NASA CR-175,286

PAR TECHNOLOGY CORPORATION

NASA-CR-175286
19850013734

Final Report for
Contract No. NAS5-28117

"Technical Support for Creating an Artificial Intelligence System for Feature Extraction and Experimental Design"

Final Report for
Contract No. NAS5-28117

"Technical Support for Creating an Artificial Intelligence System for Feature Extraction and Experimental Design"

Integrating Spatial Information into Clustering Analysis

By: Barry J. Glick

Final Report for
Contract No. NAS5-28117

"Technical Support for Creating an Artificial Intelligence System for Feature Extraction and Experimental Design"

Submitted to:

National Aeronautics & Space Administration
Goddard Space Flight Center
Technical Officer: Edward J. Masuoka, Code 922

Submitted by:

PAR Technology Corporation
Geographic Systems Section
McLean, Virginia 22102

Report No. 85-24

1. Introduction

Techniques for classifying objects into groups or classes go under many different names including, most commonly, cluster analysis. Mathematically, the general problem is to find a "best" mapping of objects into an index set consisting of class identifiers. When an a priori grouping of objects exists, the process of deriving the classification rules from samples of classified objects is known as "discrimination". When such rules are applied to objects of unknown class, the process is denoted "classification."

For this paper, our problem is to classify into groups a set of objects that are each associated with a series of measurements (ratio, interval, ordinal, or nominal levels of measurement). Each measurement produces one variable in a multidimensional variable space. Thus, objects may be represented as vectors or points in this multidimensional space and the usual multivariate statistical techniques may be used. In some applications each object also may exist in geographical space; i.e., each object is associated with a location on the earth's surface. Although an object's location in geographical space can be represented by a pair of planar or spherical coordinates (and, possibly, by a third coordinate representing height or elevation), problems exist in simply considering location as another measurement. These will be discussed below.

A basic methodological philosophy in classification is to consider the distances between objects in the multi-dimensional measurement space. It is expected that similar objects will be represented by points that lie near to one

another in this space. In this sense, clustering can be considered the process of defining regions of the measurement space that divide the points (and their associated objects) into optimal classes. New points (and objects) can then be classified by determining which region they lie in.

2. Cluster Analysis

In cluster analysis the objective is to take a set of objects with unknown classification and to group these objects into "natural" classes or clusters (Hand, 1981). The selection of measurements to use in the cluster analysis is of critical importance because the groupings that result are completely determined by the choice of measurements. If these measurements are irrelevant to the objective or application of the grouping (e.g., trying to identify groups of locations with similar remote sensor characteristics), the clustering is likely to produce irrelevant groupings.

Once the measurements have been selected, it may be desirable to reduce their number to make computation feasible and/or to eliminate variables that will not add significantly to the analysis. In order to do this a measure of how closely the reduced set of measurements or dimensions corresponds to the original set is needed along with an algorithm to find the subset of variables that optimizes this measure. The most popular approach is the method of principle components which is based on linear transformations of variables and the deletion of variables that account for very little of the total variance. More recently, non-linear methods have been proposed and are based on a wide

range of structural criteria (for example, using multidimensional scaling techniques).

A wide variety of cluster analysis techniques exist; these can be conveniently divided into two major approaches, hierarchical and optimization. In a hierarchical analysis, the final groupings are formed by iteratively grouping subclusters or by iteratively splitting parent clusters (i.e., agglomerative or divisive approaches). Optimization techniques attempt to find the clusters that result in the maximization or minimization of a clustering measurement criterion. A major difference between the two approaches is that in optimization objects can be switched between clusters if that results in an improvement in the value of the optimization criteria.

An aspect of cluster analysis that can have a significant input on the results is the choice of the multidimensional distance (i.e., similarity or nearness) measures to use. Many distance measures exist in addition to the usual Euclidean distance, some of which are generalized metrics that include the Euclidean measure as a special case (e.g., Minkowski metrics, Mahalanobis metrics). Table 1 provides a listing of selected distance measures used in clustering (Cormack, 1971).

Major categories of distance measures include those that satisfy the metric properties and those based on the correlation coefficient. For non-interval variables, other distance measures have been developed. For binary variables, similarity measures based on the 2 x 2 agreement/disagreement table are commonly

used, such as the simple matching coefficient (measure 11 in Table 1). Other measures based on this table include the Jaccard coefficient (measure 10 in Table 1) and the Dice coefficient (measure 9 in Table 1). Note that the principles behind the measures for binary variables can easily be extended to the case of nominal variables with n categories. Ordinal variables can be treated as nominal by ignoring the order information, or numerical ranks could be assigned to the orders and the ordinal variable treated as interval. Also, rank-order correlation coefficients could be used in the same way as interval correlation coefficients to define distance measures.

Another issue related to cluster analysis is variable standardization and weighting. Different sets of weights applied to the variables can lead to completely different cluster analysis results (Hand, 1981). However, this fact can be used advantageously as a means to add known or exogenous similarity or importance information to the analysis.

Hierarchical cluster analysis procedures, particularly of the agglomerative type, are well-known. At each step in the procedure, two (or more) existing sub-clusters are merged. In the final step, all objects are grouped into a single cluster. The clustering solution desired may then either be the one produced for the specified number of clusters desired, or when an a priori number of clusters is not known, an optimal clustering level can be selected using various measures of information loss or cluster compactness to select a "break point". One issue in agglomerative cluster analysis is the measurement of inter-cluster distances which are needed to decide, at each step, which

TABLE 1: SELECTED DISTANCE MEASURES USED IN CLUSTERING

(from Haggett et al, 1977)

1. Euclidean distance $\sum_{v=1}^{N} w_v(x_{iv} - x_{jv})^2$

   Unstandardized: $w_v = 1$
   Standardized by standard deviation, s: $w_v = 1/s_v^2$. Denote by Δ^2
   Standardized by range: $w_v = 1/\max_{ij}(x_{iv} - x_{jv})^2$

2. City-block metric $\sum_{v=1}^{N} w_v|x_{iv} - x_{jv}|$

   Mean character difference: $w_v = 1/N$

3. Minkowski metrics $\left[\sum_{v=1}^{N} |x_{iv} - x_{jv}|^{1/\lambda}\right]^{\lambda}$

4. Angular separation $\dfrac{\sum_{v=1}^{N} x_{iv}x_{jv}}{\left[\sum_{v=1}^{N} x_{iv}^2 \sum_{v=1}^{N} x_{jv}^2\right]^{\frac{1}{2}}}$

5. Correlation $p_{ij} = \dfrac{\sum_{v=1}^{N} (x_{iv} - \bar{x}_i)(x_{jv} - \bar{x}_j)}{\left[\sum_{v=1}^{N} (x_{iv} - \bar{x}_i)^2 \sum_{v=1}^{N} (x_{jv} - \bar{x}_j)^2\right]^{\frac{1}{2}}}$

6. Profile similarity index: $\dfrac{2k_m - \Delta^2}{2k_m + \Delta^2}$, *where*

   $P(\chi_p^2 < k_m) = 0.5$

7. Coefficient of nearness: $\{\sqrt{(2N)} - \Delta\}/\{\sqrt{(2N)} + \Delta\}$

8. 'Canberra' metric: $\sum_{v=1}^{N} |x_{iv} - x_{jv}|/(x_{iv} + x_{jv})$

9. $\dfrac{2a}{2a+b+c}$

10. $\dfrac{a}{a+b+c}$

11. Simple matching: $\dfrac{a+d}{a+b+c+d}$

---

* In this table, x_i is the value of the variate X for observation unit i, while $v = 1, 2, \ldots, N$ indexes the number of dimensions in the space. The last three indices relate to binary characters, where a, b, c, and d refer to the number of characters possessed or not possessed in a 2×2 table $\begin{array}{c|c} a & b \\ \hline c & d \end{array}$. Source: Cormack, 1971, p. 325.

sub-clusters to combine. Among the approaches possible are nearest-neighbor, furthest-neighbor, centroids, medians, group averages, or sum of squared deviations distance measures. This latter measure is used in Ward's (1963) popular method.

Optimization approaches involve the definition of the optimization algorithm. The most popular optimization criteria are those based on the matrix identity (Anderberg, 1973):

$$T = W + B$$

T is the scatter matrix which describes the overall deviation of the observations around the grand mean, W is the within-class scatter (i.e., deviation of observations around the cluster means), and B is the weighted sum describing the scatter of the cluster means about the grand mean. Given this identity, the goal of the clustering can be to maximize B or minimize W. However, in order to make the optimization meaningful, it is necessary to summarize the multivariate matrix structure.

This summerization can be performed in several ways. For example, the trace W can be used. This turns out to be identical to the sum of squared deviations from the observations to the cluster means (as in Ward's hierarchical method). However, this method is sensitive to outliers and may not result in compact clusters. In addition, trace W is not invariant to sealing or weighting (this may be an advantage if weights are to be used). Other approaches include

using the determinant of W, using the eigenvalues of the matrix W ' B (which are equal to the ratio of between-cluster scatter to within-cluster scatter), using the trace of matrix W ' B or the trace of matrix T ' W.

After an optimization criterion (and summarization method) has been chosen, the optimum clustering must be determined. The most obvious way to do this is to calculate the criterion value for every possible arrangement of clusters and to select the one with the best score. Unfortunately, the number of possible arrangements quickly becomes prohibitively large. For example, there are 10^{30} possible allocations of 100 objects into 2 classes. Therefore, either the search must be limited to some "likely" subset of arrangements or a more efficient method for complete search must be used.

One approach to limiting the scope of the problem is through the use of evolutionary search procedures. These begin with an initial arbitrary clustering and determine whether or not to switch observations to another group. If this switch will produce an improved score on the optimization criterion, it is implemented. This approach has the potential to result in a solution which is a local, non-global optimum (MacQueen, 1967). Alternatively, a steepest descent algorithm can be used following appropriate transformations of the optimization problem (Gordon and Henderson, 1977).

Branch and bound techniques have also been used to determine optimal clusters (Koontz et al, 1975). This method permits the consideration of every possible clustering arrangement without requiring the explicit evaluation of the

optimization score for each clustering (Hand, 1981). Other mathematical programming techniques have been suggested for cluster analysis, usually applied to special case problems. Rao (1971) has outlined the use of linear and non-linear integer programming techniques for several constrained clustering analyses.

3. Clustering with Spatial Constraints

When the objects to be classified are associated with geographic location, the application may require that spatial properties be explicitly considered. It is, of course, possible to modify clustering procedures so that both nearness in taxonomic space and geographic space are taken into account (Haggett et al, 1977). The simplest approach to implementing this is by the inclusion of spatial contiguity constraints into a standard clustering algorithm.

For some applications, the contiguity constraint is applied absolutely, i.e., clusters can only consist of neighboring objects or areal units. In this case, the distance function used to place an object in the multidimensional space can be defined to be an arbitrarily large number if the two objects are not neighbors; if they are neighbors, distances can be calculated in the variety of ways listed in Table 1.

The concept of variable weighting was introduced in Section 2 above. A second way to integrate spatial information into a clustering analysis is to apply weights to object pairs; these weights are related to the objects'

relative locations. For example, weights may be based on distances between objects, on the length of common boundary, or on the presence/absence of transportation/communication facilities joining the two locations.

Another approach to considering spatial contiguity in clustering relies on constraints applied during the cluster-building process. For example, in determining which object or sub-cluster to add to another in an agglomerative hierarchical clustering algorithm, the search can be restricted to contiguous sub-clusters. Thus, at each step, the goal is to find the sub-cluster such that, when merged with an adjacent sub-cluster, the overall clustering solution is improved to the greatest extent possible.

4. Measures of Spatial Patterning

The discussion of Section 3 describes how information on geographic location can be included in standard cluster analysis techniques. In this approach, geographic location (or contiguity) can be considered a characteristic of an object to be classified in much the same way as any other measured attribute of that object. In some clustering application, it may be beneficial to consider other, more complex indicators of the spatial characteristics of the areal units (i.e., objects) and their associated attributes. This is especially relevant when the object is to group the variables rather than the objects themselves. In this application we are given scores on a set of variables over a set of areal units and the problem is to determine those variables most alike in terms of their spatial distribution or patterning.

Standard correlation procedures are an obvious choice for evaluating the similarity of variables across the observation units. However, this is essentially an aspatial approach in that only areal unit to areal unit comparisons are made. To illustrate the limitations of this approach, consider two variables measured over a gridded sample area such that the value of variables A in a given cell is a function of the value of variable B in a neighboring cell. If the spatial distribution of the values of variable B is random, there will be zero correlation between variables A and B under this scenario. Clearly, we may desire a measure of similarity that could detect such "spill-over" or neighborhood effects.

One approach to doing this has been developed by analogy with time series analysis. A spatial cross-correlation coefficient can be defined as the average correlation between areal units' values on variable A and neighboring areal units' values of variable B. A coefficient with score not significantly different than zero is interpreted to mean that there are no significant "spill-over" effects between variables A and B. The averaging process causes significant loss of potentially useful information concerning directionality of effects (if any do exist). Therefore, it is possible to calculate separate spatial corss-correlation coefficients for neighbors to the east, west, north, and south of the index areal unit.

A single spatial cross-correlation coefficient includes only nearest-neighbor or contiguous units effects. The concept can be generalized to

include spatial "lags" that consider the possibility of effects of second-order neighbors (i.e., units with a single intervening unit between them), third-order, etc. Thus, in its most general form, the elements of a spatial cross-correlation matrix are the spatial cross-correlation coefficients for neighbors i units apart in the x direction and j units apart in the y direction.

For a single variable, spatial autocorrelation (coefficient and/or function) provides a descriptive measure of the overall spatial patterning of the variable. Autocorrelation functions could be calculated for each of the variables involved in a clustering problem. Comparison of these functions may assist in the elimination of variables that do not significantly differ from a remaining variable in terms of spatial pattern. In addition, clusters or groups may be created so as to maximize the degree of spatial autocorrelation in a particular pattern. An alternative approach to the same goal involves the use of trend surface modeling. In this technique, polynomial functions of x and y coordinates are used to decompose a univariate spatial pattern into linear, quadratic, cubic, etc. terms. The coefficients calculated for each of these terms can then be used to compare and classify variables according to their spatial patterns.

4.1 Spatial autocorrelation (Cliff & Ord, 1981)

Consider a study area which has been exhaustively partitioned into n nonoverlapping subareas. Suppose that a random variable, X, has been measured in each of the subareas, and that the value of X in the typical subarea, i, is

x_i . X could describe either (1) a single population from which repeated drawings are made to give the X_i; or (2) n separate populations, one for each county; or (3) a partition of a finite population among the n counties. It is important to note that the choice of population model does not affect the derivation of the measures of spatial autocorrelation, nor the method of analysis. However it does affect the inferences that can be made.

A basic property of spatially located data is that the set of values, $\{x_i\}$, are likely to be related over space. If the $\{x_i\}$ display interdependence over space, we say that the data are spatially autocorrelated. The following formal definition may be made: If, for every pair of counties i and j in the study area the drawings which yield x_i and x_i are uncorrelated, then we say that there is no spatial autocorrelation in the county system on X.

One model of the spatial interpendence among the $\{x_i\}$ is the scheme

$$X_i = p \sum_j w_{ij} X_j + e_i , \qquad i = 1,2,\ldots,n. \tag{4.1}$$

Here, the $\{e_i\}$ are independent and identically distributed variates with common variance, σ^2. The set of weights, $\{w_{ij}\}$, are any set of constants that specify which j subareas in the study area have variate values directly spatially related with X_i. The constant, p, is a measure of the overall level of spatial autocorrelation among the $\{X_iX_j\}$ pairs of which $w_{ij} > 0$. For example, we might put $w_{ij} = 1$ (unscaled) if j is physically continuous to i, and $w_{ij} = 0$ otherwise. More general sets of weights may, however, be constructed. For

example,

$$w_{ij} = (c + d_{ij})^{-a} \qquad (4.2)$$

where d_{ij} is the distance between points or areas, i and j, and a is a 'friction of distance' parameter as used in many gravity and interaction models, and c is a constant ($c > 0$). Finally, when $p > 0$ in model (4.1), we say that there is positive spatial autocorrelation among the $\{X_i\}$ whereas $p < 0$ implies negative spatial autocorrelation. The former case is characterized by similar $\{x_i\}$ values in areas with nonzero $\{w_{ij}\}$ values, and the latter by very different relationships. If $p = 0$ in model (4.1), there is said to be no spatial autocorrelation in the study area on X, and the variate values are randomly mixed.

4.2 Basic spatial autocorrelation measures

The measures of spatial autocorrelation which have been proposed in the literature are discussed according to the kind of data (nominal, ordinal, or interval scaled) to which they may be applied. This also coincides with the historical order of development of the measures.

Measures for nominal data

The simplest nominal scale is a binary classification. In each of the n counties we note whether a given event has or has not occurred. If it has, the

county is color coded black (B), and if it has not, the county is color coded white (W). If two counties have a boundary of positive nonzero length in common, they are said to be linked by a join. A join may link two B counties, two W counties, or a B and a W county. These joins are called BB, WW, and BW joins respectively. To determine whether events in neighbouring counties are spatially autocorrelated or not, we count the numbers of BB, BW, and WW joins which occur in the county system, and compare these numbers with the expected numbers of BB, BW, and WW joins under the null hypothesis, H_0, of no spatial autocorrelation among the counties. Intuitively it can be appreciated that "many" of BB joins, compared with the expected number under H_0, implies clustering of the B counties in the plane, whereas a "many" BW joins implies an alternating pattern of B and W counties as, for example, along the rows and columns of a chessboard.

The usual method employed to determine whether BB, BW, and WW depart significantly from random expectation is to use the fact that these join-count statistics are asymptotically normally distributed and to assume that these results hold approximately for moderate sized lattices. The first two moments of the coefficients are then used to specify the location (μ) and scale (σ^2) parameters of the normal distribution. The early work on these measures was carried out for rectangular lattices. The moments of the join counts were first obtained by Moran (1948).

Quite commonly the nominal scale will have classes ($k > 2$) rather than the simple binary classification discussed above. Each class may then be assigned

one of k distinct colors, and each country is called after the color of the class into which if falls. Conventionally, the analysis then proceeds by counting the number of joins between counties of (1) the same color, (2) two different colors, and (3) all counties of different colors.

Measures for ordinal and interval data

If X is ordinal scaled (ranked) or interval scaled, we could group the range of X into k classes, such as quartiles or deciles, and use the color lattice tests described above; in this case, a loss of information occurs. We now define two further coefficients which assess the degree of spatial autocorrelation between the $\{x_i\}$ in joined counties, where x_i is either the rank of the ith county (ordinal data) or the value of X in the ith county (interval data). Individual county values are therefore retained and the loss of information which occurs if the join-count statistics are employed is avoided.

The first coefficient was proposed by Moran (1950) and is denoted I. The second coefficient has been suggested by Geary (1954) and is denoted c. Both I and c are analagous to the classic form of any autocorrelation coefficient: the numerator term in each is a measure of covariance among the $\{x_i\}$ and the denominator term is a measure of variance. In terms of temporal autocorrelation, note that I reduces in one dimension to the familiar serial correlation coefficient; c corresponds in form to the Durbin and Watson d statistic (Durbin and Watson, 1971) used to search for temporal autocorrelation in regression residuals, and to the von Neumann ratio (von Neumann, 1941).

Both I and c have been shown to be asymptotically normally distributed as n increases. As with the join-count statistics, this result is assumed to hold approximately for lattices of moderate size, and I and c are tested for significance as standard normal deviates. The moments of I and c may be evaluated under either of two assumptions: normality (here we assume that the $\{x_i\}$ are the results of n independent drawings from a normal population) or randomisation. Under randomisation, whatever the underlying distribution of the population(s), we consider the observed value of I or c relative to the set of all possible values which I or c could take on if the $\{x_i\}$ were repeatedly randomly permuted around the county system. There are n! such values.

Choice of test statistic

When the researcher wishes to examine a data set for spatial autocorrelation, he will have to decide which of the coefficients defined above to use as his test statistic. The following guidelines are intended to help make that choice (Cliff and Ord, 1981).

(1) With binary (0, 1) data, the join-count statistics may be used. Alternatively, I or c could be employed by putting, say, $x_i = 1$ if an event has occurred in the ith county and $x_i = 0$ otherwise. However, with binary data, I and c reduce, apart from constants, almost exactly to the BW statistic. Thus there is little point with binary data in evaluating I or c rather than the join-count statistics. If the join counts are used, the researcher has the choice between the free and nonfree sampling models. Strictly, free sampling

choice between the free and nonfree sampling models. Strictly, free sampling may only be used if p is known a priori (exogenously). If p is estimated from the data by n_i/n, then only estimates of the moments are available. It is not known whether this would induce a serious inferential error, but in these circumstances the nonfree model may be more appropriate.

(2) With ranked or interval scaled data, I and c are preferred to the color lattice approach. In order to use the color lattice approach with these data, the $\{x_i\}$ must be grouped into classes, which results in loss of information. I and c preserve the individual x values and so avoid this problem. Results given in Cliff and Ord (1969, page 45) suggest that the variance of I is less affected by the distribution of the sample data than is the differences-squared form used in Geary's c. This is because the b_2 term in the variance of the Geary statistic has a coefficient of order n^{-1}, whereas for the Moran statistic the coefficient of the b_2 term is of order n^{-2}.

Limitations of measures

The join-count statistics, I, and c have two important limitations. First, they suffer from what Dacey (1965, page 28) has called the problem of 'topological invariance'. That is, once the connection matrix has been specified, the size and shape of counties in the system, and the relative strength of links between counties (road and rail links, for example) are completely ignored. The measures are, therefore, invariant under certain transformations of the underlying county structure.

To overcome this difficulty, Dacey (1965) suggested a measure of spatial autocorrelation where the weights are a function of county area and length of common boundary. Unfortunately it is not possible to express the moments of this measure in a usuable form, and so no test of significance is readily available.

The second limitation is one of usage. As defined, joins exist solely between physically contiguous counties. With connectivity thus specified, the measures search for spatial autocorrelation only between counties which are first nearest neighbours. Thus correlogram analysis, to determine how the autocorrelation function decays over space, was not attempted with these measures. There is nothing in the structure of the tests which prevents this kind of analysis. For example, we could define 'joins' to exist between counties which are second, rather than first, nearest neighbours. Two counties, i and k, might be called second nearest neighbours if they have no common boundary of positive nonzero length, but there exits a county j such that i and j are contiguous, and j and k are contiguous. Generalisation of the concept of a join to second and higher order neighbours in this fashion is easily performed using graph theoretic methods (see Haggett et al, 1977, pages 319-320). Even if this were done, however, all joins would still be given equal weight; and in some studies we might wish to give strong links between counties which are not contiguous, and weak links between contiguous counties.

4.3 The weighted coefficients

Instead of using binary weights to formalise the concept of a join, we can define a generalized weighting matrix W, $W = \{w_{ij}\}$, where we denote the effect of county j on county i by the weight w_{ij}. Weighted versions of the join-count statistics, Moran's I, and Geary's c statistic can be generalized from the original versions. The use of a generalised weighting matrix W, as opposed to a binary connection matrix, allows the investigator to choose a set of weights which are deemed appropriate from prior considerations. This allows great flexibility in defining the structure of the areal units and their relationships, and permits items such as natural barriers and county size to be taken into account. Further, if different hypotheses are proposed about the degree of contact between neighbouring areas, alternative sets of weights might be used to investigate these hypotheses. It is important to stress that care must be used in the choice of weights if spurious correlations are to be avoided. The factors which are most important will depend upon the study in hand. For example, the amount of interaction between any two counties may depend upon the distance between their geographical or demographic centers, the length of common boundary between the counties, and so on.

When generalized weights are employed, the join-count, I, and c statistics are still asymptotically normally distributed as n increases. An approximate test of significance is therefore provided, as with binary weights, by evaluating the coefficients as standard normal deviates. The values of the generalized or weighted moments for each of the statistics described are given

in Cliff & Ord (1981).

4.4 Interpretation of results

To interpret spatial autocorrelation coefficients generally, assume that I (or some other statistic) has been evaluated at several levels of spatial separation, such as for first, second, third,... order neighbouring cells. That is, we construct a spatial correlogram. Sokal (1979) provides the following summary of possibilities in the context of population densities, and it is possible to construct similar schemes for other spatial processes.

		Order of autocorrelation (spatial lag)	
		low	high
Sign of autocorrelation	positive	(1) dispersal from few sources (2) large favourable patches (3) gradient (trend)	(1) symmetrical surfaces (2) patchy arrangement
	negative	(1) heterogeneous study area (2) small patches	(1) gradient (trend)

In talking of patches (i.e., spatial clusters), we must consider the relative magnitudes of distances between individuals in the same patch or, alternatively, patch diameter and the magnitude (diameter) of the cell observed. Thus, when patch diameter is greater than cell diameter, we can expect positive low-order correlations, but when cell diameter exceeds patch diameter, we may get negative low-order correlations and positive higher-order correlations, depending upon the degree of regularity in the occurrence of the patches.

4.5 Spatial correlograms

Although the interaction between sites may be strongest between immediate neighbours, often the strength of interaction will vary in a complex way with distance. To detect such variations in the spatial pattern, we define a spatial correlogram by analogy with the correlogram used in time-series analysis (Kendall, 1976, page 70).

Consider a system of n cells with random variables $X_1, \ldots, X_n$ and let the cells i and j be gth-order neighbours (or g spatial steps apart). Various definitions of neighbourliness are possible. Thus, two sites i and j may be g steps apart in either of the following cases.

(a) If the shortest path from i to j on the graph connecting adjacent sites has g edges; that is, the path passes through (g-1) intervening sites ($g \leqslant D$, where D is the diameter of the graph).

(b) If the distance, d_{ij}, between sites i and j falls in the gth distance class.

Clearly the method of graph construction and the choice of distance function depend upon the investigation, so that the definition is very broad. The shortest paths for each pair of sites, as described in (a), may be evaluated using the matrix powering algorithm described in Haggett et al (1977, pages

319-320). If the variates refer to areas rather than to point locations, we may still construct graphs based upon common edges (and, possibly, vertices), or measure distances from convenient reference points such as the area centroids.

5. Two-Dimensional Spectral Analysis

By analogy to time-series analysis, a spatial correlogram in the distance domain has a corresponding spatial periodogram in the frequency domain. A double (two-dimensional) Fourier series is fitted to the values of a variate that have been collected at regular intervals on a cartesian coordinate system. The Fourier surface obtained can be viewed as analogous to a polynomial trend surface except that the surface is modeled with harmonic terms (i.e., sines and cosines) instead of polynomials.

Using a Fast Fourier Transform (FFT), a spatial spectral density estimate is computed. This results in an array of values at spatial frequencies in both the North-South and East-West directions. The interpretation of the Spectral surface is described in Rayner (1971) and Ripley (1981). Orientation is a crucial feature in the analysis; the spectral surface describes the variability in the pattern of variate values in different directions across a map. It is also possible to consider variance explained irrespective of orientation or direction. This can be done by averaging the spectral density estimates around semi-circles of constant frequency. A high value in this averaged spectrum indicates spatial periodicities at particular scales or distance intervals.

As with spatial autocorrelation and spatial correlogram techniques, spatial spectral density estimates can be used to describe the spatial patterning or interdependency of a geographically-distributed variable. Groupings of variables on the basis of similar measures can then be undertaken. Alternatively, clustering of areal units or cells can be carried out to maximize measures of spatial autocorrelation or spectral density.

References

1. Anderberg, M.R. (1973), Cluster Analysis for Applications , Academic Press, New York.

2. Cliff, A.D. & Ord J.K. (1969) "The problem of spatial autocorrelation", London Papers in Regional Science , pp. 22-55.

3. Cliff, A.D. & Ord J.K. (1981) Spatial Processes: Models & Applications , Pion, London.

4. Cormack, R.M. (1971) "A review of classification," Journal of the Royal Statistical Society (A) , Vol. 134, pp. 321-367.

5. Dacey M.F. (1965) "A review of contiguity measures for two and k-color maps" in Spatial Analysis: A Reader in Statistical Geography , ed. Berry, B.J.L. & Marble, D.F., Prentice-Hall, Englewood Cliffs, NJ., pp: 479-495.

6. Durbin J. & Watson, G.S. (1971) "Testing for serial correlation in least squared regression, III" Biometrika , Vol. 58, pp. 1-19.

7. Geary, R.C. (1954) "The contiguity ratio and statistical mapping" The Incorported Statisticain , Vol. 5, pp. 115-145.

8. Gordon, A.D. and Henderson, J.T. (1977) "An algorithm for Euclidean sum of squares classification," Biometrics , Vol. 33, pp. 355-362.

9 . Haggett, P., Cliff, A.D., and Frey, A. (1977) Locational Analysis in Human Geography , John Wiley & Sons, New York.

10. Hand, D.J. (1981) Discrimination and Classification , John Wiley & Sons, New York.

11. Kendall, M.G. (1976) Time-Series , Griffin, London.

12. Koontz, W.L.G., Narendra, P.M., and Fukunaga, K. (1975) "A branch and bound clustering algorithm," IEEE Transactions on Computers , Vol. C-24, pp. 908-915.

13. MacQueen, J. (1967) "Some methods for classification and analysis of multivariate observations" Proc. 5th Berkeley Symposium , Vol. 1, pp. 281-297.

14. Moran, P.A.P. (1948) "The interpretation of statistical maps" Journal of the Royal Statistical Society (B) , Vol. 10, pp. 243-251.

15. Moran, P.A.P. (1950) "Notes on continuous statistical phenomena" Biometrika , Vol. 37, pp. 17-23.

16. Rao, M.R. (1971) "Cluster analysis and mathematical programming" Journal of the American Statistical Association , Vol. 66, pp. 622-626.

17. Rayner, J.N. (1971) An Introduction to Spectral Analysis , Pion, London.

18. Ripley, B.D. (1981) Spatial Statistics , John Wiley & Sons, New York.

19. Sokal, R.R. (1979) "Ecological parameters inferred from spatial correlograms" in Contemporary Quantitative Ecology , ed. Patil, G.P. and Rosenzweig, M.L., International Cooperative Publishing House, Fairland, MD, pp. 167-196.

20. Von Neumann, J. (1941) "Distribution of the mean square successive difference to the variance" Annals of Mathematical Statistics , Vol. 12, pp. 367 -395.