WHAT IS GISABLE SPATIAL ANALYSIS?

Stan Openshaw

WORKING PAPER 94/15

SCHOOL OF GEOGRAPHY • UNIVERSITY OF LEEDS

1 Introduction

The GIS revolution that started in many countries in the mid-1980s is creating an immense wealth of spatial information in a large number of different application areas. More and more users will want to analyse their data with increasing degrees of imperative, whether for good public, community, commercial, research or other operational reasons. It is a 'scandal' that many key, important, databases are not already being properly analysed; for example, morbidity, mortality, cancer data, or crime databases; whilst commercial concerns and government agencies probably waste millions if not billions by poor spatial analysis of needs, inefficient spatial planning of facilities and poor targeting of resources. If better, more relevant, spatial analysis tools were available then they would almost certainly be used. Some even claim that their non-availability threatens the viability of GIS itself. This paper attempts to address some of these concerns by focusing on the general and generic aspects in an attempt to define in a general way the type of new exploratory spatial data analysis that is likely to be most useful for use within GISs whilst at the same time recognising the usefulness of some of the more traditional spatial analysis tools if they can be put into a form that users can apply.

The early 1990s seems a good time for a far reaching and fundamental re-think on GIS functionality because spatial analysis is entering a new era in the development of quantitative geography. The new analysis needs are being created and stimulated as a by-product of GIS. This is quite different from more traditional quantitative geography concerns. This change in emphasis and its implications have not yet been properly understood. The lack of recognition of these new needs can be seen in the continuing confusion as to what spatial analysis in GIS actually means. Moreover whilst there is a general concensus that spatial analysis functionality is needed in GIS toolkits, there is no agreement about what kinds of spatial analysis methods are most relevant to GIS and which ones are not. The various Spatial Analysis and GIS initiatives (NCGIA I-14) workshops and conferences and the ESRC's RRL initiative have not only failed to clarify the situation but have probably, yet unwittingly, contributed to the confusion. It is useful to note at this point that despite considerable publicity for the NCGIA I-14 Spatial Analysis and GIS initiative, it was later admitted to be primarily a public relations exercise. The total resources expended probably amounted to less than £35,000

(Fotheringham and Rogerson, 1993). Maybe the current state of spatial analysis confusion is, therefore, understandable.

The question "What kind of spatial analysis do we want in GIS" has to be tempered by the feasibility constraint of what kinds of spatial analysis can be implemented in GIS and the sensibility constraint, what kinds of spatial analysis is appropriate for use in GIS environments. Another set of general design constraints reflect other considerations such as who are the likely users, what is it they want, and what sort of analysis technology can they handle given likely levels of statistical knowledge and training. It is likely that the answers to these sorts of questions will shape the particular direction future GIS applications will take. **Table 1** summarises the principal design questions. It is noted that userability criteria are very important.

When these design principles are applied to the set of available spatial analysis, spatial statistical, quantitative or statistical geographic, and spatial modelling tools; then it is quite clear why many will never fit. It is also the major source of confusion. As a result the phrase "Spatial Analysis and GIS" means many different things depending on your viewpoint. Likewise, the debate as to whether these methods should be accessed from within or without a GIS package is quite irrelevant. There is no reason to insist on only one form of integration or interfacing, except the obvious point that to be a GIS tool the spatial analysis operation has to be called from, and end within, a GIS. However, this does not mean that only methods that can be put into a GIS are GIS relevant, but it should restrict and define the area of interest. Hopefully, by being more precise about what might constitute GISable spatial analysis, then much of the historical confusion in this area can be removed.

2 Statistical hangovers

It is useful to start by trying to put many existing traditional spatial analysis methods in a proper GIS context. Quite simply, they do not belong in the GIS era. Similar unkind thoughts can also be applied to many aspatial statistical methods originally designed for use in survey sampling. These are not usually spatial analysis relevant methods because they treat spatial information as if it were equivalent to survey data, and generally totally fail to both handle any of the key features that make spatial data special or provide results

Table 1 Basic design questions

What kinds of spatial analysis are:

- relevant to GIS data environments
- sensible given the nature of GIS data
- reflect likely end user needs
- compatible with the GIS style
- capable of being used by end users
- add value to GIS investment
- can be an integral part of GIS
- offer tangible and significant benefits

that are sensitive to the nature and needs of geographical analysis. The vast literature of statistical geographical books should in general be seen for what they really are; introductions to statistical methods that are of geographic interest mainly because they have been applied to spatial data examples. The application of aspatial statistical methods in GIS might appear sensible but it has to be recognised that the technology may often be inherently unsuitable and, whilst most may appear harmless, they are not at all safe. Many of these problems associated with the analysis of spatial data are not new. As a result, most of the statistical geographical literature is simply not relevant to GIS. Key texts such as Ebdon (1977) <u>Statistics in Geography</u> or Gregory (1963) <u>Statistical Geography</u> contain little or nothing relevant to GIS. They are not appropriate as a source of GISable spatial analysis tools, even if some of them are excellent texts describing the use of aspatial statistical methods to spatial data; for example, Taylor (1977) <u>Quantitative Methods in Geography : an introduction to spatial analysis.</u> They belong to a different era in geography when large spatial databases usually consisted of less than 30 zones.

Anselin (1989) clearly recognises that this problem exists when he argues: "With the vast power of a user-friendly GIS increasingly in the hands of the non-specialist, the danger is great that the wrong kind of spatial statistics will become the accepted practice. Since the 'easy' problems have more or less been solved, a formidable challenge lies ahead." (p14-15). Indeed knowing how to develop new and more relevant methods looks like being too difficult for current technology particularly when it has to be preformed in a rigorous fashion within a classical statistical framework. But is this really what GIS needs? Might it not be easier to contemplate more data descriptive and less statistical theory related technologies? It should be possible to do this because purely spatial statistical analysis is probably not what most GIS end-users want or need.

So whilst it is being increasingly recognised that aspatial classical statistical methods are not all that useful, there is still a strong lobby in favour of continuing to develop more spatial data relevant classical statistical methods. Griffith and Anselin (1991) neatly summarise the situation when they write "Classical statistical techniques have often been found to be woefully inadequate for solving geographic problems. The realisation that geo-referenced data are special has given rise to an increased integration of spatial statistical concerns into classical statistical frameworks. No longer does the training of geographers strictly in techniques of aspatial analysis constitute an adequate education, nor do brief excursions into the applications of classical statistics to spatial problems" (p. xiii). However, their version of Statistical Analysis for Geographers offers little more than what they argue is inadequate; that is, the application of classical methods to spatial problems with only minor changes to handle spatial data. The inclusion of spatial concepts into statistical methods may appear useful but will probably eventually 'fail' for one reason or another to be GIS relevant due to the complexity of the problems.

Haining (1992) also argues in favour of a classical approach. He writes "Whilst there are certainly important problems that are specific to analysing spatial data the position taken in this book is that many of these can be tackled within existing statistical methodologies." (p10). Haining is certainly correct if he thinks in terms of the simple analysis of (spatial) data. In Haining's view of spatial analysis, the GIS revolution seems irrelevant other than as a source of spatial data. Haining (1992) provides a good statistical geography book, but whilst it is not all aspatial technology, there is little that is really relevant to GIS.

It is unfortunate that people persist in asking the wrong questions. For example, Griffith's (1993) has a paper entitled "Which spatial statistics techniques should be converted to GIS functions?". The concern here is with developing gateways into GIS for spatial statistics, in Griffith's case they are Moran Coefficient related regression methods. However, this is fundamentally flawed technology, at least from a geographical perspective even if the statistics are correct. The problem here is a fundamental and deeply rooted geographical misunderstanding that dates back to Cliff and Ord (1969). The original mistake was to continue the 1950s spatial statisticians' assumptions that it makes sense to summarise whole map patterns or test whole map statistics for departures from randomness. The classic Irish County (N = 27 observations) data was too small for the fundamental limitations of interpreting whole map statistics to be evident. Scale the methods up by applying to much larger sets of zones and all manner of geographical problems appear.

The real challenge that GIS presents is to re-think what GIS relevant spatial analysis is needed, free from bias about the importance of one or other favourite method. It is becoming important to mentally dump much of the obsolete baggage of the statistical geography past and try to develop some spatial analysis tools that are simultaneously: GISable, relevant to the needs of GIS based users, and appropriate for spatial data.

The problem of geographical common sense also needs to be addressed. Far too many statisticians seem to hold the naive view that the role of the geographer is to define the problem, pass them some data, and wait for the definitive statistically approved results to appear. It is not satisfactory to separate spatial analysis functions from an understanding of the geography of the problem. Neither statistical nor geographical knowledge are by themselves sufficient, both are necessary. it is likely that the needs of spatial analysis and GIS will not be well served either by statisticians who admit to being ignorant of geography and who generally appear unaware of the nature of the spatial information they are concerned with, or by so-called statistical geographers who fail to realise that their statistical skills are probably inadequate for handling the problems they are interested in, who have forgotten it seems most of their geographical knowledge and who typically claim no great knowledge of the GIS world in which their techniques are now suddenly supposed to operate or be relevant. There is a growing recognition that there is no easy, soundly based in theory, statistical solution to many of the problems of spatial analysis. However, we should at least know what we need. It is useful, therefore, to define basic GISability criteria that is reasonable to expect that GIS appropriate spatial analysis methods should meet.

3 Openshaw's ten basic rules for identifying future GISable spatial analysis technology

It is useful to try and being quite clear as to what is needed without being too concerned with how to achieve it. The aim is to create a list of criteria able to discriminate between GISable and GIS irrelevant technology.

3.1 A GISable spatial analysis method can handle large and very large N values

In the GIS era data sets typically have large N values. It is interesting that the N values in many databases will soon reach their theoretical maximum values. For example, in the UK there are 150,000 census enumeration districts, 1.6 million postcodes, and a limit of 53 million people points. This may seem a lot but computer hardware can now handle these dimensions with ease. Hardly any GIS users will be interested in 50 or 100 zones; if they are, then they should realise that data of two or three order of magnitude greater resolution exist and might be preferable. Any statistical method that involves the use and manipulation of N by N matrices is probably not sensible to include in GIS. When N = 27 fine, but such data sets are rare. Of course it is possible to perform matrix operations on quite large scale N values but perhaps not within GIS. As computer power increases then the problem of matrix size in spatial modelling will diminish but it will still be a concern for a long time yet.

3.2 The most useful GISable analysis and modelling tools will be study region independent

Spatial pattern analysis results that depend on, or are determined by, the arbitrary definition of a study region are often not useful. This observation is a simple extension of the modifiable areal unit problem. It is also geographical common sense. At one level it is a trivial and obvious problem. If you study data just for Scotland and then repeat the analysis for England and Scotland, or just Leeds, then the results will probably differ. The differences reflect variation in the phenomenon under study, the choice of study region, and interactions between the two. However, the effects of this are so obvious that it is often completely ignored. For example, if you wish to summarise the distribution of pubs in Leeds by a nearest neighbour statistic then the results depend on the definition of Leeds and on the patterning of pubs within Leeds. In the north, it might be systematically and spatially even, in the south it might be random, and in the city centre clustered. The result for the whole of Leeds averages out all these differences and will almost certainly be a meaningless outcome unless the spatial pattern of pubs is more or less the same everywhere.

Another example might help clarify this further. Imagine a cancer data set with one very strong cluster. If a small study region is selected that includes the cluster (probably by chance) then virtually any spatial pattern statistic will detect its existence. If this is done deliberately then it is a form of gerrymandering the results by boundary tightening. Now increase the study region in steps and there comes a point when the cluster is lost in a sea of randomness. As this study region decisions are arbitrary and seldom part of the analysis process, so it follows that whole map statistics should not be used. They are not as helpful in GIS except where the whole map region is characterised by only one distinctive type of spatial pattern.

There are also scale effects that provide an additional level of complexity. For example, if the answer to the question 'does a disease cluster at the District scale' is no, this does not mean that some Districts will not have localised clustering within them or that at a small area level there might be considerable clustering. What it indicates is that when viewed at an aggregate District scale, clustering is not a characteristic phenomenon of the disease under study. This may be a useful result but from a geographical point of view it is the localised clustering of small scales that are of greatest interest. Indeed, it is hard to imagine why District level clustering (given the heterogenous nature of Districts) should be of interest to anyone. the disease processes simply do not operate at that scale.

It is argued, therefore, that in most analysis applications spatial pattern detectors are needed which can look within the map to find patterns and relationships that are geographically localised in extent and not global. This is important to avoid the modifiable study region effect and because geographic space is not characterised by uniform densities of anything. In a spatial context techniques that can detect or model only global patterns or relationships need to be used with great circumspection.

3.3 GIS relevant methods needed to be sensitive to the special nature of spatial information

To be relevant to GIS the spatial analysis tools should be able to handle in a reasonable manner spatial information, taking into account rather than ignoring any special features that matter. The corollary of this statement is that if an analysis tool has been developed for use with aspatial data then it will probably be inappropriate for use with spatial data without major modification.

Table 2 lists some of the features that characterise spatial data. Of particular importance is the spatially structured nature of data precision and errors, they are not spatially random. Database anomalies might be detected as patterns but could just as easily be data error and reflect other data inconsistencies. Also, the data to be analysed are often known to be wrong but still have to be analysed. An example would be the analysis of cancer data consisting of incidents for 1969-84 using denominator for either 1971 or 1981. The temporal resolution of the data differ. The data riches of the GIS world is often achieved at the expense of accuracy and the spatial analyst often faces the prospect of what might be termed 'poor data analysis', and of having

Table 2 **Features that characterise spatial data in the GIS era**

many cases/objects/points
many variables
large data volumes
spatial autocorrelated values
spatial data error of various types
non conformity with standard statistical distribution
data precision can be spatially structured
errors need not be random
not samples in usual sense
surrogate variables abound
nonlinearity is the norm
a high degree of complexity
modifiable areal unit, scale and aggregation effects
mixtures of scale and data resolution
mixtures of measurement
small number problems can be important

to analyse data containing a mixture of different levels of error and uncertainty. Likewise in the analysis of data for areas, more extreme results are often found in areas with the smallest denominators. Such areas tend to be geographically structured; i.e. rural rather than urban. Likewise the internal heterogenity of all zoning systems varies and this must influence the results. If the data are converted into continuous form, as surfaces, then the problems do not disappear. They are just hidden and present problems in a different way.

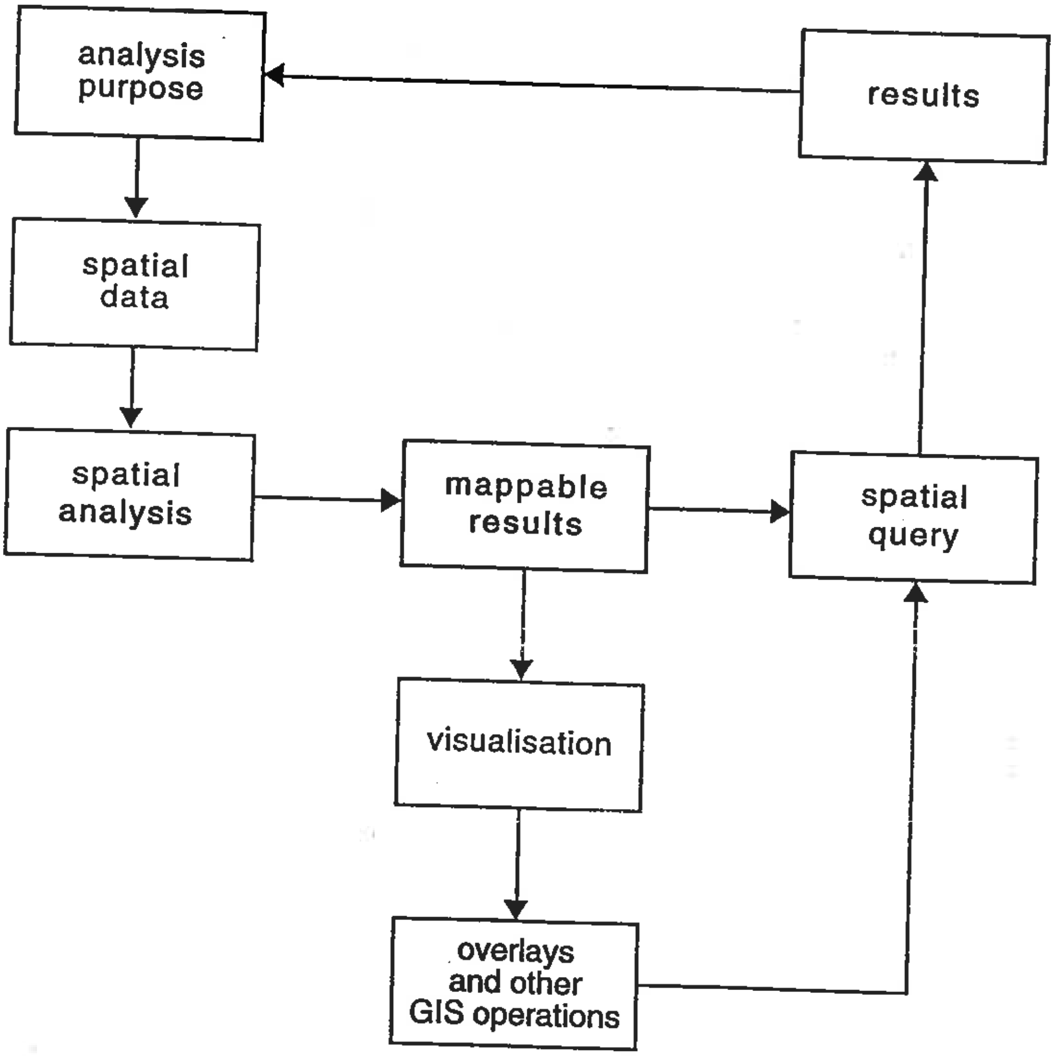
Openshaw (1989) argues that the real secret of spatial analysis in GIS is to handle rather than ignore the problems. It is simply unacceptable to assume that the modifiable areal unit problem does not exist. In many instances, a handle on the problems can be obtained via Monte Carlo simulation, by sensitively analysis, by boot-strapping, and seeking out high leverage points or data values. Some of the extra CPU power that is now available should be used to make the spatial analysis tools less naive and more robust in the face of the known characteristics of spatial information.

3.4 The results should be mappable

GIS is a highly visual and graphics orientated technology. The results of GIS spatial analysis operations should also be available in a graphic and mappable form. This is very important and in some ways represents a major design constraint on the forms of spatial analysis that might be considered as GIS relevant. The output cannot be a set of statistics or model parameter values embedded in a text report. The outputs also have to be visual and mappable. This provides for a different spatial analysis paradigm in which the emphasis is firmly on visibility and visualisation. Figure 1 summarises what is intended.

The role of spatial analysis is essentially that of a filter designed to remove the rough from the smooth or to highlight areas of 'unusualness'. Far too many geographers seek complex statistical solutions to problems that are best visualised, at least initially, so that when the time comes to apply more sophisticated methods there is at least some idea as to what to look for. The spatial analysis paradigm in Figure 1 should be applied iteratively; each iteration has an input of human intuition and knowledge; and with each successive iteration more insight is gained.

Figure 1. GIS spatial analysis paradigm



3.5 GISable spatial analysis is generic?

The GIS toolkit consists of a set of generic, application independent, tools for data capture, storage, manipulation, mapping and analysis. Very little, if anything, is application specific. The tools are generic. It follows that spatial analysis methods of the greatest claim for inclusion in proprietary GIS should be similarly application independent. It is claimed here the generic, general purpose, data invariant spatial analysis methods can in fact be defined. For example, methods that examine point data for clustering can be applied to disease data, crime data, earthquake, lightening, gas or water bursts, telephone faults, traffic accidents, etc. The function is application dependent. It was with concept in mind that Openshaw (1991a, 1991b) listed a set of general spatial analysis procedures; these are reproduced in Table 3. At the present point in time it is probably more useful to agree on broad areas of analysis functionality than it is detailed how to achieve it.

3.6 GISable spatial analysis methods should be useful and valuable

If users are to be lured into using spatial analysis then there has to be a good and compelling reason. The GIS relevant tools will not merely meet academic research needs but should have something tangible and valuable to offer genuine end users in applied contexts. The benefits have to outweigh the end users' perceptions of the 'costs' of applying spatial analysis; an area where perceptions of statistical and mathematical complexity might give the impression of outstanding difficulty. This mis-impression is not helped by the usually complex nature of spatial analysis conferences and workshops.

It also follows that if GISable spatial analysis methods can be defined, then they should not be free. In the GIS world where all aspects of everything are owned, copyrighted, licensed, and commercial, it is a mistake for vendors and system developers to assume that spatial analysis tools should be 'free'. It cannot be left to the research sector to develop, because what they will concentrate on is the research-interesting methods that are probably least likely to be of general interest or meet end usability criteria.

Table 3 Basic generic spatial analysis procedures relevant to GIS

pattern spotters and testers
relationship seekers and provers
data simplifiers
edge detectors
automatic spatial response modellers
fuzzy pattern analysis
visualisation enhancers
spatial video analysis

Source: Openshaw (1991a)

3.7 Interfacing issues are initially irrelevant and subsequently a problem for others to solve

In seeking to develop relevant spatial analysis methods for GIS, the nature of the linkage is irrelevant. At the present moment in time, it is far more important to demonstrate the utility of GISable and GIS relevant spatial analysis tools. The philosophical rights and wrongs of programming in this or that trendy language or system, of embedding or not embedding, of tight or loose coupling, of integrated or stand alone, is largely if not completely an irrelevant distraction. The computing world is moving towards open, distributed, heterogeneous systems with perhaps special or single function hardware being invisibly interlinked. It is not the spatial analyist's concern. Their task is to create the most useful and relevant technologies for analysis. Likewise, compute intensiveness is no longer a critical issue. If it can be done at all at present and there is a good case for doing it, then it will one day become common practice on affordable hardware located somewhere on a global network.

3.8 Ease of use and understandability are very important

GISability is not just about relevance to GIS environments and relevance to the needs of GIS users, it is also about the ability of the typical end users to use it. For this reason highly statistical technology and raw statistical methods will not usually be appropriate. They fail on ease of use and usability grounds. The statistical packages demonstrated long ago how even the most mathematically and complicated of methods can be presented in a fairly easy to use form. There are dangers in hiding complexity and allowing the statistically ignorant and unskilled to use highly advanced methods; but no more so here than in the other areas of GIS. The converse is also unproven. There is no guarantee that statistical experts will only perform sensible analysis with GIS data. GIS is based on the proposition that the complexity is hidden; for example how many users understand the computational geometry that lies behind many of the methods in their GIS packages. There is no good reason why the complexity of any relevant spatial analysis technology cannot also be hidden behind suitable user interfaces.

However, it should also be stated that there is no reason to assume that GISable spatial analysis tools must only involve statistically and mathematically complex methods. It is possible that the most relevant spatial analysis need not use any advanced forms of conventional statistical technology. It is not clear that any conventional statistical analysis skills are needed. Maybe the GIS vendors lack of interest in spatial analysis reflects a mistaken assumption, that advanced statistical skills are needed before even the simplest method can be used. Indeed, it might be observed that quite often there is an inverse relationship between the degree of mathematical complexity present in a spatial analysis method and the associated level of applied geographical usefulness. The proof of this is left to the reader but Haining (1992) and Griffith and Anselin (1991) provide clear examples of such an inverse relationship. On the other hand the statement may be misleading because maybe those researchers who are most likely to indulge in the statistical spatial analysis may not see themselves as doing anything applied.

3.9 GISable analysis should be safe technology

The applied nature of GIS makes it important that naive methods with a high innate propensity to generate spurious results should be avoided. Spatial analysis results should be reliable, robust, resilient, error and noise resistant, non-parametric, and not based in any important way on standard distributions. Simple tests of inference should be avoided. Monte Carlo significance tests, boot strapping, jack knifing and other computational statistics tools should be used wherever important analysis is being performed. GIS databases contain errors and uncertainties of various kinds and it is important that they do not mislead the innocent. The basic null hypothesis in GIS is not of randomness but of database error. Only if this can be rejected is it worth applying anything more sophisticated. The main problem with the analysis of GIS data is not that the data contain error and uncertainty or is wrong in various ways, it is that the analysis technology does not know or expect or understand such problems. The errors and uncertainties inherent in spatial databases can be handled only if allowance is made for them (Openshaw, 1989). Methods should be self-checking numerically and the basic rule of the user having to validate any really interesting results via independent data needs to become an established rule of thumb. In a GIS context, conventional significance testing is a very minimalistic barrier to cross.

3.10 GISable methods should be useful in an applied sense

It has to be faced that GIS is primarily and predominantly an applied technology. Those spatial analysis methods most relevant to it will be applied and seek to meet generic, important in the sense of being useful, and valuable applied goals. They should also focus on spatial analysis tasks that are relevant to GIS environments. Academic research needs fall into a different category. It has to be recognised and accepted that there are many wholly GIS inappropriate methods of analysis lying around that may well need to be discarded or never introduced into GIS toolkits. Indeed, there would appear to be many more of these designated unsuitable for GIS use methods than those that are. However, we should not throw away the baby with the bathwater - there must be room for both new generic spatial analysis routines which solve applied problems as well as conventional spatial analysis and spatial modelling methods which has a proven track record of applied problem solving.

4 Developing a spatial analysis paradigm that is appropriate for GIS

It would seem to follow from the previous discussion that future GISable spatial analysis methods will be:

(i) essentially descriptive;

(ii) essentially exploratory;

(iii) probably not inferential in a traditional spatial hypothesis testing sense; and

(iv) they may have to be invented from scratch although often the basic techniques already exist.

On the other hand, it is apparent that some suitable methods do exist and that these can be readily ported into GIS environments without too much difficulty. The principal problem is that of discriminating between those that are useful, those that might usefully be reconditioned or refurbished for spatial data applications and those that are not. Table 4 outlines an example of a set of spatial analysis methods being developed for Arc/Info by a consortium of users.

Table 5 outlines some of the basic design principles that are deemed important in. Clearly not all GIS users will agree with them and probably even fewer statisticians, but they do at least draw attention to the need to think about the underlying design aspects; in particular who the end users are, their levels of skills and what

Table 4 SAT/1 contents

```
Diggles point density method
a simple GAM
Besag-Newell method
cartogians
Kernel estimated surfaces
Stones method
Getis-Ord g statistics
zone design methods
regionalisation and classification
zone ranking
spatial regression modelling
```

Table 5 **Technical specification**

high degree of automation
self-checking
constantly seeking to improve performance
capable of real-time operation
can be presented as a black-box but is understandable
contains mechanisms for user interaction and incorporating human knowledge
can handle localised spatial heterogeneities
not hindered by lack of knowledge of what, where, or when to look for
handles rather than ignores spatial data characteristics
can respond to the unexpected in intelligent ways
can generate new hypotheses as an output from the analysis process

they see GIS as providing them with. The typical GIS end user should not be assumed to be an academic mainly interested in high level research activities that need extensive research training and high skill levels. Equally, designers of GIS relevant spatial analysis technology should not be mesmerised by a search for the 'holy grail' of soundly based statistically theory rich methods; such things in general do not yet exist for use with spatial data.

The real problem is that users do not know what patterns or relationships exist in the data. They do not know where to look for them assuming they know when to look or even what to look for. GIS databases are multivariate data rich. There are also now lots of them and it is likely that whatever prior knowledge the user has will either be wrong (maybe only partially) or inappropriate (due to the unexpected presence of much more distinctive but initially unknown orderliness).

The goal is easily stated: a search for orderliness and regularities that are either recurrent elsewhere or are of local importance. The problem is, as Openshaw (1993) explains, we are blind to the spatial patterns and processes that exist in our 'geocyberspaces'. We have probably only ever found a minuscule fraction of the regularities that exist and worse still what we have found so far might well be artefacts created by the arbitrary selections and restrictions imposed on the data just so that it can be analysed. There is a desperate need for new smart spatial data explorers and modellers that can function without being told, in advance, what to look for.

AI technology hold some promise and there are various potentially interesting ways forward based on neurocomputing and artificial life, as well as other forms of machine learning (Openshaw, 1992). However, in many ways these methods still need to be developed. Here it is more appropriate to develop a more general specification of the sort of GIS relevant technology that meet the design criteria that have been discussed. Table 5 outlined a technical specification which the new tools should try to meet.

A final aspect is a key need is the ability to create a means of building into the spatial analysis technology human knowledge, human skills, and human intuition. No automated spatial analysis technology will develop far by itself, but equally it does not make sense to cripple the new methods by expecting all the critical and hard decisions to be made by humans. There is need for a conceptual view of exploratory analysis and modelling in GIS that can combine the strengths of both the human being and the power of AI based machine learning technology, whilst simultaneously attempting to avoid their weaknesses. The ultimate endstate is the development of smart spatial analysis technology, based on the best of statistical based tools, on new AI novelties, and incorporate human knowledge and reasoning. The whole system should then be controlled by a feed-back loop that can learn. Clearly such a system cannot yet be built but may be we can reach an intermediate state by simply indulging in multiple-method analysis. Instead of using only one method, K different methods can be used and the best or safest selected (Openshaw, 1992b). Old methods are used as benchmarks or challengers against which newer methods have to compete. The emphasis now changes to evaluation and choice from amongst alternatives. Hopefully, with a little luck; good, safe results will be produced. The concept of a general purpose spatial analysis machine is, perhaps, something that needs to be revived.

References

Anselin, L., 1989, What is special about spatial data?', NCGIA Technical Report.

Cliff, A., Ord, K., 1969, Spatial Autocorrelation, Pion, London

Ebdon, D., 1977, Statistics in Geography, Blackwell, Oxford.

Fotherham, S., Rogerson, P., 1993, 'Letter to the editor', Environment and Planning A, 25

Gregory, S., 1963, Statistical methods and the geographer, Longmans, London.

Griffith, D. A., Anselin, C. G., 1991 Statistical Analysis for Geographers. Prentice Hall, New Jersey

Griffith, D. A., 1993, "Which spatial statistical technicques should be converted to GIS functions?" in M. M. Fischer, P. Nijkamp (eds) GIS, Spatial Modelling and Policy Evaluation Springer Verlag, Berlin p103-114

Haining, R., 1992, Spatial data analysis in the social and environmental sciences Cambridge University Press, London

Openshaw, S., 1989, 'Learning to live with errors in spatial databases', in M.F. Goodchild and S.Gopal The Accuracy of Spatial Databases. Taylor and Francis, London

Openshaw, S., 1990, Automating the search for cancer clusters: a review of problems, progress and opportunities, in R. W. Thomas (ed) Spatial Epidemiology, Pion, London p48-78

Openshaw, S., 1991a, 'A spatial analysis research agenda', in I. Masser and M. Blakemore (eds) Handling Geographical Information: methodology and potential applications Longmans, London, p18-37

Openshaw, S., 1991b, 'Developing appropriate spatial analysis methods for GIS', in D. J. Maguire, M. F. Goodchild, D. W. Rhind (eds) Geographical Information Systems: principles and applications Longmans, London, p389-402

Openshaw, S., 1992, 'Some suggestions concerning the development of AI tools for spatial analysis and modelling in GIS', Annals of Regional Science 26, p35-51

Openshaw, 1992b, 'A review of the opportunities and problems in applying neurocomputing methods to marketing applications', Journal of Targeting, Measurement and Analysis for Marketing, 1, 170-186.

Openshaw, S., 1993, 'Two exploratory space-time attribute pattern analysers relevant to GIS', in S. Fotheringham (ed) Spatial Analysis and GIS, Taylor and Frances (forthcoming)

Taylor, P.J., 1977, Quantitative methods in geography, Houghton Mifflin Co, Boston.

Produced By
School of Geography
University of Leeds
Leeds LS2 9JT
From Whom Copies May Be Ordered

July 1994