

- N., 1979, Thermal effects of ridge subduction: Earth and Planet. Sci. Letters, v. 44, p. 239-246.
- HASEGAWA, A.; UMINO, N.; and TAKAGI, A., 1978a, Double-planed structure of the deep seismic zone in the northeastern Japan arc: Tectonophysics, v. 47, p. 43-58.
- ; —; and —, 1978b, Double-planed deep seismic zone and upper mantle structure in the northeastern Japan arc, Geophys. Jour. Roy. Astr. Soc., v. 54, p. 281-296.
- MARSH, B. D., 1979, Island-arc volcanism: Amer. Scientist, v. 67, p. 161-172.
- MCKENZIE, D. P., 1969, Speculations on the consequences and causes of plate motions: Geophys. Jour. Roy. Astr. Soc., v. 18, p. 1-32.
- MYSEN, B. O., 1978, The role of descending plates in the formation of andesitic melts beneath island arcs: Carnegie Inst. Washington Year Book, 77, p. 797-801.
- SCHOLZ, C. H., 1968, The frequency-magnitude relation in microfracturing in rock and its relation to earthquakes: Bull. Seis. Soc. Amer., v. 58, p. 399-415.
- WATTS, A. B., 1978, An analysis of isostasy in the world's oceans, 1. Hawaiian-Emperor seamount chains: Jour. Geophys. Res., v. 83, p. 5989-6004.
- WYLLIE, P. J., 1979, Magmas and volatile components: Amer. Mineral, v. 64, p. 469-500.
- YOSHII, T., 1975, Proposal of the "Aseismic Front" (in Japanese): Jour. Seism. Soc. Japan, ser. II, v. 28, p. 365-367.

NEW STRATIGRAPHIC CORRELATION TECHNIQUES¹

TEMPLE F. SMITH AND MICHAEL S. WATERMAN

Northern Michigan University, Marquette, MI 49855; Los Alamos Scientific Laboratory, Los Alamos, NM 87545

ABSTRACT

Cross-association is a standard correlation technique. It is valid for theoretically complete stratigraphic sequences, but not for stratigraphic sequences as they commonly occur. A major problem is its inability to include correlations across gaps caused by local nondeposition or eroded strata. The methods proposed in this paper do include correlations across gaps and allow for completely general measurements of strata similarity. In addition, they are extended to include correlation of a fragmentary sequence with a longer complete sequence and to include correlations of more than two sequences. Using these techniques, most problems can be handled by hand calculation.

Correlation of stratigraphic sequences is an important problem. (In this paper, correlation is always used in the geological sense, not the statistical.) Stratigraphic sequences can be obtained from drill holes, outcroppings, or even road cuts. The data, in the simplest cases, are stratigraphic sequences of lithologic units, but may include much additional information such as strata thickness, geochemical or mineral assay, fossil occurrence and abundance, and electric, gamma, or gamma-gamma logs. A few elementary automated correlation techniques have been developed to utilize this information. The main one currently in use seems to have originated with Sackin, Sneath, and Merriam

(1965). This technique is called cross-association (Harbaugh and Merriam 1968; Merriam 1971; Davis 1973). In this paper, correlation techniques are presented which overcome some of the shortcomings of cross-association.

Cross-association is applied to pairs of sequences. The two sequences are "slid" by one another to locate the maximum of the ratio of matches to the number of comparisons (Davis 1973). The position of this maximum defines the correlation between the sequences. For theoretically complete stratigraphic sequences the method is valid, but it is inadequate for stratigraphic sequences as they commonly occur. The method is inadequate because of its inability to include correlations across gaps (caused by local non-deposition or eroded strata) or to include correlations of repetitive or missing sequences (caused by faulting or folding). All

¹ Manuscript received July 25, 1979; revised February 12, 1980.

gaps in the stratigraphic record are equivalent in their effect on the information content of the record. Thus the first problem is locate the gaps themselves and leave the study of their causes to a later time. Cross-association and other methods have been applied to cases where multiple parameters of each strata are known (Sackin et al. 1965; Sneath 1974, 1975) but the same limitations hold.

A variety of correlation techniques have been presented but none overcomes the difficulties mentioned above (Gill 1970; Matuszak 1972; Neidell 1969; Preston and Henderson 1964; Rudman and Lankston 1973). However, a related gap problem has been successfully treated in studies of evolution of genetic sequences in molecular biology (Sellers 1974; Waterman et al. 1976). Such studies form the basis of all molecular taxonomy and these efforts motivate the algorithms of this paper.

Let $s = s_1 s_2 \dots s_n$ and $r = r_1 r_2 \dots r_m$ be two stratigraphic sequences where s_i and r_j represent lithologic characteristics of strata i and j . The basic problem is to find the best alignment or correlation between any two such sequences. We define an alignment between s and r by $a(s, r) = \{(s_{i_1}, r_{j_1}), (s_{i_2}, r_{j_2}), \dots$ where $i_1 < i_2 < \dots$; $j_1 < j_2 < \dots$ and the pair (s_{i_1}, r_{j_1}) indicates correlation between s_{i_1} and r_{j_1} . For display of an alignment, we include the null element Δ to indicate gaps. For example,

$$\begin{array}{cccccccc} \Delta & \Delta & s_1 & s_2 & s_3 & s_4 & & \\ r_1 & r_2 & r_3 & r_4 & \Delta & r_5 & & \end{array} \quad (A_1)$$

can be written as $a(s, r) = \{(s_1, r_3), (s_2, r_4), (s_4, r_5)\}$ and means s_1 is correlated with r_3 , s_2 with r_4 and s_4 with r_5 .

While the methods we present can include general measurements of strata similarity, we begin with a discussion of a simpler situation. Here, the object is to find the alignment with the maximum number of perfect matches. Also, gaps here will denote single missing strata. Now, notice that, with sequences of length n and m , the number of gaps is given by

$$\text{gaps} = n + m - 2 \text{ (correlated pairs)}. \quad (1)$$

The factor of two comes in because each correlated pair involve two strata elements, while gaps involve only one. If all correlated

pairs represent perfect matches, we obtain the relationship

$$\text{max (matches)} = \frac{n + m}{2} - \frac{1}{2} \text{ min (gaps)} \quad (2)$$

In this case, obtaining the maximum number of matches is equivalent to minimizing the number of gaps. In the more general case of allowing mismatching (correlation, e.g., where lithologic types are identical but thicknesses are not) among the correlated pairs, we obtain

$$\begin{aligned} \text{max (matches)} \\ = \frac{n + m}{2} - \frac{1}{2} \text{ min (gaps + 2 mismatches)}. \end{aligned} \quad (3)$$

The minimum of the number of gaps plus twice the number of mismatches is a measure of distance between the sequences. Equation (3) shows the equivalence of maximizing the matches (or homology) and minimizing the distance.

The above analysis is the basis for our approach to a correlation which finds the minimum distance between sequences. In simple situations, such as considered above, the minimum distance approach is equivalent to the maximum homology approach. More generally, minimum distance is a (mathematical) metric and has the advantage of understood properties. If $D(r, s)$ is the minimum distance between two sequences, then

$$\begin{aligned} D(r, s) &= D(s, r) \\ D(r, s) &= 0 \text{ if and only if } r = s \end{aligned} \quad (4)$$

$$D(r, s) \leq D(r, t) + D(t, s) \text{ for any sequence } t.$$

Use of distance correlation has the advantage that it allows the direct inclusion of degree of mismatch between the correlated pairs in an alignment. This is clearly necessary as different strata seldom are perfect matches and the degree of mismatch should be included in any general analysis. We accomplish this by introducing two distance (or weighting) functions. First, let d_{ij} be the distance between strata s_i and strata r_j . Also, let g_l be the distance associated with a gap opposite the strata l . The distances themselves should obey the properties given in Equation (4).

An algorithm (or method) for computing distance correlation is now presented. To begin, let D_{ij} be the distance $D(s_1 s_2 \dots s_i, r_1 r_2 \dots r_j)$ between the initial segments of s and r , ending at i and j . The following equation gives the algorithm for obtaining the D_{ij} values from earlier computed values (Sellers 1974),

$$D_{ij} = \min \{D_{i,j-1} + g_j, D_{i-1,j-1} + d_{ij}, D_{i-1,j} + g_j\}. \quad (5)$$

The algorithm is initialized by setting $D_{0k} = D_{k0} = k$ for all $k \geq 0$. Equation (5) follows from the fact that either (1) r_j is associated with a gap where the distance is

$$D_{i,j-1} + g_j,$$

or (2) s_i matches r_j where the distance is

$$D_{i-1,j-1} + d_{ij},$$

or, finally, (3) s_i is associated with a gap and the distance is

$$D_{i-1,j} + g_i.$$

To illustrate this algorithm we consider an example from Harbaugh and Merriam (1968, p. 173). Two sequences from Kansas roadcuts are analyzed in figure 1 where the entire matrix of D_{ij} values are given. For this example, the values of d_{ij} and g_i are given by

$$d_{ij} = \begin{cases} 0 & \text{if } s_i = r_j \\ 3 & \text{if } s_i \neq r_j \end{cases}$$

and

$$g_i = 1 \text{ for all strata } t_i.$$

We note that any value of $d_{ij} > 2$ prevents the inclusion of mismatches in the alignments.

As the matrix in figure 1 shows, $D(s, r) = D_{14,13} = 5$. The optimal alignment associated with $D_{14,13} = 5$ is obtained by tracing back through the matrix to obtain the set $a(s, r)$ of correlated pairs. This traceback procedure is shown in figure 1 and results in the alignment

$$\begin{array}{cccccccccccc} E & F & A & B & F & E & F & B & \Delta & \Delta & A & B & F & E & F & B \\ E & F & \Delta & B & F & E & F & B & E & F & A & B & F & E & \Delta & \Delta \end{array} \quad (A_2)$$

The alignment A_2 should be compared with the two alignments that might result from a cross-

association analysis. Each of these two alignments have six matches:

$$\begin{array}{cccccccccccc} E & F & A & B & F & E & F & B & A & B & F & E & F & B \\ E & F & B & F & E & F & B & E & F & A & B & F & E \end{array} \quad (A_3)$$

and

$$\begin{array}{cccccccccccc} E & F & A & B & F & E & F & B & A & B & F & E & F & B \\ E & F & B & F & E & F & B & E & F & A & B & F & E \end{array} \quad (A_4)$$

The second of these alignments, A_4 , would be preferred under the assumption that record gaps are rare. Near the end of this paper we give evidence based on another Kansas sequence that A_2 should be prepared.

An extension of the above algorithm to include contiguous gaps of length greater than one can easily be made (Waterman et al. 1976). This algorithm is given by

$$D_{ij} = \min \left\{ \min_{1 \leq k \leq j-1} \left\{ D_{i,j-k} + g_{j-k+1, j} \right\}, D_{i-1,j-1} + d_{ij}, \min_{1 \leq k \leq i-1} \left\{ D_{i-k,j} + g_{i-k+1, i} \right\} \right\}. \quad (6)$$

Here $g_{l,m}$ denotes a gap of length $m - l + 1$ associated with sequence elements l through m . If gaps of length k are given weight $1 + 0.1(k - 1)$, then the above two sequences have distance $D_{14,13} = 3.2$ and the alignment is

$$\begin{array}{cccccccccccc} E & F & A & B & F & E & F & B & (\Delta) & A & B & F & E & (F & B) \\ E & F & \Delta & B & F & E & F & B & (E & F) & A & B & F & E & (\Delta) \end{array} \quad (A_5)$$

There are three gaps: one of length one and two of length two.

Often stratigraphic sequences are rather fragmentary and the problem is to identify the correlation of a shorter fragment with a longer complete sequence. The proper task is *not* to find the minimum distance between the fragment and the complete sequence, but rather to find the segments of the long sequence which are the minimum distance from the fragment (Sellers in press). The idea is that regions preceding and following a potential correlation should not be weighted as geological gaps, but as gaps in our information. This can be accomplished in the matrix by initializing the first row associated with the longer sequence with zeros.

Wynandotte County Roadcut

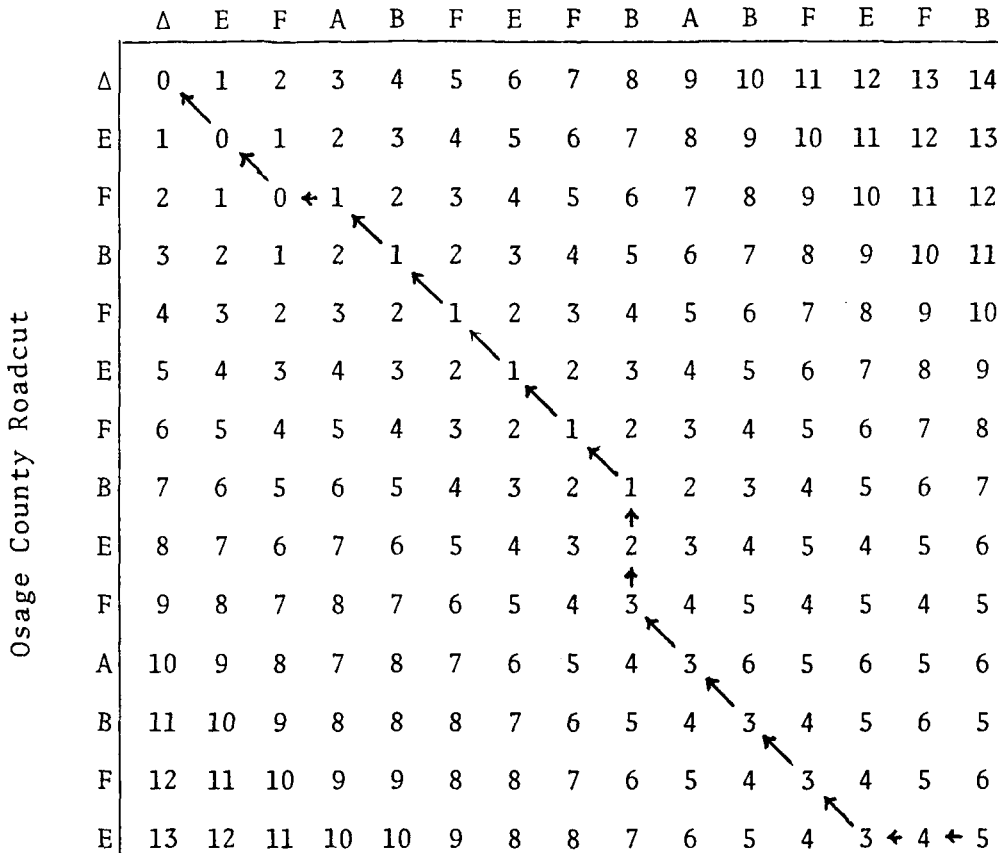


FIG. 1.—Calculation of distance or correlation between a stratigraphic sequence EFABFEFBABFEFB from a Wynandotte County, Kansas roadcut with the sequence EFBFEFBEFABFE from an Osage County, Kansas roadcut. The letter code is F = limestone, E = marine shale, D = coal, C = underclay, B = non-marine shale, and A = sandstone (Harbaugh and Merriam 1968; p. 173). The traceback to determine alignment is the path shown by arrows in the figure.

An example of correlation of a shorter sequence with a long sequence occurs in Davis (1973, p. 255). The example has a mine section (the long sequence) from central England and a quarry section six miles away. Our analysis with the traceback is given in figure 2. The two optimal alignments are

245634531453453454532 4534531454563

45345453212453 (A₆)

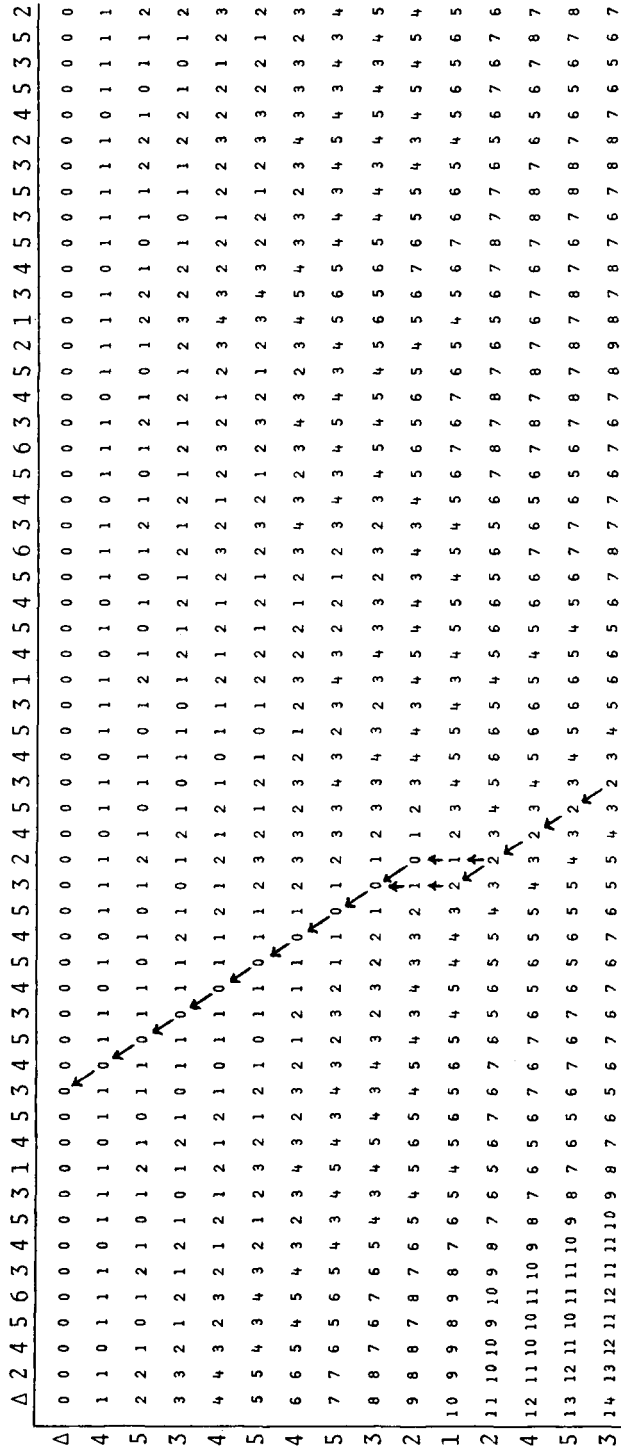
and

24563453145345345453 24534531454563

45345453212453 (A₇)

Obtaining correlations between more than two sequences is of considerable interest and is our last topic. While no automated techniques have been developed, attempts have been made by considering all pairs of sequences. While

Mine Section



Quarry Section

FIG. 2.—Calculation of minimum distance or correlation between a quarry section (the short sequence) and any segment of a mine sequence (the longer sequence) from Central England. The code is 1 = sandstone, 2 = siltstone, 3 = nonfossiliferous shale, 4 = underclay, 5 = clay, 6 = fossiliferous shale, 7 = limestone (Davis 1973, p. 255). The traceback to determine alignments is the paths shown by arrows.

pairwise correlations may have several optimal alignments, it is assumed that additional pairs will reduce these degeneracies. Attempts have been made on this problem by using the pairwise correlations in a dendrographic analysis but the approach has not been too satisfactory (Merriam and Sneath 1966; Harbaugh and Merriam 1968).

Fortunately, these problems have been considered for molecular sequences and can be adapted for geological use (Waterman et al. 1976). The algorithm is an extension of Equation (5). As computation time grows rapidly with the number of sequences, we only present the algorithm for correlation of three sequences. To present this algorithm we need to define a new set of lithology unit weighting functions $d(a, b, c)$ analogous to the d_{ij} 's and g_k used for pairwise correlations. One can think of the $d(a, b, c)$'s as the "distance" among three stratigraphic elements (including gaps). For example, we might choose

$$\begin{aligned} d(a, \Delta, \Delta) &= d(\Delta, a, \Delta) = d(\Delta, \Delta, a) = 1 \\ d(a, a, \Delta) &= d(a, \Delta, a) = d(\Delta, a, a) = 1 \\ d(a, a, a) &= 0 \\ d(a, b, c) &> 3 \text{ for no two elements equal} \end{aligned} \tag{7}$$

where the proposed values are just a count of the number of elements which would have to be replaced to result in three identical elements. If we restrict our considerations, as before, to only perfect matches and/or gaps, the value for $d(a, b, c)$ in the last line must be set greater than three. For these sequences, $s = s_1 s_2 \dots s_n$, $r = r_1 r_2 \dots r_m$, $t = t_1 t_2 \dots t_l$, the algorithm is given by

$$\begin{aligned} D_{ijk} &= \min\{D_{i-1, j, k} + d(s_i, \Delta, \Delta), \\ &D_{i, j-1, k} + d(\Delta, r_j, \Delta), D_{i-1, j-1, k} + d(s_i, r_j, \Delta), \\ &D_{i, j, k-1} + d(\Delta, \Delta, t_k), D_{i-1, j, k-1} + d(s_i, \Delta, t_k), \\ &D_{i, j-1, k-1} + d(\Delta, r_j, t_k), \\ &D_{i-1, j-1, k-1} + d(s_i, r_j, t_k)\}. \end{aligned} \tag{8}$$

For an instructive application, we add a Crawford County, Kansas open pit sequence to the pair of Kansas sequences considered in

figure 1. Let $D_{i, j, k}$ be the 3-distance among the first i strata from Osage County, the first j strata from Wynandotte County, and the first k strata from Crawford County. As the Crawford County sequence is longer and not too closely related to the other two sequences, we solve the problem of best correlation of the two shorter sequences with a segment of the longer one. This is handled by choosing

$$D_{0, 0, k} = 0, D_{0, j, 0} = j, \text{ and } D_{i, 0, 0} = i.$$

Applying Equation (8) yields the values $\min_k D_{13, 14, k} = 11$ with three alignments. One of the alignments is

$$\begin{aligned} &BDBFBDCBDC\Delta F\Delta BFDC\Delta FB\Delta\Delta ABFECDFBD \\ &EFABF\Delta\Delta EFBA\Delta\Delta ABFE\Delta\Delta FB \\ &EF\Delta BF\Delta\Delta EFBEFABFE, \end{aligned} \tag{A_8}$$

while the other two can be obtained by easy rearrangements. Notice that this alignment is consentaneous with the earlier distance correlation alignment A_2 for the two shorter sequences.

These algorithms should make automated correlation much more practical and include gaps in that analysis. In addition, these algorithms can be performed by hand for the size problems presented in this paper. Of course, only the simplest of geological transformations have been used here. To include transformations such as folding much more work needs to be done to develop algorithms. The values of d and g can only be assigned by utilizing geological experience and information. If maximum homology is desired rather than minimum distance, then, by changing minimum to maximum, the above algorithms are easily adapted.

Finally, the potential of the proposed methods is apparent in the agreement noted between the alignment of A_8 of three sequences and the pairwise alignment A_2 . Alignment A_2 now seems more probable than the cross-association alignment A_4 which might have been preferred in the absence of A_8 . Full validation awaits application by geologists to the wide variety of existing data.

REFERENCES CITED

- DAVIS, J. C., 1973, *Statistics and data analysis in geology*: New York, John Wiley, 550 p.
- GILL, D., 1970, Application of a statistical zonation method to reservoir evaluation and digitized log analysis: *Am. Assoc. Petroleum Geologists Bull.*, v. 54, p. 719-729.
- HARBAUGH, J. W., and MERRIAM, D. F., 1968, *Computer applications in stratigraphic analysis*: New York, John Wiley, 282 p.
- MATUSZAK, D. R., 1972, Stratigraphic correlation of subsurface geological data by computer: *Jour. Math. Geology*, v. 4, p. 331-343.
- MERRIAM, D. F., 1971, Computer applications in stratigraphic problem solving in *Decision Making in the Mineral Industry*: Canadian Inst. Min. Metal, v. 12, p. 139-147.
- , and SNEATH, P. H. A., 1966, Quantitative comparison of contour maps: *Jour. Geophysical Res.*, v. 71, p. 1105-1115.
- NEIDELL, N. S., 1969, Ambiguity functions and the concept of geological correlation: *Kansas Geol. Survey Computer Contributions* 40, p. 19-29.
- PRESTON, F. W., and HENDERSON, J. H., 1964, Fourier series characterizations of cyclic sediments for stratigraphic correlation: *Kansas Geol. Survey Bull.*, v. 169, p. 415-425.
- RUDMAN, A. J., and LANKSTON, R. W., 1973, Stratigraphic correlation of well logs by computer techniques: *Am. Assoc. Petroleum Geologists Bull.*, v. 57, p. 577-588.
- SACKIN, M. J.; SNEATH, P. H. A.; and MERRIAM, D. F., 1965, ALGOL program for cross-association of nonnumeric sequences using a medium size computer: *Kansas Geological Survey Sp. Dist. Publ.* 23, 36 p.
- SELLERS, P. H., 1974, On the theory and computation of evolutionary distances: *SIAM Jour. Appl. Math.*, v. 26, p. 787-793.
- , 1980, Pattern recognition in genetic sequences: *Natl. Acad. Sci. Proc.*, in press.
- SNEATH, P. H. A., 1975, Quantitative method for lateral tracing of sedimentary units: *Computers and Geosciences*, v. 1, p. 215-220.
- , 1976, Clarification on a quantitative stratigraphic correlation technique: *Computers and Geosciences*, v. 1, p. 353-354.
- WATERMAN, M. S.; SMITH, T. F.; and BEYER, W. A., 1976, Some biological sequence metrics: *Adv. Math.*, v. 20, p. 367-387.

TOPOLOGIC ASYMMETRY OF DRAINAGE NETWORKS: THE L INDEX AND ITS APPLICATIONS¹

R. I. FERGUSON

Department of Earth & Environmental Science, University of Stirling, Scotland FK9 4LA

ABSTRACT

Most drainage networks are right- or left-handed to some extent, but extreme or widespread asymmetry may indicate systematic distortion of local or regional drainage patterns by geologic or geomorphic complications. The maximum excess of interior over exterior links as a network is scanned in clockwise maze order provides a simple topologic index of asymmetry, low for right-handed and high for left-handed networks. Its sampling distribution under Shreve's random-topology model is known, allowing assessment of the unusualness of individual asymmetric networks or the significance of regional departures from randomness. Scottish examples of both types of application are given. Glaciation can lead to significant asymmetry.

INTRODUCTION

In a pioneering paper Shreve (1966) demonstrated that Horton's (1945) "laws" of drainage composition can be explained by assuming that drainage networks are topologically random. Many other observed regularities of drainage basin morphology have since

been explained by extensions of Shreve's basic model (see Smart 1973 for a review), and several workers have tested it directly by classifying mapped drainage networks according to topologic criteria and comparing observed with expected frequencies. With few exceptions these empirical tests have confirmed that "the random topology model correctly describes drainage networks in a wide variety of lithological, structural, and climatic regions" (Werritty 1972, p. 192), so much as indeed that Werritty expressed concern at the apparent insensitivity of topologic indices to geological controls.

¹ Manuscript received August 2, 1979; revised March 10, 1980.