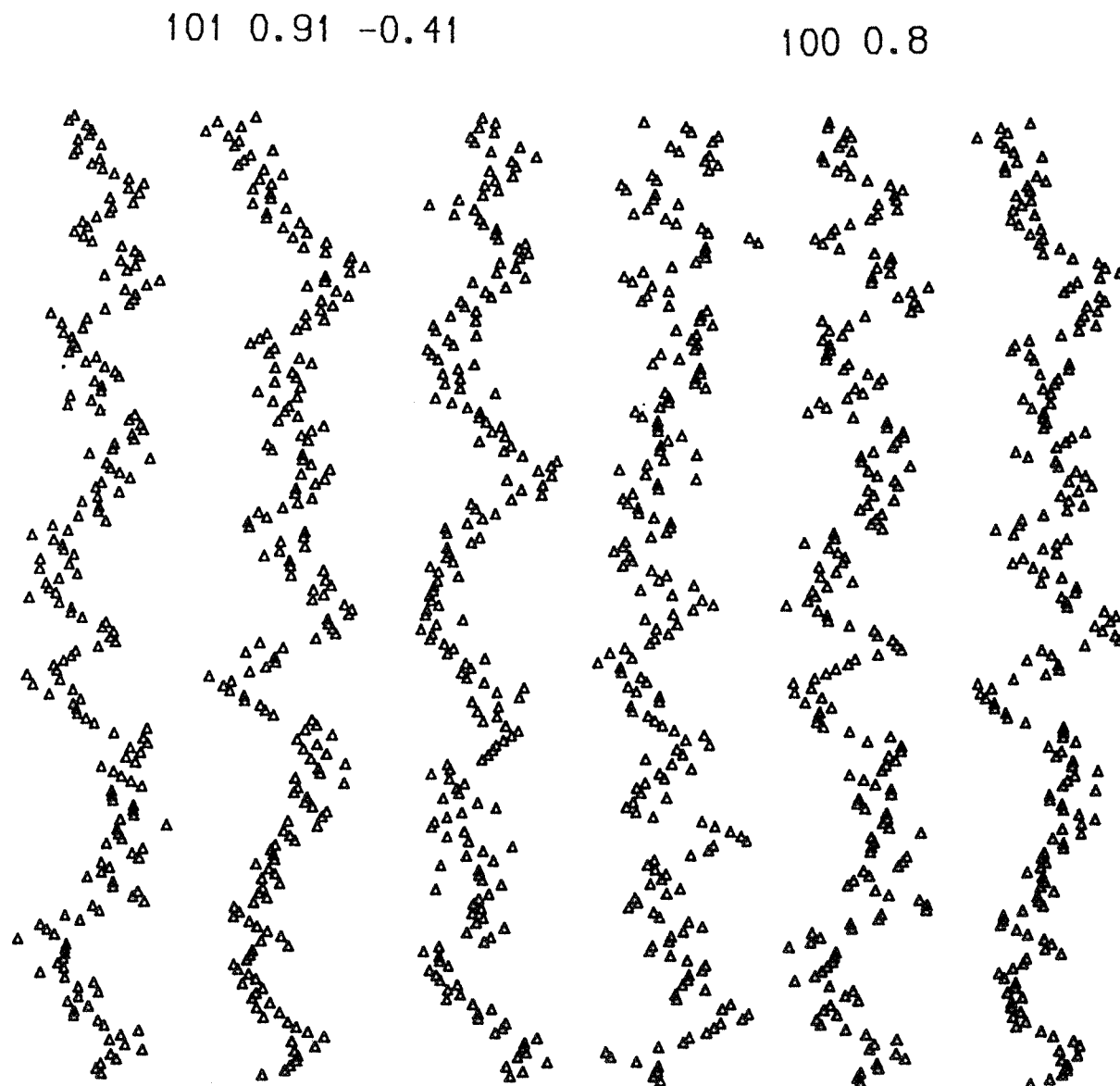


Fig. 7. ARIMA series. Data sets generated to mimic palaeomagnetic secular variation time series on the sphere. The three left hand examples shown were generated using a combined autoregressive moving average process with coefficients of 0.91 and  $-0.41$  respectively, while the three right hand examples were produced from a 100 process with an AR coefficient of 0.8. The choice of coefficients is not especially critical for generating series which resemble down core palaeomagnetic records.

This procedure is then repeated, say 40 or 50 times. The test then measures how well the original sequences and the combined sequence are reconstructed from these random sub-sequences. If the combined sequence can *not* be reconstructed as well as the original sequences individually, this indicates that the sequences do

*not* arise from a common signal, i.e., the matching is spurious.

In principle this test leads to *exact* significance levels. It is easy to implement, does not involve any assumptions regarding the form of the fitted signal or the stretching function, or any simulation of data according to some assumed model. The

examples in Clark (1988) show that the test works well and can be remarkably sensitive.

A variety of techniques have been developed in connection with the alignment or correlation of time series that can be applied to stratigraphic data. We finally briefly discuss the differences between three such techniques and note some of their advantages and disadvantages.

1. Clark & Thompson (1978) – linear stretching with cubic spline smoothing of the observed data
2. Martinson *et al.* (1982) – Fourier series estimation of the mapping function with coherence to be maximised
3. Present Paper – Sequence-slotting with unblocking.

These three methods were developed to do three different things namely:

*Method 1*: to estimate a linear stretching function from noisy discrete data in which both 'signals' are distorted versions of some unknown reference signal;

*Method 2*: to estimate a general mapping function from two continuous signals with high signal-to-noise ratio, with one signal regarded as a known reference signal;

*Method 3*: to match two ordered sequences of observations (with multivariate responses), using only the relative *order* of the observations, and taking account of other stratigraphic constraints.

All three methods require, in their implementation, a number of more-or-less arbitrary decisions, e.g.,

*Method 1*: the number and location of the knots used to define the cubic splines fitted to the data.

*Method 2*: the number of Fourier coefficients, the constants  $\Delta a$  and  $k$  in equation 10, the degree of smoothing or filtering of the data prior to maximizing the coherence.

*Method 3*: the choice of distance measure and of 'local discordance', the preliminary standardization, the weights to be assigned to different variables.

Methods 1 and 2 require an iterative trial-and-error approach starting from an initial guess; in contrast method 3 uses a direct and efficient algorithm with no iteration (once the maximum block-length is specified).

Method 1 has the following advantages:

Being based on well-known and well-understood least-squares methods, its assumptions can be readily tested, and confidence limits for both the stretching function and the fitted signal can be found. The method treats the sequences symmetrically (unlike Method 2), in the sense that both sequences are treated as distorted versions of a common unknown signal. The least-squares criterion is more sensitive than the correlation coefficient, (see the examples and discussion in Clark & Thompson, 1978). In principle the method can test the adequacy of ANY given monotonic stretching function, but in practice this is difficult to do.

Its disadvantages are:

1. The choice of knots is arbitrary, although some guidelines are available to assist this choice.
2. A large number of parameters must be fitted.
3. A trial-and-error approach is needed to find the best values of alpha and beta (which define the stretching function).

Method 2 clearly works well in the situation for which it was designed. Also it has been claimed to be superior to earlier versions of sequence slotting for matching palaeomagnetic data. However, it is not clear how well it would work with discrete data with a low signal-to-noise ratio or how it should be modified in order to deal with multivariate data. It has the following disadvantages:

1. Like method 2, it requires several arbitrary decisions and an iterative approach starting from an initial guess.
2. It cannot handle (i) cases in which only the order of the observations is known, (ii) stratigraphic constraints, and (iii) simultaneous matching of several sequences.
3. There is no guarantee that the fitted mapping function will turn out to be monotonic.
4. No statistical test is provided for interpreting or assessing the final maximised coherence  $C$ .

Method 3 has the following advantages:

1. It directly produces the optimum 'path' as well as an estimate of the stretching or mapping function.

2. It uses a fast and efficient algorithm, which requires no iteration, no initial guess, and no arbitrary prior smoothing of the data.
3. It can handle multivariate signals or responses.
4. Additional stratigraphic information is easily incorporated.
5. Only the *order* of the points in each sequence need be known; the corresponding depths or times are not needed.
6. The randomization test provides a conceptually-simple and easily-implemented statistical test of how well the series are matched or slotted together.
7. In its extended version (see Gordon, Clark & Thompson, 1988) it is possible to assess which parts of the slotted sequence are well-determined (and which parts are not), and to investigate the influence on the results of deleting single possibly-aberrant observations.

Method 3 has the following disadvantages:

1. The choice of distance function, relative weighting for different variables (in the multivariate case), initial standardization and maximum block length are arbitrary.
2. Sometimes it is not easy to interpret the final delta-value.
3. The algorithm requires a lot of computer memory, and in its extended form would be impractical on a mainframe computer if the sequences had more than 1000 points each.
4. It does not produce confidence bands for the fitted path or the implied stretching function.
5. Continuous signals would have to be digitized before analysis – so introducing another unwanted arbitrary choice.

## References

- Anderson, N.J., 1986. Diatom biostratigraphy and comparative core correlation within a small lake basin. *Hydrobiologia* 143: 105–112.
- Clark, R.M., 1985. A FORTRAN program for constrained sequence-slotting based on minimum combined path length. *Computers & Geosciences* 11, 605–617.
- Clark, R.M., 1988. A randomisation test for the comparison of ordered sequences. Research Report 177. Dept. Mathematics, Monash Univ.
- Clark, R.M. & R. Thompson, 1979. A new statistical approach to the alignment of time series. *Geophys. J. R. Astron. Soc.* 58: 593–607.
- Clark, J.S., J.T. Overpeck, III, T. Webb & III, W.A. Patterson, 1986. Pollen stratigraphic correlation and dating of barrier beach peat sections, *Rev. of Palaeobot. Palynol.* 47: 145–168.
- Delcoigne, A. & P. Hansen, 1975. Sequence comparison by dynamic programming. *Biometrika* 62: 661–664.
- Embleton, B.J.J., N.I. Fisher & P.W. Schmidt, 1983. Analytic comparison of apparent polar wander paths. *Earth Planet. Sci. Letters* 64: 276–282.
- Emiliani, C., 1966. Paleotemperature analysis of Caribbean cores P6304-8 and P6304-9 and a generalized temperature curve for the past 425 000 years. *J. Geol.* 74: 109–124.
- Emiliani, C., 1978. The cause of the ice ages. *Earth Planet. Sci. Lett.* 37: 349–352.
- Gordon, A.D., 1973. A sequence-comparison statistic and algorithm. *Biometrika* 60: 1, 197–200.
- Gordon, A.D., 1982. An investigation of two sequence-comparison statistics. *Aust. J. Statist.* 24: 332–342.
- Gordon, A.D. & H.J.B. Birks, 1974. Numerical methods in Quaternary palaeoecology. II: Comparison of pollen diagrams. *New Phytologist* 73: 221–249.
- Gordon, A.D., R.M. Clark & R. Thompson, 1988. The use of constraints in sequence slotting. In *Data analysis and informatics 5*. (Eds. E. Diday *et al.*) North Holland (In Press).
- Gordon, A.D. & R.A. Reymont, 1979. Slotting of borehole sequences. *Mathematical Geology* 11: 3, 309–327.
- Holmquist, B., 1989. Sequence-comparison statistics and dissimilarity measures for slotting two sequences. (Unpub. Manuscript).
- Lowe, J.J. & M.J.C. Walker, 1984. *Reconstructing Quaternary Environments*. London: Longman.
- Martinson, D.G., W. Menke & P. Stoffa, 1982. An inverse approach to signal correlation. *J. Geophys. Res.* 87: 4807–4818.
- Rummel, R.J., 1970. *Applied Factor analysis*: Northwestern Univ. Press, Evanston, Ill., 617 p.
- Zhou, D., T. Chang & J.C. Davis, 1983. Dual extraction of R-Mode and Q-Mode factor solutions, *Math. Geol.* 15: 5, 581–606.