The construction of hierarchic and non-hierarchic classifications

By N. Jardine and R. Sibson*

Many of the cluster methods that are used in the construction of classificatory systems operate on data in the form of a dissimilarity coefficient on a set of objects. In this paper we outline a theoretical framework within which the properties of such methods may be discussed. Certain conditions that a cluster method should satisfy are suggested, and a particular sequence of cluster methods which satisfies these conditions is described. The application of the sequence of methods is illustrated by a simple example.

Numerous methods for the derivation of classificatory systems from data in the form of a dissimilarity coefficient on a set of objects have been proposed. Some of these methods are described in: Sokal and Sneath (1963); Williams and Dale (1965); Lance and Williams (1967a, 1967b). Despite the very considerable number of papers describing and applying such methods, there have been relatively few attempts to construct a mathematical framework within which the properties of such methods may be investigated. Exceptions are the papers of: Bonner (1964); Watanabé (1965); Estabrook (1966); Johnson (1967); Jardine, Jardine, and Sibson (1967); and Jardine and Sibson (1968). Likewise, there have been relatively few reports of comparative studies in which a variety of methods were applied to the same data. Exceptions are the papers of: Boyce (1964); Minkoff (1965); Watson, Williams, and Lance (1966); and Sokal and Michener (1967).

If progress is to be made in the understanding of these potentially very useful methods it is important that analytical and empirical investigation of the properties of cluster methods should go hand in hand. Suppose, for example, that we wish to investigate the *stability* of the classification produced by a given cluster method as we increase the quantity of information used in calculating the dissimilarity coefficient on the set of objects to be classified. Before we can interpret the changes produced in the classification we must know whether the transformation of the dissimilarity coefficient by the cluster method is a continuous transformation. Similarly, we may wish to be able to compare the goodness-of-fit to given data of the classifications obtained using a variety of cluster methods. In this case we need some measure of the distortion imposed by a classification on a dissimilarity coefficient, and we must set up the analytical framework required to determine for each cluster method whether it minimises the distortion under certain conditions.

In Jardine, *et al.* (1967) and in Jardine and Sibson (1968) a formal model was established within which the mathematical properties of any cluster method operating on a dissimilarity coefficient may be investigated. Criteria of adequacy for a cluster method were suggested, and

* *King's College, Cambridge, England.*

it was shown that the majority of the cluster methods currently in use fail to satisfy these criteria. A particular sequence of cluster methods which satisfy the criteria was described. In the following sections an informal account of the model is given, and a graph-theoretic description is used to illustrate the sequence of cluster methods and their application. First we consider cluster methods that may be used in obtaining a hierarchic classification, then we examine the more general problem of characterising non-hierarchic classifications, and describing cluster methods that may be used in their construction.

1. Hierarchic classificatory systems

The derivation of a hierarchic classificatory system from a dissimilarity coefficient is a two-stage process. The first stage is the derivation of a *dendrogram*. A dendrogram may be described informally as a hierarchy with numerical levels. The levels at which each pair of objects meet in a dendrogram, the *splitting-levels*, are determined by the dissimilarity coefficient from which the dendrogram is derived. The way in which the splitting-levels are determined depends upon the cluster method used. A hierarchic classification may be derived from a dendrogram by identifying the ordinal levels (ranks) of the hierarchy with numerical levels in the dendrogram. We shall call the sets of objects which are grouped at or below some numerical level in a dendrogram *clusters*. The sets of objects that are grouped at some rank (ordinal level) in a classificatory system we shall call *classes*. The classes of a given rank in a hierarchy consist of just those objects that cluster at or below the corresponding level in the dendrogram. Various rules may be devised for identifying the ranks of a hierarchy with the numerical levels in a dendrogram. For example we might make the identification correspond to some observed clumping of the splitting-levels in the dendrogram. Alternatively we may, as was suggested by Wirth, Estabrook, and Rogers (1966), use the dendrogram, together with suitable measures of the isolation and homogeneity of the clusters in the dendrogram, as a guide to the construction of a hierarchic classification.

It is the first stage, the passage from a dissimilarity coefficient to a dendrogram, that constitutes what is generally called cluster analysis (or a sorting-strategy). A hierarchic dendrogram may be given a numerical characterisation by indicating the splitting-level for each pair of objects (see Fig. 1). In general a hierarchic dendrogram may be characterised as a pair $S = (P, d)$, where P is a set and d is a function from pairs of elements of P to the non-negative real numbers, satisfying the conditions:

(1) $d(a, b) \geqq 0$ for all $a, b \in P$;
(2a) $d(a, b) = 0$ if $a = b$;
(2b) $d(a, b) = 0$ only if $a = b$;
(2c) there exist $a, b \in P$ such that $d(a, b) \neq 0$ if $|P| > 1$;
(3) $d(a, b) = d(b, a)$ for all $a, b \in P$;
(4) $d(a, c) \leq \max \{d(a, b), d(b, c)\}$ for all $a, b, c \in P$.

In other words a hierarchic dendrogram is characterised by an *ultrametric*. A similar characterisation is given by Johnson (1967).

A *dissimilarity coefficient* will, in general, satisfy conditions (1), (2a), (2c) and (3); it will not necessarily satisfy condition (4) (the ultrametric inequality), or condition (2b). It may satisfy some weaker condition analogous to (4) such as the metric inequality:

(5) $d(a, c) \leq d(a, b) + d(b, c)$ for all $a, b, c \in P$.

A cluster method which transforms a dissimilarity coefficient into a hierarchic dendrogram may therefore be regarded as a method whereby the ultrametric inequality is 'imposed' on a dissimilarity coefficient. There are certain simple conditions which we may reasonably require any such transformation to satisfy.

(A) A unique result should be obtained from given data; that is, the transformation should be *well-defined*.

(B) Small changes in the data should produce small changes in the resultant dendrogram; that is, the transformation should be *continuous*.

(C) If the dissimilarity coefficient is already ultrametric it should be unchanged by the transformation.

(D) In some sense the result obtained should impose the minimum distortion upon the dissimilarity coefficient, subject to conditions (A)–(C) and (E)–(G).

A possible family of measures of the distortion imposed by a hierarchic dendrogram is given by:

$\hat{\Delta}_\mu(d, D^*(d)) =$
$[\Sigma|d(a, b) - D^*(d)(a, b)|^{1/\mu}]^\mu/[\Sigma d(a, b)^{1/\mu}]^\mu \ (0 < \mu \leq 1)$

$\hat{\Delta}_0(d, D^*(d)) =$
$\qquad [\max|d(a, b) - D^*(d)(a, b)|]/[\max d(a, b)]$

where summation and maximisation are taken over all pairs $(a, b) \in P \times P$; d is the dissimilarity coefficient; and $D^*(d)$ is the ultrametric characterising the hierarchic dendrogram obtained by some transformation of the dissimilarity coefficient. Some of the $\hat{\Delta}_\mu$ are familiar in form: $\hat{\Delta}_0$ (obtained in the limit as $\mu \to 0$) is a normalised

'maximum modulus'; $\hat{\Delta}_1$ is a normalised 'mean modulus'; and $\hat{\Delta}_{1/2}$ is a normalised 'root mean square'. A similar suggestion was made by Ward (1962).

(E) The operation of the transformation should commute with multiplication of the dissimilarity coefficient by any strictly positive scalar; that is,

$$D^*(kd) = kD^*(d)$$

for any $k > 0$. The effect of this is to ensure that the transformation is independent of scale.

(F) The operation of the transformation should commute with any permutation of P: that is,

$$D^*d[(\rho \times \rho)] = [D^*(d)](\rho \times \rho)$$

where ρ is any permutation of P. The effect of this is to ensure that the transformation is independent of any preliminary labelling of the objects to be classified.

(G) If a cluster is excised and the transformation is applied to the restriction to it of the dissimilarity coefficient, the resultant dendrogram should be the restriction to that cluster of the original dendrogram; that is, (in view of condition (F)) if d has the form, as a matrix
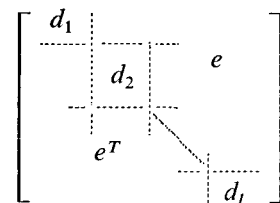
where the minimum entry outside the square diagonal blocks exceeds the maximum entry within them, then $D^*(d)$ has the form
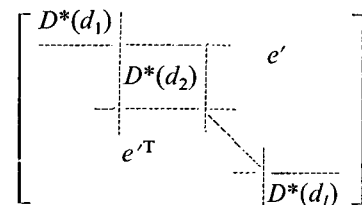
where this matrix satisfies the same condition. This condition guarantees the consistency of the dendrograms obtained when the set P is extended or restricted in suitable ways.

It can be shown that the majority of the cluster methods currently in use fail to satisfy these conditions. For example, of the cluster methods described by Lance and Williams (1967a) all but one fail to satisfy the conditions; the *complete-link* (furthest-neighbour) method originally proposed by Sørensen (1948) fails by being ill-defined; the *centroid* method of Sokal and Michener (1958), the *median* method, the *group-average* method, and the *flexible* method suggested by Lance and Williams (1966b), all fail by being discontinuous. The flexible method may also fail to satisfy condition (C). Most known methods satisfy conditions (E)–(G).

178

The *single-link* (nearest-neighbour) method is the only method that we have seen that satisfies conditions (A)–(G). It can be given a simple graph-theoretic description which makes clear its defects as a method of classification (cf. Wirth *et al.*, 1966). Any dissimilarity coefficient on a set of objects can be characterised by a set of graphs, one for each value taken by the dissimilarity coefficient. The graph for a given value h of a dissimilarity coefficient on a set of objects has as vertices points representing the objects and edges joining just those pairs of points representing pairs of objects with dissimilarity $\leq h$. The single-link method produces clusters at each level that correspond to the components of the graph for that value of the dissimilarity coefficient (see Fig. 1). The defect of the single-link method is that it clusters together at a relatively low level objects linked by chains of intermediates. This defect is generally called *chaining*, but to call chaining a defect of the single-link method is rather misleading; the graph-theoretic description makes it clear that chaining is simply a description of what the method does.

The various average-link and centroid methods attempt to avoid this 'defect' by picking out clusters which are in some sense more homogeneous than those obtained by the single-link method, but in doing so they fall prey to the defect of discontinuity. In Jardine and Sibson (1968) it is suggested that the defects of the single-link method should be regarded as defects of hierarchic classification itself. It is further suggested that the best way to recover information of the kind that is concealed by chaining, for example information about the relative homogeneity of clusters, is to consider cluster methods which lead to overlapping (non-hierarchic) classificatory systems.

2. Non-hierarchic classificatory systems

The fact that there is a considerable body of consistent usage for hierarchic classificatory systems facilitates the construction of an appropriate logical model for such systems, for the terms which must be defined and the theorems which must be proved are known in advance. The situation is not as simple for non-hierarchic classificatory systems, for in this case there is no well-established consistent usage. Some of those who have discussed methods leading to systems of overlapping clusters are: Olson and Miller (1951); McQuitty (1956); Parker-Rhodes and Needham (1960); Needham (1961, 1965, 1967); Jancey (1966); and Lance and Williams (1967b). Intuitively we should expect that as the degree of overlap is allowed to increase so the accuracy of representation of the data should increase, although at the cost of increased complexity. In the limiting case where arbitrary overlap is allowed an exact representation of the original data should be obtained. These intuitions are precisely expressed in the generalised model covering both hierarchic and non-hierarchic classificatory systems given in Jardine and Sibson (1968).

A hierarchic classificatory system may be considered as a nested sequence of partitions of a set of objects. In

Jardine and Sibson (1968) the notion of a partition is generalised by defining a *k-partition*. A *k*-partition allows a maximum of $k - 1$ objects in the overlaps between the classes that belong to it. A classificatory system may therefore be considered as a nested sequence of *k*-partitions; the system will be hierarchic in case $k = 1$ and overlapping in case $k > 1$. By a corresponding generalisation we may define the notion of a *k-dendrogram*. In a *k*-dendrogram clusters at a given level may overlap to the extent of $k - 1$ objects.

The single-link cluster method can be generalised and it can be shown that each of the sequence of cluster methods so defined satisfies conditions (A), (B), and (D)–(F) given on p. 178, and suitable generalisations of conditions (C) and (G). This sequence of methods we denote by (B_k). The first member of the sequence, B_1, is the *single-link* method leading to a hierarchic dendrogram (1-dendrogram). The second member of the sequence, B_2, may be called the *double-link* method, and leads to a dendrogram in which clusters may overlap to the extent of one object (a 2-dendrogram), and so on. If P is the set of objects, and $|P| = p$, where $|P|$ is the number of elements of P, then B_{p-1} gives an exact representation of the dissimilarity coefficient. It can be shown that the family of measures of distortion, $\hat{\Delta}_\mu(d, B_k(d))$, is monotone decreasing with increasing k, becoming zero in case $k = p - 1$. It can be shown also that the use of this sequence of methods enables us to recover information about the homogeneity of clusters that the single-link method fails to reveal.

The sequence of cluster methods (B_k) can be given a simple graph-theoretic description which generalises that given for the single-link method (see above). The clusters at level h in $B_k(d)$, the kth member of the sequence of *k*-dendrograms, are obtained as follows. A graph is drawn whose vertices represent the objects and whose edges join just those pairs of points which represent objects with dissimilarity $\leq h$. The maximal complete subgraphs (maximal subsets of the set of vertices in which all possible edges are present) are marked, and wherever the vertex sets of two such subgraphs intersect in at least k vertices further edges are drawn in to make the union of the two vertex sets into a complete subgraph. The process is repeated until there is no further alteration. If this process is carried out for each of the values taken by the dissimilarity coefficient the graph representation of $B_k(d)$ is obtained. This algorithm is not suitable for computation, but is useful since it illustrates the way in which the sequence of cluster methods (B_k) operates on a dissimilarity coefficient. The application of this algorithm is illustrated in **Fig. 1**. An algorithm suitable for computation is given on p. 180.

The maximal complete subgraphs given by each value of the dissimilarity coefficient can be regarded as the 'nuclei' of the clusters formed at each level in the sequence of *k*-dendrograms $(B_k(d))$. Several authors, notably Bonner (1964) and Needham (1961), have suggested that the recognition of maximal complete subgraphs should be the first stage in a cluster method. The
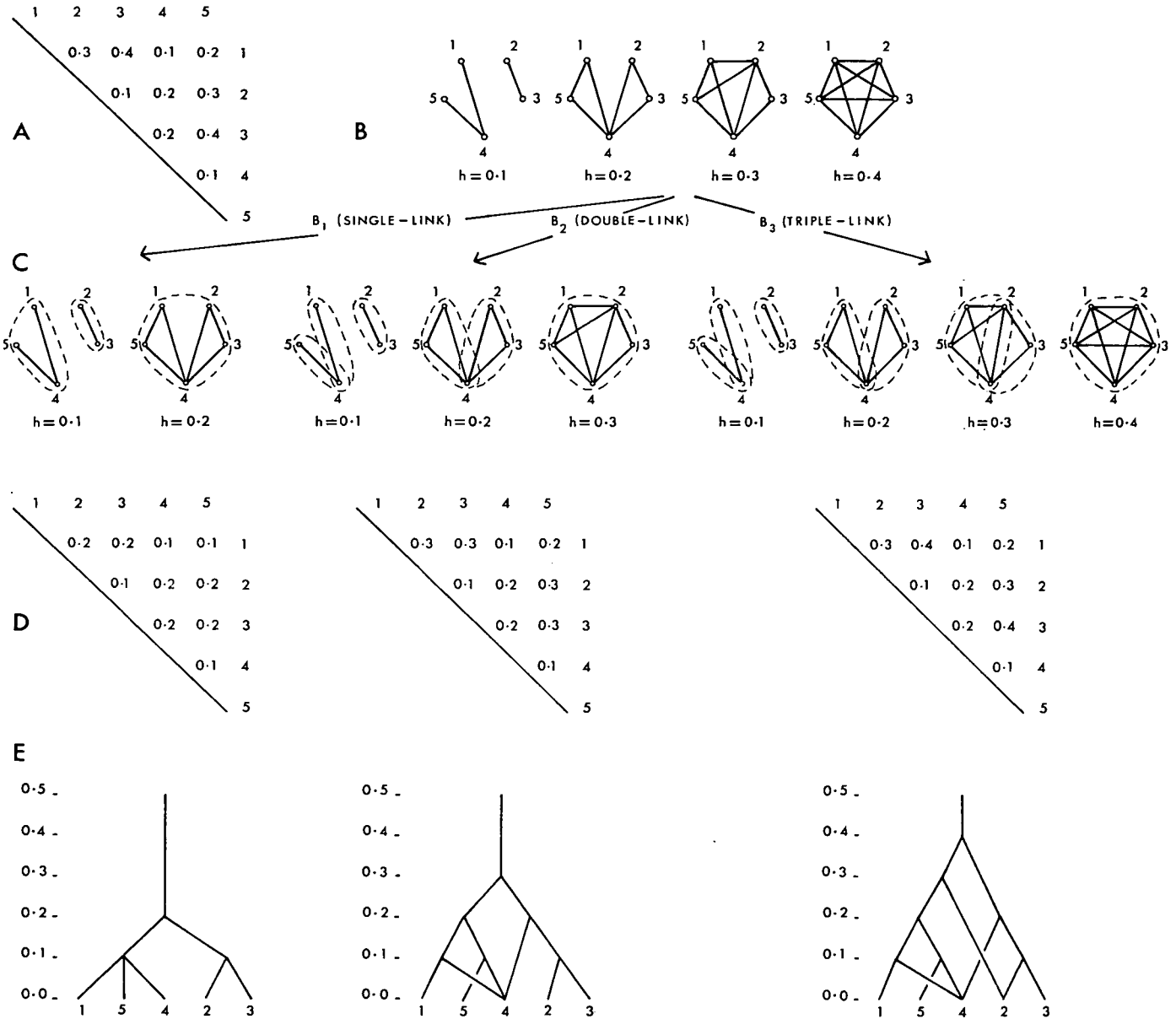
179

Fig. 1. The derivation of $(B_k(d))$, the sequence of k-dendrograms, from a dissimilarity coefficient. **A.** The numerical characterisation of a dissimilarity coefficient on five objects; **B.** A graph representation of the dissimilarity coefficient; **C.** A graph representation of $(B_k(d))$; **D.** A numerical characterisation of $(B_k(d))$; **E.** 'Tree' diagrams representing $(B_k(d))$.

advantages of the particular sequence of methods described here are as follows.

1. Calculation of the distortion imposed on the data by successive members of the sequence $(B_k(d))$ makes it possible to decide how far it is useful to depart from a hierarchic classification. In other words, the distortion measures make it possible to choose a reasonable compromise between complexity and accuracy of representation of the data.

2. Since each member of sequence of methods is a continuous transformation it is possible to investigate the stability of the classifications obtained when further information about the objects is used in computing the

dissimilarity coefficient. If discontinuous cluster methods are used such investigation poses great difficulties.

3. The application of the sequence of cluster methods (B_k)

For a set of objects P, with $|P| = p$, a dissimilarity coefficient can take a maximum of $\frac{1}{2}p(p - 1)$ distinct values. In cases where the number of distinct values approaches this maximum it becomes unfeasible to carry out the sequence of cluster methods (B_k) by hand, using the graph-theoretic algorithm, for more than about 10 objects. The following terminating algorithm is suitable for purposes of computing for up to about 25 objects.

1. List the subsets of P with exactly $k + 2$ elements in an arbitrary order.

2. Consider the value of d taken on each pair in the first subset. If d takes a unique maximum value on a single pair reduce this value to the next largest value taken by d on any pair from the subset. Otherwise leave d unchanged.

3. Repeat the process on the next subset starting with the modified d. Continue until all the subsets have been considered.

4. Repeat 2 and 3 until the list can be run through without further modification of d. The resultant dissimilarity coefficient is $B_k(d)$. This algorithm, together with the calculation of $\hat{\Delta}_{1/2}(d, B_k(d))$ for each member of the sequence, has been programmed for use on the Titan computer at the Cambridge University Mathematical Laboratory by Miss A. A. Houston.

Two practical points should be noted in using this algorithm. Firstly, the computation time increases rapidly with p (approximately as p^{k+2} for $k \ll p/2$). It is probable that a more economical algorithm could be devised. Secondly, certain difficulties arise in cases where condition (2b) on p. 178 is not satisfied by the dissimilarity coefficient: i.e. $d(a, b) = 0$, but $a = b$. Two options are available. One is to treat the objects as distinct and to apply the algorithm directly. Alternatively single-link clustering at level $h = 0$ may be applied to produce preliminary clusters. The dissimilarity between two such clusters is taken to be the minimum value of the dissimilarities between pairs, one member of the pair being from each cluster. The algorithm is then applied with those clusters as objects, and dissimilarities calculated as above. This process corresponds to the preliminary identification of objects not distinguished by the original dissimilarity coefficient. In general the first process leads to smaller distortion by the resultant k-dendrograms; if the dissimilarity coefficient satisfies the metric inequality (condition (5), p. 178) the two options are equivalent.

The representation of the sequence of k-dendrograms $(B_k(d))$ is itself a serious problem. The numerical characterisations are not very informative, and do not readily suggest useful classifications of the objects. The 'tree' representations, whilst useful for $k \le 2$, become increasingly difficult to draw and interpret for larger values of k (see Fig. 1). The most useful way of presenting the information seems to be to indicate the clusters recognised at each level on graph diagrams (see Figs. 1 and 2). The difficulty in this method of presentation lies in finding a two-dimensional arrangement of points representing the objects which minimises the tangling between edges in the graphs for each level. A convenient way of obtaining such a representation is to apply non-metric multidimensional scaling to the dissimilarity coefficient (see Kruskal, 1964a, 1964b. The two-dimensional arrangement obtained by non-metric multidimensional scaling may then be modified to eliminate collinearity of points and to obtain a neater arrangement for purposes of display. The arrangement

I

Table 1

A dissimilarity coefficient (Mahalanobis' generalised distance) on nine populations of *Sagina apetala*

1	2	3	4	5	6	7	8	9	
	9·28	6·53	9·73	1·79	6·41	6·98	13·04	3·24	1
		4·65	5·21	10·21	7·30	3·66	4·72	10·09	2
			6·01	7·55	2·98	2·63	9·30	6·21	3
				11·20	7·73	8·37	4·48	10·15	4
					4·62	6·90	12·45	4·68	5
						3·68	11·21	3·80	6
							8·38	7·13	7
								12·75	8
									9

Table 2

The numerical characterisation of the first four members of the sequence of k-dendrograms, $(B_k(d))$

1	2	3	4	5	6	7	8	9	
	3·80	3·80	4·72	1·79	3·80	3·80	4·72	3·24	1
		3·66	4·72	3·80	3·66	3·66	4·72	3·80	2
			4·72	3·80	2·98	2·63	4·72	3·80	3
				4·72	4·72	4·72	4·48	4·72	4
					3·80	3·80	4·72	3·24	5
						2·98	4·72	3·80	6
							4·72	3·80	7
								4·72	8
									9

$k = 1$

$\hat{\Delta}_{1/2}(d, B_1(d)) = 0.528$

1	2	3	4	5	6	7	8	9
	6·21	6·21	6·21	1·79	4·68	6·21	6·21	3·24	1
		4·65	5·21	6·21	4·65	3·66	4·72	6·21	2
			6·01	6·21	2·98	2·63	6·01	6·21	3
				6·21	6·01	6·01	4·48	6·21	4
					4·62	6·21	6·21	4·68	5
						3·68	6·01	3·80	6
							6·01	6·21	7
								6·21	8
									9

$k = 2$

$\hat{\Delta}_{1/2}(d, B_2(d)) = 0.367$

1	2	3	4	5	6	7	8	9
	7·30	6·53	7·73	1·79	6·41	6·90	8·38	3·24	1
		4·65	5·21	7·30	7·30	3·66	4·72	7·30	2
			6·01	6·53	2·98	2·63	8·38	6·21	3
				7·73	7·73	7·73	4·48	7·73	4
					4·62	6·90	8·38	4·68	5
						3·68	8·38	3·80	6
							8·38	6·90	7
								8·38	8
									9

$k = 3$

$\hat{\Delta}_{1/2}(d, B_3(d)) = 0.227$

1	2	3	4	5	6	7	8	9
	9·28	6·53	9·28	1·79	6·41	6·98	9·30	3·24	1
		4·65	5·21	9·28	7·30	3·66	4·72	9·28	2
			6·01	7·13	2·98	2·63	9·30	6·21	3
				9·28	7·73	8·37	4·48	9·28	4
					4·62	6·90	9·30	4·68	5
						3·68	9·30	3·80	6
							8·38	7·13	7
								9·30	8
									9

$k = 4$

$\hat{\Delta}_{1/2}(d, B_4(d)) = 0.146$

$h = 5.50$
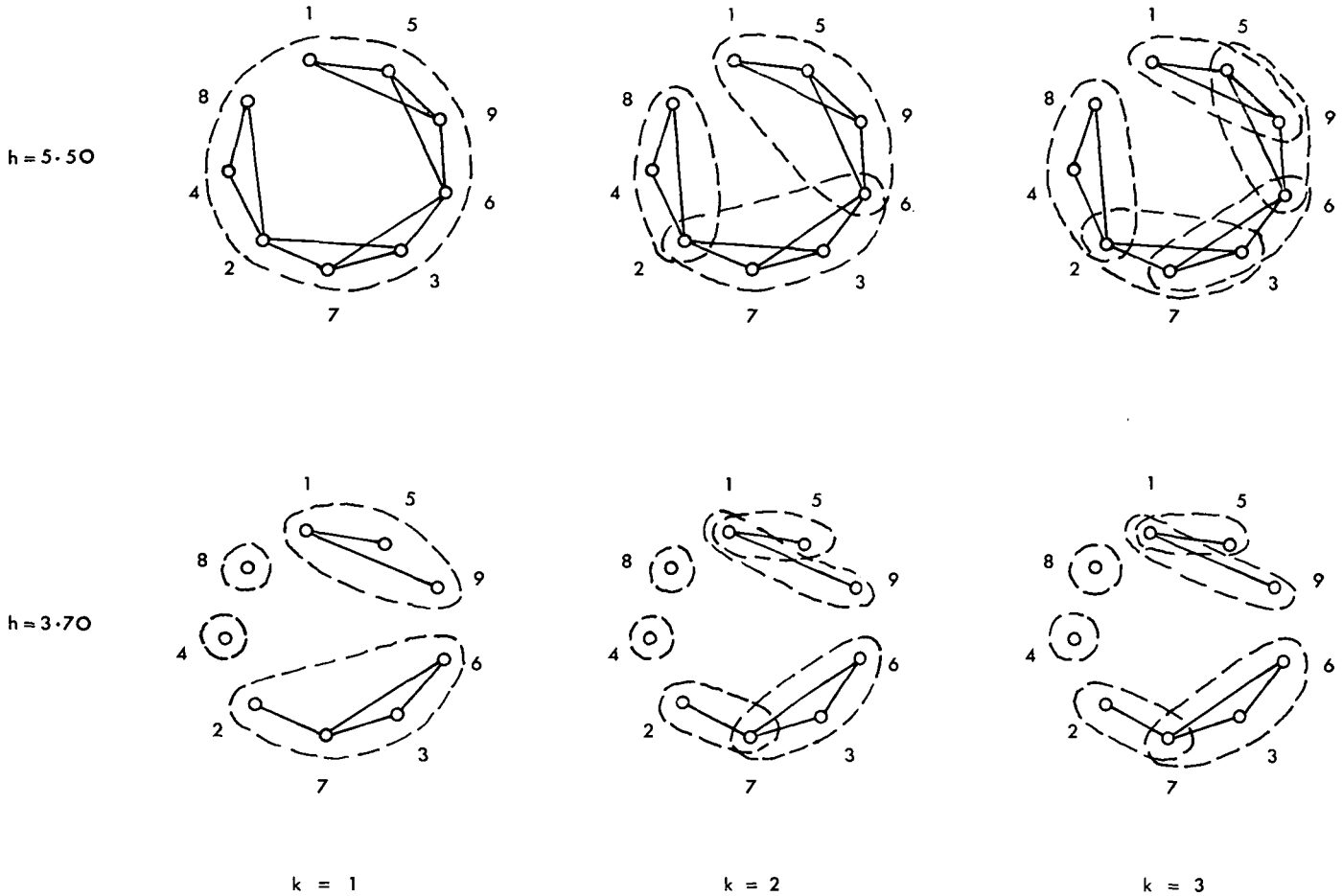
$h = 3.70$

$k = 1$ $k = 2$ $k = 3$

Fig. 2. The graph representations of $(B_k(d))$ for $k = 1, 2$ and 3 at levels $h = 3.70$ and 5.50.

of points in Fig. 2 was produced by modification of the two-dimensional arrangement obtained using a program for non-metric multidimensional scaling written by J. B. Kruskal.

In deciding on the level or levels at which clusters obtained in a k-dendrogram should be recognised as classes in a classification of the set of objects, it is helpful to devise measures of the relative *isolation*, and of the relative *internal cohesion* (homogeneity) of the clusters. One possible measure of isolation is given by $(h_1 - h_0)$, where $h_1 = \min \{d(a, b): a \in P-S, b \in S\}$, and h_0 is the lowest level at which the cluster S appears in the k-dendrogram; this measure generalises a measure suggested by Wirth *et al.* (1966). An appropriate measure of the internal cohesion of a cluster in a k-dendrogram at level h is given by

$$[\Sigma_+(d(a, b) - h)]/\tfrac{1}{2}s(s - 1)$$

where $s = |S|$, $d(a, b)$ is the value of the dissimilarity coefficient on the pair of objects (a, b) from S, and Σ_+ is summation over positive terms.

4. An illustrative example

In **Table 1** a dissimilarity coefficient on nine populations of the plant *Sagina apetala* is shown. The coefficient is the Generalised Distance computed from the values taken by twenty parameters on thirty individuals in each population; pooled variances and covariances were used (see Mahalanobis, 1936; Rao, 1952). The aim of the study was to find out if the data supports the recognition of three geographical subspecies within *Sagina apetala*.

In **Table 2** the numerical characterisations of the k-dendrograms obtained are shown for $k = 1 \ldots 4$. In each case $\hat{\Delta}_{1/2}(d, B_k(d))$ is given. In **Fig. 2** the clusters obtained for $k = 1 \ldots 3$ at levels $h = 3.70$ and 5.50 are shown.

In the hierarchic dendrogram the distortion is very high and the clusters obtained shown low homogeneity. In $B_2(d)$ there is a marked drop in distortion and at $h = 5.50$ three clusters having relatively high homogeneity are obtained. Subsequent members of the sequence reveal no useful clusters. The clusters obtained

in $B_2(d)$ at $h = 5 \cdot 50$ assign the populations correctly to the subspecies recognised on independent grounds in Clapham and Jardine (1964). Populations 2 and 6 which occur in the overlaps between the clusters are intermediates which occur where the geographical ranges of two of the subspecies overlap with that of the third. In this case, at least, the use of the sequence of cluster methods (B_k) is vindicated.

5. Conclusions

The theoretical model outlined in this paper is one in which classification is regarded as a two-stage process. The first stage is the derivation of a dissimilarity coefficient. The second stage is the transformation of a dissimilarity coefficient into a classificatory system, and it is this stage that we have considered in detail. The derivation of dissimilarity coefficients from discrete-state (attribute) data, quantitative data, and mixed data is the subject of a forthcoming paper by Jardine and Sibson. An apparently very different kind of theoretical model was constructed by Watanabĕ (1965); his information-theoretic model covers a variety of methods that go directly from attribute data to a hierarchic classificatory system (for example the methods of Alexander and Mannheim, 1962; Alexander, 1963; Macnaughton-Smith, Williams, Dale, and Mockett, 1964; Lance and Williams, 1966a). The relation between such methods and methods of the kind that are covered by our model needs further investigation.

One obvious omission in our discussion of the sequence of methods described is the absence of any tests for the significance of the clusters obtained. A promising approach to devising such tests is to compare the clusters obtained by the sequence of cluster methods (B_k) on a given dissimiliarity coefficient, with those obtained by application of (B_k) to dissimilarity coefficients generated from some kind of random data.

We do not claim that the sequence of methods described here constitutes any kind of unique solution to the problems of cluster analysis. We do, however, suggest that the construction of mathematical models of the kind outlined here is essential if cluster methods are to be used in scientific investigations; for very little can be done to establish the empirical structure of data until the mathematical properties of the methods of analysis used are known.

References

ALEXANDER, C. (1963). HIDECS 3: four computer programs for the hierarchical decomposition of systems which have an associated linear graph, *M.I.T. Civil Engineering Systems Lab. Res. Rept.* R 63–28, Cambridge, Mass.

ALEXANDER, C., and MANNHEIM, M. (1962). HIDECS 2: a computer program for the hierarchical decomposition of a set with an associated graph, *M.I.T. Civil Engineering Systems Publ. No.* 160, Cambridge, Mass.

BONNER, R. E. (1964). On some clustering techniques, *IBM Journ. Res. Devpt.*, Vol. 8, p. 22.

BOYCE, A. J. (1964). The value of some methods of numerical taxonomy with reference to hominoid classification, in: *Phenetic and phylogenetic classification* (eds. Heywood and McNeill), Syst. Assoc. Publ. No. 6, p. 47.

CLAPHAM, A. R., and JARDINE, N. (1964). '*Sagina*' in: *Flora Europaea*, Vol. 1 (eds. Tutin *et al.*), Cambridge University Press, p. 148.

ESTABROOK, G. F. (1966). A mathematical model in graph theory for systematic biology, *Journ. Theor. Biol.*, Vol. 12, p. 297.

JANCEY, R. C. (1966). Multidimensional group analysis, *Aust. J. Bot.*, Vol. 14, p. 127.

JARDINE, C. J., JARDINE, N., and SIBSON, R. (1967). The structure and construction of taxonomic hierarchies, *Math. Biosc.*, Vol. 1, p. 173.

JARDINE, N., and SIBSON, R. (1968). A model for taxonomy, *Math. Biosc.* (in press).

JARDINE, N., and SIBSON, R. The measurement of dissimilarity, (forthcoming paper).

JOHNSON, S. C. (1967). Hierarchical clustering schemes, *Psychometrika*, Vol. 32, p. 241.

KRUSKAL, J. B. (1964a). Multidimensional scaling by optimising goodness-of-fit to a nonmetric hypothesis, *Psychometrika*, Vol. 29, p. 1.

KRUSKAL, J. B. (1964b). Nonmetric multidimensional scaling: a numerical method, *Psychometrika*, Vol. 29, p. 115.

LANCE, G. N., and WILLIAMS, W. T. (1966a). Computer programs for monothetic classification ('association analysis'), *Computer Journal*, Vol. 8, p. 246.

LANCE, G. N., and WILLIAMS, W. T. (1966b). A generalised sorting strategy for computer classifications, *Nature*, Vol. 212, p. 218.

LANCE, G. N., and WILLIAMS, W. T. (1967a). A general theory of classificatory sorting strategies. I. Hierarchical systems, *Computer Journal*, Vol. 9, p. 373.

LANCE, G. N., and WILLIAMS, W. T. (1967b). A general theory of classificatory sorting strategies. II. Clustering systems, *Computer Journal*, Vol. 10, p. 271.

MACNAUGHTON-SMITH, P., WILLIAMS, W. T., DALE, M. B., and MOCKETT, L. G. (1964). Dissimilarity analysis: a new technique of hierarchical subdivision, *Nature*, Vol. 202, p. 1034.

McQUITTY, L. L. (1956). 'Agreement analysis', classifying persons by predominant pattern of response, *Brit. Journ. Stat. Psychol.*, Vol. 9, p. 5.

MAHALANOBIS, P. C. (1936). On the generalised distance in statistics, *Proc. Nat. Inst. Sci. India*, Vol. 2, p. 49.

Minkoff, E. C. (1965). The effects on classification of slight alterations in numerical technique, *Syst. Zool.*, Vol. 14, p. 196.

Needham, R. M. (1961). The theory of clumps II, Cambridge Language Research Unit, M.L. 139, (mimeo).

Needham, R. M. (1965). Application of the theory of clumps, *Mechanical Translation*, Vol. 8, p. 13.

Needham, R. M. (1967). Automatic classification in linguistics, *Statistician*, Vol. 17, p. 45.

Olson, E. L., and Miller, R. L. (1951). A mathematical model applied to a study of the evolution of species, *Evolution*, Vol. 5, p. 256.

Parker-Rhodes, A. J., and Needham, R. M. (1960). The theory of clumps, Cambridge Language Research Unit, M.L. 126, (mimeo).

Rao, C. R. (1952). *Advanced statistical methods in biometric research*, John Wiley, New York.

Sokal, R. R., and Michener, C. D. (1958). A statistical method for evaluating systematic relationships, *Univ. Kansas Sci. Bull.*, Vol. 38, p. 1409.

Sokal, R. R., and Michener, C. D. (1967). The effects of different numerical techniques on the phenetic classification of bees of the *Hoplitis* complex (Megachilidae), *Proc. Linn. Soc. Lond.*, Vol. 178, p. 59.

Sokal, R. R., and Sneath, P. H. A. (1963). *Principles of numerical taxonomy*, W. H. Freeman & Co., San Francisco and London.

Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology . . ., *Vidensk. Selsk. Biol. Skr.*, Vol. 5, No. 4, p. 1.

Ward, J. H. (1963). Hierarchic grouping to optimise an objective function, *Journ. Amer. Stat. Assoc.*, Vol. 58, p. 236.

Watanabé, S. (1965). Une explication mathématique du classement des objets, in: *Information and prediction in science* (Eds. Dockx and Bernays), Academic Press, New York and London.

Watson, L., Williams, W. T., and Lance, G. N. (1966). Angiosperm taxonomy: a comparative study of some novel numerical techniques, *Journ. Linn. Soc. (Bot.)*, Vol. 59, p. 491.

Williams, W. T., and Dale, M. B. (1965). Fundamental problems in numerical taxonomy, *Adv. Bot. Res.*, Vol. 2, p. 35.

Wirth, M., Estabrook, G. F., and Rogers, D. J. (1966). A graph theory model for systematic biology, with an example for the Oncidiinae (Orchidaceae), *Syst. Zool.*, Vol. 15, p. 39.

Book Review

Process Control Systems, by F. G. Shinskey, 1967; 367 pages. (*McGraw-Hill Book Company*, £5 12s.)

The gap between theory and practice in Process Control continues to grow. Although elegant techniques exist for the solution of many control problems the pay-off, in terms of industrial applications, so far reflects scant reward for the time and effort which have been spent to date on the application of modern control theory in the process industries. Although it has been suggested that the mathematics of modern control theory is a hindrance to its use by practising engineers I believe that very often the theories themselves do not meet the needs of real process control problems. Before much headway can be made towards the analytical design of more advanced process control systems a far greater understanding is required, among the control theorists, of the nature of physical and chemical processes, their dynamics and the real problems associated with the running and control of these processes.

A new look at process control is needed which combines that which is worth saving from the largely empirical control techniques which are currently employed in the industry with the advantages of a methodical approach to analysis and design which are claimed for modern control theory.

This book by Shinskey obviously sets out to provide such an approach. In particular it contains much useful classification of process characteristics and of the systems which are known to give satisfactory control characteristics. In this respect the book is admirable. However, the way in which control system design is approached is too sketchy and relies too much on the acceptance of results which have been obtained by others either for different plants or from mathematical generalisations of the problems. I believe the dangers of this kind of approach are severe, that each system must be treated on its merits and, therefore, that this textbook would be better if greater attention were paid to the methods by which results may be derived.

The book is to be completely recommended to control theorists as offering some insight into the nature of practical control problems. As a guide for process engineers into the methods of control system design it is less good but is still a vast improvement on most texts of its kind.

Difficult aspects of control such as nonlinearity, interaction, adaptation, optimisation are mentioned but only briefly and, while not providing all the tools necessary to deal with these problems, the author does at least survey the inherent advantages and disadvantages of particular control schemes.

J. M. Nightingale (Southampton)