A CONCEPTS-RICH APPROACH TO SPATIAL ANALYSIS, THEORY GENERATION, AND SCIENTIFIC DISCOVERY IN GIS USING MASSIVELY PARALLEL COMPUTING

Stan Openshaw

WORKING PAPER 93/27

SCHOOL OF GEOGRAPHY • UNIVERSITY OF LEEDS

WORKING PAPER 93/27

A CONCEPTS-RICH APPROACH TO SPATIAL ANALYSIS, THEORY GENERATION, AND SCIENTIFIC DISCOVERY IN GIS USING MASSIVELY PARALLEL COMPUTING

Stan Openshaw

School of Geography
University of Leeds
Leeds LS2 9JT

1. Introduction

A major criticism of geographical analysis has been its inability to provide results that are rich in a traditional conceptual sense. Quantitative geographers often claim to be concerned about developing and testing theory but this task is hard, confirmatory methods do not always mean much in a geographical setting, and there is a notable lack of major theoretical achievements that can be explicitly traced back to nearly 30 years of quantitative analysis and mathematical modelling of spatial databases. The GIS revolution does not appear to be of much assistance here because it is creating an immensely spatial data-rich situation but without providing sufficiently powerful spatial analytic and spatial process modelling tools to allow users to cope with the new opportunities. Increasingly, it seems that it is not only that we can see the forest for the trees but we cannot even see the tree we are sitting in. Quite simply we currently lack tools that are sufficiently powerful to release the empirical regularities that exist amongst the noise in our spatial databases.

Furthermore, there is a wide gap between the apparent intellectual elegance and generality of those abstract spatio-theoretical constructs that exist as a legacy from the pre-GIS years and the relatively theory, poor, by comparison, empirical results obtained from the geographical analysis and the mathematical modelling of spatial data. Also those, often qualitative researchers, who are seemingly most able to develop descriptive theoretical and pseudo-theoretical conceptualisations of the processes and dynamics of spatial phenomenon, usually fail to deliver testable hypotheses and computer models capable of either validating their theories or yielding useful and testable predictions in any remotely scientifically acceptable manner. As a result much of the current process knowledge base in geography is extremely fragile due to this history of untestable hypothetical speculation about the nature of the world. Likewise the computer based empiricists who are seemingly expert at developing models and performing detailed computer analyses are often quite unable to generalise their results as useful theory. They too have been poor performers at the task of scientific discovery. Indeed, it is difficult to see how this theoretical and technical gap might be bridged especially as the various parties on either side of the divide often seem to belong to separate ontological universes, characterised by very different philosophies and research paradigms. Both sides seem to ignore the problems in what they advocate by adopting a philosophical stance that deems such concerns to be largely irrelevant; see Openshaw (1991, 1992).

Yet it is essential to consider how in the 1990s the computational geographer might start to bridge this conceptual chasm. This is important if only to avoid a rerun of the quantitative revolution disaster of the 1970s when many of the first generation of quantitative geographers wandered off into purely social theoretical domains never to return to spatial analytical concerns. Arguably one reason for this was their dissatisfaction with the conceptual poverty of the newly found quantitative analysis tools. Maybe also the justification involved a search for an improved social behaviouralist understanding of the world, following disillusionment about the shallowness of the spatial analytical paradigm at a time when computers were extremely slow, data were extremely poor, and the available analytical methods were often inappropriate. However, the computing and data situations are today vastly different from the 1960s and 1970s. Indeed the GIS revolution is generating vast amounts of spatial information and geography relevant data. However, GIS is still extremely poor in terms of its representation of basic spatial concepts and its ability to generate geographical theory although very useful as an applications orientated tool.

It is with some justification that the critics of GIS regard the developments in the area as a re-birth of quantitative geography. Viewed in a narrow historic way it can be seen as suffering from similar limitations. Progress in basic digital map information handling technology has far outstripped the ability of geographers to do much that is new, creative or conceptually useful with the databases. Indeed, the current era of GIS induced data richness is emphasising our inability to generalise pattern and process, and synthesize new concepts and discoveries from the minutiae of detail found in spatial databases. GIS can, therefore, be regarded as conceptually shallow and as a result it gives the impression of being not particularly intellectually satisfying in terms of either its ability to explain the observed spatial reality it can represent or as an aid to theoretical understanding. Additionally, its strongly applied focus sometimes fosters the misleading impression that is not even a particularly scholarly area; although, of course, it is essentially an enabling technology and not an end in itself. So to some extent, sometime in its future, GIS faces the spectre of a possible rerun of the great quantitative revolution disaster of the 1960s unless a much greater degree of development occurs to the spatial analytics theory generation, and scientific discovery side of GIS.

The challenge for the 1990's is to try and discover how this might be achieved and to identify ways by which the gap between the theoretical and empirical sides to geography and GIS can be narrowed, perhaps by the development of a new style and new forms of computer analysis. It is noted that the developments in computer hardware and computing methods that has stimulated the GIS revolution have also sparked off major, new, and often revolutionary developments in many other areas of computer science, pattern recognition, and artificial intelligence. In particular, there is an increasing set of practical AI based tools that can in principle handle soft information and abstract concepts. The implication is that it is no longer necessary to only think in terms of fairly crude, from a theoretical point of view, data orientated, exploratory spatial analysis of the positivist sort dominated by number crunching applications and statistical methods dating mainly from the 1960s. Of course, there is no need to be exclusive and it is extremely important to continue the development of computational approaches to exploratory geographical analysis making use of new methodologies and massively parallel hardware as and when they becomes available during the 1990s. However, it is also noted that as it becomes possible to handle abstract information in the form of concepts and theories, that this may be regarded as providing a means of opening a "second front" in spatial analysis based on an inversion of the normal perspectives by seeking to develop a new form of exploratory spatial analysis that is both concepts rich and dedicated to theory generation and scientific discovery.

There are seemingly at least two ways forward. First, it would be feasible to consider the development of a computer language designed from first principles to handle the basic theoretical concepts of geography imbedded in a space-time framework (Sture, Holm, Goddard, Openshaw, 1990). This is, however, likely to be a slow and lengthy process. The second strategy is to find a means of directly using existing theoretical and pseudo-theoretical concepts thought to be relevant to geographical analysis by the development of new procedures that can take them into account when performing spatial analyses. A corollary is the corresponding requirement to try and rectify the surprising scarcity of relevant spatio-theoretical notions in geography by seeking to discover new ones by induction from the data rich GIS environments of the 1990s. The need for the creation of new knowledge by discovery is obvious; for example, it seems that our knowledge of the spatial patterning that characterise our towns and cities has advanced little in the last 20-30 years despite massive improvements in data provision. Seemingly most towns exhibit strongly recurrent regularities in many key variables. We know these patterns exist but until recently we did not have access to the pattern recognition technologies needed to identify them in a formal scientific fashion.

The purpose of this article is to consider how to develop from scratch a concept rich form of spatial analysis by seeking to exploit developments in AI and computer vision in an era of massively parallel computing. The analysis process is designed to be based on concepts that derive from geographical theory; and where these are poorly developed, to discover new theory by exploratory data analysis. It is noted that a theory based concepts rich style of spatial analysis is surprisingly undeveloped. Traditionally, this area has been almost entirely dependent on the application of inferential statistical methods which are often too narrowly focused to be of any great value in geography. Additionally, there was always the lingering problem of what inference actually means in a spatial context. It should now be possible to do much better by the application of AI methods and by seeking geographical uses for recent developments in computer vision. In some senses also the new opportunities concern how best to use the impending next generation of massively parallel computers expected to be capable of sustained terraflop speeds by the late 1990s, in order to make the most of the richness in spatial databases that GIS is providing. In some ways it might soon be possible to become smarter by becoming dumber in what spatial analysis tools are expected to produce. Why this might be considered useful and how in general terms it can be applied is the subject of this Chapter.

2. Towards a concept based spatial analysis paradigm

2.1 Spatial autocorrelation as a concept

It is perhaps useful to start be focusing on a space only form of spatial analysis. This is the traditional map metric of geography and is where GIS offers many manipulative and visualisation tools. The extension into space-time is left for later study on the grounds that geographically speaking it is much less urgent.

Consider an example of a apparently simple spatial analysis task. For a long time it has been a widely practiced procedure to look for evidence of spatial autocorrelation in univariate mappable data. Various methods have been derived to test for its presence and to quantify its extent. However, a far more basic spatial question is not how much there is but where it might be found or not found. The so-called "First Law of Geography" (Bunge, 1966) is based on a very simple concept; that nearby areas are more likely to be similar than areas far apart. So why not operationalise this geographical concept as a concept rather than as a precise, assumption dependent, statistical test of a hypothesis. The statistical test approach requires a much greater input of knowledge than is contained in the original theoretical notion; for example, a weights matrix and a null hypothesis to compare the observed result against. The weights matrix has to be chosen in advance of the analysis and reflects the form of spatial autocorrelation that is to be detected. It is also sometimes necessary to provide a detailed specification of the alternative hypothesis; for instance that the value in one area is negatively correlated with neighbouring areas or that the spatial autocorrelation follows a first order Markovian process. Even when a "significant" level of spatial autocorrelation is found to exist, this result contains no information about what might have caused it. If the null hypothesis is not rejected, then there may still be spatial autocorrelation present but the test statistic may have simply failed to detect it; for instance, if the spatial autocorrelation is localised rather than widespread throughout the study region. In both instances, the geography of the problem is lost in the statistical analysis. It is not very interesting to have an answer that merely says that there is evidence of positive spatial autocorrelation throughout (or averaged over) an arbitrarily defined study region or that even it exists within x-km of a specified point. The former is far too geographically vague to be of much assistance and the latter is too specific, too scale and aggregation dependent, and too parametric to be much help in an exploratory context where prior knowledge is both limited and uncertain. It is the non-stationarities or the spatial heterogeneities in the GIS database that often are of greatest interest. It should be relatively easy to spot spatial patterns that repeat throughout a study region. There is really little of interest in many such phenomenon; mapping the data will usually identify them. Instead, it is the much more localised, perhaps surprisingly, neighbourhood based patterns or anomalies that are often of the greatest interest; and it is these that conventional methods miss almost completely. The geographical question is much vaguer than the statistical tests can handle. It can in essence be reduced to a series of linked questions: (a) Is there any spatial autocorrelation? (b) Where is it? And (c) What form does it take? Ideally, the answer should be visible on a map for visualisation purposes. Question (b) is also the answer to questions (a) and (c). The hope is also expressed that the resulting map patterns may contain useful insights about underlying process.

2.2 Spatial association as a concept

Similar considerations apply to concepts of spatial relationships although here the problem is far more data rather than concept orientated as the idea of spatial association is usually an empirical notion that is application specific. Nevertheless, there is a whole plethora of statistical and mathematical modelling techniques that seek to identify, specify, and quantify spatial relationships. Again, they are all extremely precise; for example, a regression model requires a parameterised specification of the explicit mathematical form of the assumed relationship; for example a linear function or one specific nonlinear one. Of course, when this precise functional information is missing, then various exploratory methods can be used to help the researcher specify and test plausible or possible functional relationships. Unfortunately, once again when viewed from a concepts perspective, the questions that the statistical technology can answer are simultaneously too specific, too precise, too narrowly defined and too assumption dependent. The concept of spatial association is much vaguer; for instance, that there is a spatial relationship without necessarily being any more specific in the first instance. It is not necessary to start by assuming a functional form; for example, that it is linear. There is no idea of what form or shape or nature that the relationship should or must take in the concept of spatial association, only that there may either be one (or not) as the case may be. Testing a linear regression model only indicates whether a global linear association exists throughout the study region. This is extremely specific and only one of a myriad of different possible functional forms. It is also not very interesting because it contains so little theoretical information. Far more relevant would be the discovery that a spatial relationship exists but without necessarily knowing precisely how to specify or parameterise it, in the first instance. It would also be nice to know where abouts on a map of the world or the region of interest such relationships exist and where they do not. Of course, still keeping things simple and general, the mathematical nature or shape of the relationship in terms of abstract functional forms might also be interesting to know; for example, do spatial variables X and Y increase together in some way, or is there an inverse relationship, or do they go up and down together, or do they merely fluctuate randomly in their values, or is there a small number of classes of different type of relationship. Again the aim is to categorise the map space by the general nature of the association that is found; free of having to be any more precise or being forced to assume that the spatial domain of the relationship is the entire map area. Modelling the relationship is a subsequent activity. Yet at present the ability to detect the presence of a relationship entirely depends on our ability to model it in some highly specific way before we can know what it is, and this restriction is quite unnecessary from a broader theoretical point of view.

2.3 Distance decay concepts

Another basic geographical concept is that of distance decay; for example, this is regarded as applicable to spatial interaction data and to urban population density patterns of cities around a central point. It is also implicit in the spatial autocorrelation concept. The classical quantitative geographic approach would involve trying to fit one or more a priori, supposedly globally defined, distance decay functions to a whole data set. Indeed there is a battery of mathematical models that could be used. However, the basic conceptual concern is whether or not there is any evidence of distance decay patterns in an abstract and highly generalised non-parametric sense. There is no need to worry about what might or might not be the precise nature of its functional specification, at least initially. Does it exist and if so, whereabouts. If the concept is in fact a valid generalisation, then there should be evidence that it is applicable in many different parts of the world and also of some areas where it does not apply at all. By adopting only a statistical approach, there is a danger of being too precise in applications where the prevailing theoretical knowledge is general and far vaguer than the data or current knowledge about form of distance decay relationships. In short spatial analysts have traditionally ended up testing highly specific bits of far more general theories, or of performing highly specific and detailed analyses in ways that make it difficult or impossible to subsequently express the results as generalisable concepts. No wonder the first quantitative revolution in geography failed.

2.4 Social pattern concepts

More complex and more abstract theoretical notions are also relevant here as a source of inspiration. Unfortunately, and once again, the classical Park and Burgess (1925) concentric ring model, Hoyt's (1939) sector model of theoretical urban social structure, the multiple nuclei model of Harris and Ullman (1945), and the Shevsky-Bell (1955) models of multivariate urban structure are far removed from either the micro-scale, small area, based empirical results of the factorial ecologists (Davies, 1984) or, more latterly, the residential area taxonomists who analyse small area census data; for instance, Charlton et al (1985). Traditionally, social geographers have attempted to relate their results in a vague and general way to these theories; see for example, Clark and Gleave (1973) or Johnston (1979); but the task is extremely difficult. Somehow the researcher has to try and project the immensely spatially detailed and multivariate complex results of small area census analyses back into a much more abstract and generalised conceptual framework of theories about the structure of the urban mosaic; see for instance, Robson (1975). Seemingly this is an almost impossible task using conventional statistical approaches. The differences in scale and detail between general theory and the immensely detailed empirical analysis of census data are just too great to be easily reconciled, if at all. The microscopic nature of the data analysis emphasises the unique, whereas the macroscopic and abstract nature of the theories of urban social structure emphasise generality, recurrency, and ignores the micro details. It is extremely difficult to relate one to the other in any sort of direct manner, and thus obtain useful and testable hypotheses or even a typology or library of recurrent urban area pattern types. It seems that all too often the uniqueness of place has combined with the micro detail of large scale analyses to hide the macroscopic generality from the spatial analyst. We know it exists but we have so far largely been unable to systematically seek it out and identify the precise nature of the recurrent patterns in a scientific manner. As a result, quantitative geography is rich in statistical and mathematical technology but poor in terms of its concepts. Again the challenge is to try and develop a different approach to spatial analysis that is based more on basic geographical concepts than on statistics and which directly addresses these concerns.

3. A pattern recognition approach

Consider again the Burgess concentric ring theory as a concepts based model of a particular general form of spatial pattern. The question is which, if any cities, exhibit concentric ring patterns in their social structure? The traditional inferential statistical approach would be to test its existence in a data set by a statistical test of some kind. However, in practice it is difficult to devise a relevant null hypothesis to test. The statistical approach is far too precise for the abstract nature of the query given the scale-free nature of the original theoretical concept; for instance, what scale of data should be used, how wide should the rings be, around which point should the circles be formed. It is noted that many urban geographers in the 1960's and early 1970's did treat the concentric ring theory completely literally and did try and fit idealised rings to cities; see for example, Murdie (1969). Additionally, it is possible to imagine situations where the concept applies but only partially, or only in certain directions or else has bits missing or added. No statistical method is going to be clever enough to handle these problems and as a result there are going to be problems with whatever null hypotheses are devised by the researcher. A conventional statistical test is also completely aspatial, in that it does not treat patterns in spatial data as a spatial entity. However, if it is treated as a pattern recognition problem, then the question is now simply whether or not a concentric ring type of spatial pattern can be found in urban social structure. The scale at which such a patterning exists is not known and has to be determined but need not be fixed. There is no need to specify the key dimensions as a priori information. Again the question is simply whether any kind of concentric structure can be found without worrying too much initially as to its precise morphology or to worry about parameterising its size and shape. The latter are statistical concerns that are initially of secondary importance. If such patterns exist, then where do they exist and at what scale. If a concentric ring model is inappropriate then which of a small library of alternative urban spatial structuring concepts might be more appropriate. If different or new urban pattern types are found to be relevant, then they can be added to the library.

Another way of asking similar questions is to reformulate the task as an exploratory analysis problem. An even simpler question might be "Do cities exhibit any kind of geometric or geographical pattern based regularities in their spatio-social structure at all?" If so, what do they look like, at what scale, and how recurrent are they. Another example might be even vaguer. It has long been speculated that cultural and socio-economic differences between countries, and sometimes within them, have produced characteristic but different types of urban form, social spatial structure, and urban development. For example, the concepts of New Towns or colonial cities or port towns or industrial towns abound in the old geographical literature. It would be possible using multivariate statistical methods to try and identify what these characteristics might be. However, such methods cannot handle any realistic representation of a town's spatial pattern. All the within-city geographical detail is lost once the city is represented as a list of values for M variables. Alternatively, you could identify from the literature a library of different city spatial pattern types and then use this as a reference list for template matching against particular cities. Note that obviously no two cities are going to be identical in their layout and structure for all sorts of unique reasons; for instance, there will often be effects caused by site, with preferences for development in different directions, differences in size, topological controls etc. The pattern recognition question is whether their overall patterning is similar in an abstract, scale and rotation invariant, and size independent way. You could also ask the more focused question as to which cities are similar in a spatial pattern sense to Manchester or Newcastle or New York. In essence, the objective is once again to convert existing theoretical concepts into generalised abstract spatial patterns and then to search for their existence in a similarly abstract and generalised manner in GIS databases.

Pattern searching is not the same as hypothesis testing because there is no relevant null hypothesis. This point was lost on the original quantitative geographers. For instance, despite a superficially promising title (Patterns in Human Geography), Smith (1975) in common with other statistical geographers failed to develop a statistical theory of spatial analysis as distinct from providing examples of statistical methods being applied to spatial data in the search for largely aspatial patterns. The danger now is that the same mistake will be repeated 20 years later in the GIS era by a failure to appreciate that spatial patterns are themselves geographical objects that can be recognised and extracted from spatial databases. The question is does a theoretically derived abstract spatial pattern of urban social structure fit spatial reality as modelled in a GIS database? If it does, where does it, and at what scale. If it does not, are there any alternatives that might be more relevant? In this context then, the question does Leeds show a similar spatial arrangement as Newcastle; or are major British cities characterised by a small number of different theory based abstract spatial pattern types and, if they are, at what spatial scales and levels of multivariate generalisation are applicable.

In essence the objective is to develop a concepts rich form of spatial analysis by looking for the existence of abstract spatial patterns obtained initially from theoretical geography, by searching for the existence of bits of them. In short, the gap between theoretical and empirical geographical analysis may be bridged by the development of a concepts based pattern recognition approach to spatial analysis.

A final example concerns the concept of clustering in space. A test of spatial clustering is not very helpful at all. If the null hypothesis of some-kind of randomness is disproven, then we still know very little about the nature or extent or location or intensity of the clustering; see Openshaw and Craft (1991) or Besag and Newell (1991). It is also possible that the form of spatial randomness assumed by the null hypothesis is wrong; for example, the usual Poisson assumption is very much the assumption of last resort because we do not know any better. If the test of some kind of non-randomness is not disproven, then the data may still be clustered but all that may be happening is that the test may be too insensitive or the clustering too spatially localised to be detected. Disproving a null hypothesis of spatial randomness is likewise not very helpful because it provides no clues about what might be a more appropriate pattern. Additionally, the results can be affected by choice of study region and scale of data being analysed.

It would be much more useful to test or search for the existence of different types of clustering phenomenon, without expecting that the same type of clustering may be applicable to the entire map space. Some of these ideas are incorporated in the GAM-K cluster searcher of Openshaw and Craft (1991), although here only one type of cluster pattern could be detected (i.e. non-Poisson) in data viewed as excess cancer incidence; the effects of variations in population at risk, age-sex covariates, and spatial aggregation effects having been removed. There is no reason why the underlying principles cannot be developed further and the map space classified into regions of randomness and other areas where different types of scale and rotationally invariant clustering phenomenon might be more relevant. For instance, are there areas where the Seascale or Dounreay or Gateshead "type" of clustering phenomenon seem to be applicable. As the library of different "types" of clustering phenomenon is established, so the ability to recognise the locations of the different types becomes important and useful. However, in this context, there is no well established theory or concepts that can be applied "off the shelf" and it would be necessary to create a library of different theoretical clustering pattern types; for example, single centered with rapid distance decay or an elevated region of uniform clustering. The patterns of spatial clustering contained in such a library would provide the basis for further theoretical developments. In this way, it is also possible to develop a concepts-rich approach to exploratory spatial analysis and in the process develop appropriate theoretical generalisations.

There are a number of approaches that can be identified as a means of achieving these goals. Section 4 looks at production rule based approaches designed to identify generic features and Section 5 considers neurocomputing methods. At present knowing more or less what you want to do is more important than how it is done. Nevertheless, a brief outline is useful to establish feasibility.

4. Rule based approaches

Clearly there is no single approach that can tackle all these issues and different techniques may be appropriate in different applications. The simplest and most direct approach is to encode the basic concepts as a general rules which can then be applied to a GIS database if this is possible, bearing in mind the requirement that the rules will have to be scale, rotation, and aggregation invariant. It is possible that this strategy will not always work; for instance, if the theory based spatial patterns are highly complex. However, the GAM-K cluster searcher contains an example of this approach, in that there is a cluster recogniser that examines the spatial evidence for clustering and then looks for the presence of a certain type of clustering (viz. a peak value of excess with a distance decay structure). The rules for this recognition procedure are extremely simple (Openshaw, 1992) but this particular approach may not always be generalisable to all types of spatial theory. It also assumes the pre-existence of idealised, theory or concepts based, notions of spatial pattern. An initial library of types could, however, be created by computer based experimentation.

4.1 Identifying areas of positive spatial autocorrelation

A useful candidate for developing this rules based approach is the concept of spatial autocorrelation. This may be encoded in a completely generalised fashion as follows. The spatial autocorrelation concept implies that the similarity of areas diminishes as the distance separating them increases. A simple algorithm to detect this type of spatial structuring is as follows.

Step 1. Select a free area, perhaps ranked in descending size order

Step 2. Rank all other areas by distance from it

Step 3. Identify those areas whose data values (or more generally, a measure of multivariate similarity) compared with the central case diminish with distance and remove these cases from the database

Step 4. Link and plot these areas on a map

Step 5. Return to Step 1

It may also be useful to impose some continuity criteria on Step 3; ie the distance increments should be less than some maximum. Also the inverse (ie negative spatial autocorrelation) can likewise be of defined; that is the values will increase with distance. The no spatial autocorrelation case would imply that there is little change in the values with distance. Distance can also be measured in different metrics. The use of a topological definition (ie first, second, third etc order contiguities) might also be useful. In all cases, the results can be visualised as a map.

Note that there is no notion of scale or size. The method would be applied to basic spatial data and then to a sample of aggregation and scale changes. There is no notion in the theory of spatial autocorrelation that it should work at particular scales; maybe it will work at all scales, or maybe only at certain macroscales. An important task is to find out, where and when it works best by operationalising the search process on a massively parallel computer. A general principle is that with terraflop parallel processors likely by the late 1990s, we can seek to become clever about our spatial analysis by global brute force searches that can exploit massively parallel hardware. It is useful to remember that map searching is an explicitly parallel process. Indeed computational geography is one of the few subject areas that has several problems that are explicitly suitable for parallel computation, and which cannot be adequately handled on serial machines.

Note also that in Step 3, there is no attempt to quantify the distance decay effects. It would be possible but is it necessary as it would (1) be far too specific, (2) require the definition of some globally applicable model or function (viz. a particular autoregressive function that is assumed to be applicable throughout the study region), and (3) assumes a level of knowledge and statistical sophistication that is probably not necessary in the first instance in testing concepts whilst probably also being geographically inappropriate. It is argued that it should be possible to verify the reality of the concept of spatial autocorrelation in the same abstract fashion as it is expressed. To go further than this, represents a test of much more specific and less general concept. Finally, the relationship between the concepts of spatial autocorrelation and distance decay are clearly seen; viz. there is no real difference since the same rules would identify both.

4.2 Identifying areas where there is spatial association

Another rules based approach could be devised to detect spatial association. Here the search is for areas of the map that can be categorised in the simple bivariate case as being: areas where variables X and Y increase together, area where one or other decreases, and areas where they merely fluctuate in an uninteresting fashion. A suitable algorithm might be.

Step 1. Rank unallocated data by Y values

Step 2. Identify area i and flag all areas where X increases with increasing Y values. Compute a score for area i; viz. number of areas or map area of linked zones.

Step 3. Increment i and repeat until list of data exhausted

Step 4. Display areas associated with largest score and remove from list

Step 5. Repeat Steps 1 to 4 until no further allocation possible

In Step 2 it may be useful to impose a minimum increase threshold (eps) to focus on strong rather than weak association. The procedure can also be run to detect areas of negative association. The unclassified areas are assumed to be regions of no or little (ie less than eps) association. Note also that the procedure is scale and aggregation free. One of the objectives would be to search for map association that was indeed scale specific; merely by re-running this procedure on a large sample of scales and aggregations. Alternatively, the effects of area size and varying levels of data precision could be removed by first engineering a zoning system that removed any of the relevant confounding factors; ie variations in zone size.

4.3 Identifying areas where Burgess and Park's concentric ring theory applies

This is a more complex search pattern. The following algorithm can be used to find areas where it may fit. It is assumed that a small number of diagnostic variables are defined for the smallest available data zones. The algorithm is described in its univariate form. The search pattern assumes a gradual increase in wealth with distance

Step 1. Select a point x,y.

Step 2. Rank all data zones by distance from this point, out to a maximum window distance.

Step 3. It is assumed that social class increases within distance; so try and find a "low" value distance band by aggregating consecutive bands.

Step 4. Try and find a "middle" value band.

Step 5. Try and find a "high" value band terminated by a fall to background values

Step 6. Score the resulting "fit". If the score exceeds a critical threshold then save the details, else try somewhere else.

The search process is explicitly suitable for parallel computing. It could be systematic on a two dimensional grid or under the control of a intelligent search heuristic. Note that the scale at which any pattern is found is left undefined and is determined by interaction between the pattern rules and the data. Additionally, if there is more than one surrogate variable, perhaps with an assumed different pattern, (for example, an increase, then a decay, or an inverse pattern to that defined here), then the "goodness of pattern fit" can be modified accordingly. Maps showing areas of relatively high fit or goodness of fits that are unusually good (in some senses) might then be useful as an exploratory analysis of the data within respect to the assumed theoretical pattern.

4.4 Searching for a multiple sector model pattern

This is more complex because the pattern consists of one or more separate pieces. A possible algorithm involves the following.

Step 1. Define a suitable multivariate data set

Step 2. Define a point x,y and search window with side length z

Step 3. Regionalise the data within this window using a multivariate, contiguity constrained grouping procedure

Step 4. Score the results by comparing the map pattern of the groups with a library of theoretical target map pattern.

Again the search process can be systematic or under the control of an intelligent method. The search need not be too smart provided it is dumb in a parallelisable fashion. Note that here the spatial arrangement of the groups may no longer by relevant be as relevant as previously. On the other hand, all the urban model theories do specify topological constraints; for instance, low class residential is nearer the centre than high class and that there are social class gradients; for instance, low class adjoins medium class, medium class adjoins high class, and commuter zones may adjoin high class. These notions can also be built into the goodness of fit assessment. Additionally, the search rules will need to be adaptive. In the light of running with the simplest rules, new rules (perhaps embodying different concepts from those expressed in the original urban structure theories) may be needed (for instance, to allow for differences between US cities for which the theories were originally designed and UK cities to which they are being applied). These modified rules will themselves contain extremely useful information.

5. Neurocomputing and template matching

5.1 Template matching

An alternative strategy to that of coding general rules, is to search for defined target patterns using pattern recognition technology. This is far less sophisticated in that complex, data invariant, pattern recognition rules are no longer needed. On the other hand, once a library of spatial pattern forms are established, then there is a basis here for a far more general strategy. Spatial analysis now becomes a search of GIS data for previously identified types of spatial patterning that may be of some analytical, or theoretical, value. The problem now is how to achieve this goal in a completely general and spatial data invariant fashion. For instance, there is no value in searching for a sector model pattern with sectors of a fixed width and orientation, instead the essential characteristics of a sector feature need to be defined in some highly generalised and abstract manner that is scale, sector size and rotation invariant.

5.2 Black and white pattern search

In essence the pattern search process involves using a presence or absence template of the sort used in character recognition systems. Such a template would be of a fixed size. In a spatial context there would usually be several different templates depending on the nature of the search pattern; (cf. an alphabet of letters). The search window would also have to be stretched and distorted in various ways to see whether it can be made to fit. Various methods exist for doing this; for example, James (1987), Hussain (1991); and most are inherently suitable for parallel processors.

A search for positive spatial autocorrelation would now involve identifying areas where there is a pattern of "high" values embedded in an region of "low" values. Negative spatial autocorrelation is the converse. Zero or low spatial autocorrelation implies a different type of spatial pattern. The size of the areas of "high" or "low" value would be varied and would reflect the sensitivity of the pattern detector and the scale at which the analysis was to be performed. A search for a pattern that indicates the presence of a spatial relationship might also assume that areas which are "high" on one variable are either "high" (positive association) or "low" (negative association) on another. The black and white patterns can be coded accordingly. Multivariate equivalents might also be imagined.

A search for a concentric ring pattern or a multiple nuclei pattern or a sector pattern can likewise be coded as a series of black and white templates. This time there would be sets of target patterns, each of which forms an idealised spatial pattern 'alphabet'. The data would be matched against a complete set.

A general search algorithm would be as follows.

Step 1. Define a search window centred around point x,y.

Step 2. Compare the data patterns found here against a pre-defined library of reference patterns

Step 3. Compute a pattern goodness of fit.

Step 4. Keep if there is a good match, else either store the unmatched pattern in the spatial pattern library or discard.

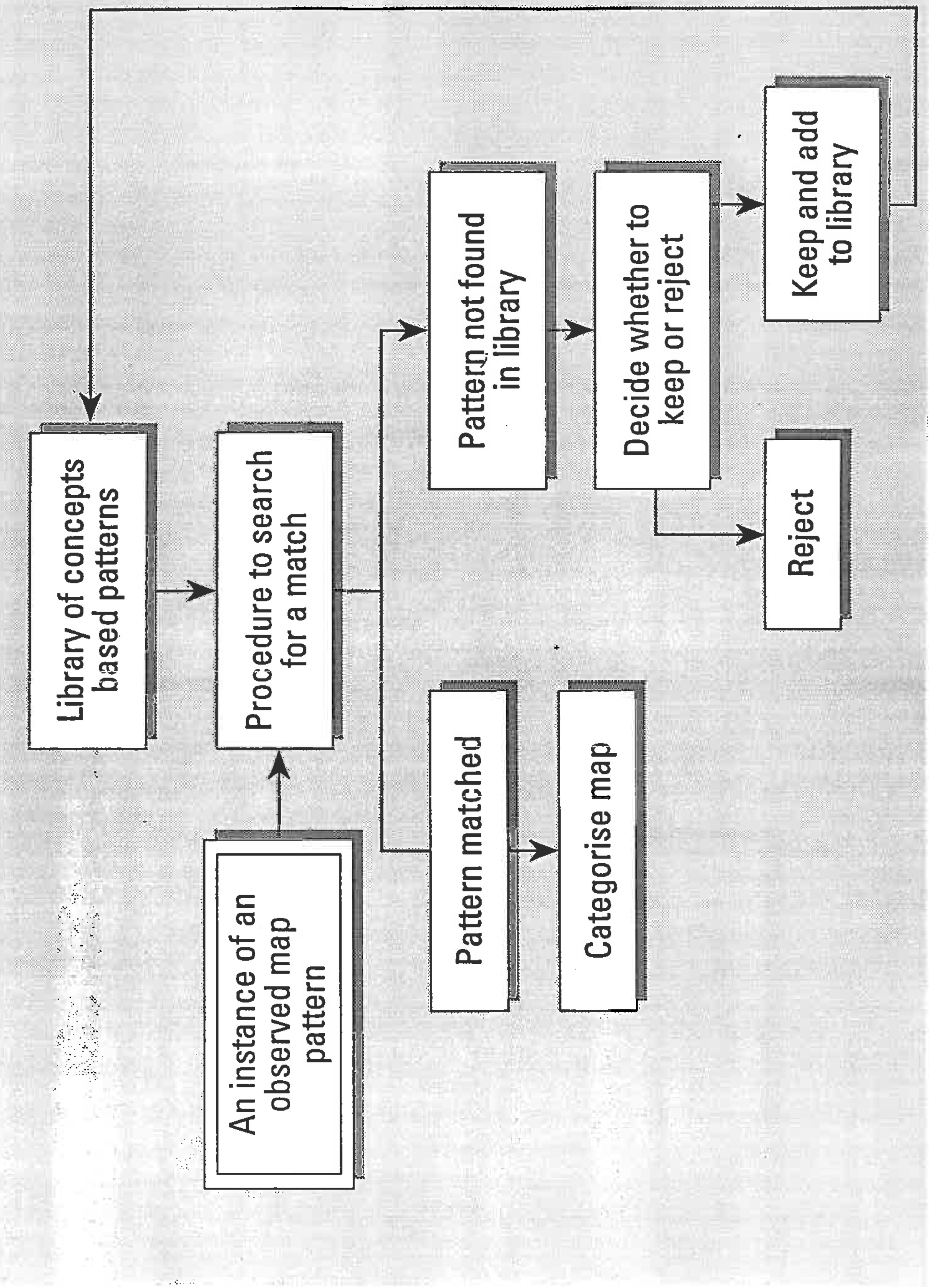
The spatial search in Step 1 could either be systematic (ie over a two-dimensional grid) or based on an intelligent search heuristic. The pattern goodness of fit in step 3 could be either a template match statistic or based on scale and rotation invariant measures derived from fast fourier or Hough transforms. The Step 4 stage offers the possibility of keeping patterns that appear to be present in the data but have not previously been specified in the spatial pattern library. This has the attraction of building up knowledge about "unexpected" spatial patterns, albeit at the cost of discovering how to discriminate between rubbish and interesting results.

5.3 Creating spatial pattern libraries

The discovery of apparently meaningful spatial patterns that do not fit theoretical expectations would be of great interest. They should be retained for detailed subsequent analysis and added to the spatial pattern library. Patterns that are frequently recurrent may well be of considerable theoretical interest because they clearly contain useful information; for instance, why else would they be so recurrent? Maybe they can be re-categorised as belonging to one or other of the hypothesised spatial pattern types; maybe they are indicative of other types of spatial pattern that have so far passed unnoticed. The general process outlined in Figure 1 would seem to be appropriate. This is clearly not a new idea; for instance, Wilson (1963) used a similar approach to generate hypotheses about bubble chamber events. Here the keep or reject decision would probably have to be based either on some Monte Carlo simulation to investigate whether the pattern was in some unusual (ie non-random) or on an arbitrary decision as to whether it possesses strong structural characteristics that make it interesting. The alternative criteria would be to store nearly all exceptions (based on very weak selection criteria) and then discard those which were either unique or had very low frequency counts later. Whether this is feasible depends on the speed of the matching process and the nature of the search. However, a limited sample of the study region could be scanned to establish an initial library of pattern types. Indeed it might well be conjectured that once the uniqueness of site, size, rotation, and local topologically induced distortions are removed, there might well be only a limited number of different spatial pattern types of any particular class of phenomena in the world. The pattern recognition task would be to create a cumulative and comprehensive database of the more important ones. The analyst "just" needs to find, define, and generalise the most important ones on a global scale; the most important ones are those that occur most often!

5.4 <u>Neurocomputing</u>

The development of practical neurocomputing based tools is a feature of the late 1980s. Some of these developments concerning associative memory are of considerable relevancy here as a means of spatial pattern recognition. Of particular interest here is the associative memory nets; specifically Hopfield nets and


Library of concepts based patterns
Procedure to search for a match
An instance of an observed map pattern
Pattern matched
Categorise map
Pattern not found in library
Decide whether to keep or reject
Keep and add to library
Reject

Bidirectional Associative Memories (BAMs) (Kohonen, 1988; Kosko, 1992). These nets can handle high levels of noise and would be used to search for theory based pattern keys again using multi resolution rasterised spatial data. Their main attraction is that the speed of recall is independent of the number of patterns stored, but they can also handle noisy data, and they are currently fashionable technology

5.5 Machine learning and computer vision

Another set of tools come from a machine learning and computer vision area. Exemplars, or idealised types of pattern template can be used as models against which data may be fitted via various deformations (Hill et al, 1992; Kipson et al, 1990). Genetic programming is another possible tool of considerable generality and potential usefulness here. It is noted that many of the pattern recognition procedures have been developed to assist in the task of battlefield target recognition at night. The problems of spatial analysis presented by GIS are probably considerably easier and to the extent that the problems overlap, there would appear to be a set of applicable tools just waiting geographical application; for example, Daniel et al (1992). It is imagined that the civilianisation of military technology may release many similar methods that might well be extremely useful here.

5.6 Searching for particular patterns or seeking to exhaustively classify entire map spaces

There are clearly two very different objectives of search that underlie both the rules based and the pattern search. The first, is a spatially exhaustive systematic search for all occurrences of a particular spatial pattern or library of patterns. The alternative, is to search for particularly good instances of one particular pattern type; without necessarily performing a space exhausting search. This might well be a good way of determining whether it is worthwhile performing a systematic search for a particular pattern type. Additionally, it would also allow the analyst to speculate about the presence of particular forms of spatial pattern and then see whether they exist, perhaps subsequently modifying the idealised or hypothesised form in an interactive fashion. This suggests that there are three modes of operation: (1) categorising a map in terms of a library of one or more model spatial patterns based on pre-existing concepts and theory; (2) theorising by extending the library to include new spatial pattern types that are found to exist in the database or need to be added after the categorising process is complete; and (3) speculating about what new types of idealised spatial pattern type may exist and then attempting to explain them in terms of theory. The result is a seamless hypothetico-deductive-inductive spatial analysis system.

6. Conclusions

It would appear that the GIS modelled world may be full of recurrent and regular spatial patterns in whatever domains we care to look. Maybe they remain undetected because the technology for "looking" for them does not exist in the computer era. Maybe this is one function that the traditional, pre-computer, geographer was able to perform with much greater efficiency when there was no compelling need to be statistical or computer based. It is also clear that today's spatial patterns often need to be found. The patterns that matter are not self-evident and the limited sensory bandwidth provided by GIS makes it difficult to employ the full range of human cognitive skills. As a result the sensory deprived geographers of the 21st century needs advanced pattern recognition tools to make the most of the rich spatial databases available to them. It is also apparent that there is probably a tremendous amount of geographical knowledge just waiting to be discovered; and that given the right tools geographers are about to embark upon a new age of exploration and discovery.

Undoubtedly also some of the apparent patterns will have been produced by essentially random processes; ie chaos theory but they may also be the product of underlying generalisable theories. What ever the cause, it is essential to have spatial analysis methods that can try and make them visible and then identify what they are by reference to both existing theoretical knowledge and empirically established spatial pattern libraries. A concepts rich, spatial pattern recognition approach powered by massively parallel computing with libraries of pattern exemplars, is one way forward that can effectively combine the soft and hard forms of spatial analysis in a new approach to GIS based exploratory descriptive geography. Clearly, there are some problems. Maybe this strategy assumes too much and attempts to be too clever. Perhaps the available spatial pattern related concepts and theories are too crude to be of much practical value, whilst many contemporary theories are perhaps too complicated to be easily or readily expressed as simple models of idealised spatial patterns. Yet it is seemingly a valid challenge to return to basic geographic principles and then seek to link the theoretical spatial patterns with empirical reality. It is not a matter of coarsening the data or of seeking to convert abstract theories into explicit tests of hypotheses. Rather it is a matter of looking for evidence that the abstract theories fit the observed spatial data patterns in a suitably abstract manner and if they do not, then lets find some that do. This does not trivialise the analysis task; indeed, it greatly complicates it. Moreover it will only be practicable if the associated massive computational problems can be resolved. A start has been made, but much remains to be done.

References

Besag, J. & J. Newell, 1991, 'The detection of clusters in rare diseases', J. Royal Stat Soc Series A, 154, 143-155

Bunge, W., 1966, Theoretical Geography, Lund Studies in Geography, Lund

Charlton, M., S. Openshaw & C. Wymer, 1985, 'Some new classifications of census enumeration districts in Britain: a poor man's ACORN', J. of Econ. and Social Measurement, 13, 69-98

Clark, B. D. & M. B. Gleave, 1973, Social Patterns in Cities, IBG, London

Daniel, C.E., Kemsley, D.H., Lincoln, W.P., Tackett, W.A., Baraghimian, G.A., 1992 'Artificial neural networks for automatic target recognition', Optical Engineering 31, 2521-2531

Davies, W. K. D., 1984, Factorial Ecology, Gower, Aldershot

Hill, A., F. T. Cootes, C. J. Taylor, A Generic System for Image Interpretation Using Flexible Templates, British Machine Vision Conference 1992, ed D. Hogg, R. Boyle, Springer-Verlag, 1992

Hussain, Z., 1991, Digital image processing, Horwood, New York

James, M., 1987, Pattern Recognition, BJP Professional Books, Oxford

Johnston, R. J., 1979, 'On the characterisation of urban social areas', TESG. 70, 232-238

Kohonen, T., 1988, Self-organization and associative memory, Springer-Verlag, Heidelberg

Kosko, B., 1992, Neural Networks and Fuzzy Systems, Prentice-Hall, New Jersey

Murdie, R.A., 1969, Factorial Ecology of Metropolitan Toronto, 1951-61. University of Chicago, Geography Research Paper No 116

Openshaw, S., 1991, 'A view of the GIS crisis in geography, or, using GIS to put Humpty-Dumpty back together again', Environment & Planning A, 23, 621-628

Openshaw, S. & A. Craft, 1991, 'Using Geographical Analysis Machines to search for evidence of cluster and clustering in childhood leukaemia and non-Hodgkin lymphomas in Britain', in G. Draper (ed) The Geographical Epidemiology of childhood leukaemia in GB, 1966-83, OPCS Studies on Medical and Population Subjects, HMSO, London

Openshaw, S., 1992, 'Some suggestions concerning the development of AI tools for spatial modelling and analysis in GIS' Annals of Regional Science, 26, 35-51

Openshaw, S., 1992, 'Further thoughts on Geography and GIS : a reply', Environment and Planning A, 24, 463-466

Lipson, P., A. L. Yuille, D. O'Kvanaugh, J. eefe, J. CaTaaffe, D. Rosenthal, Deformable Templates for Feature Extraction from Medial Images, Computer Vision -ECCV90, ed. O. Faugeras, Springer-Verlag, 1990

Robson, B. T., 1975, Urban social areas, Oxford University Press, London.

Shevky, E., and Bell, W., 1955, Social Area Analysis, Stanford University Press

Smith, D.M., 1975, Patterns in Human Geography, Penguin, London

Sture, Holm, E., Goddard, J. B., Openshaw, S., 1990, 'Towards a time geographical computer language for social science'. Unpublished MS.

Wilson, A. G., 1963, 'Hypothesis assigning in the testing of bubble chamber events', The National Institute for Research in Nuclear Science, Report 42, HMSO, London