tion and their change detection are the fundamental topics. Although they have been studied for over half a century, rooms are still left for improvement. *Pu and Gong* introduce a new approach based on gray systems theory to predict land cover change from historical aerial photographs.

VI. EPILOGUE

In this editorial, we feel very inconvenient to write down the full names of "geographical information science" and "geographical information systems" in most cases in order not to mix them up. Since "GIS" has been booked to refer to "geographical information systems", it might be a good idea to adopt another word for "geographical information science".

Looking at the development and the context of geographical information science, it may be regarded as a discipline which integrates theories, techniques and methodologies in spatial information processing. From this point of view, the term *Geoinformatics* might be more appropriate. This word consists of "geo-" and "informatics". The latter comes from German "Informatik" which refers to a discipline that studies the coding, transmission, process, retrieval and utilisation of information. Therefore, it is quite natural to refer "Geoinformtics" to geographical information science. Indeed, "Geoinformtics" has already been adopted in the Netherlands, although in Germany this term is not adopted because of the strong objection from the "informatik" community.

At last, we wish that everyone, either researcher or practitioner, could find this issue informative, useful and worthy of reading.

REFERENCE

[1] Abler, R. F., 1987. The national Science Foundation National Center for Geographic Information and Analysis. *International Journal of Geographical Information Systems*, Vol.1, No.4, pp.303-326.

[2] Chrisman, N., 1988. The risk of software innovation: a case study of the Harvard Lab. *The American Cartographer*, Vol.13, No.3, pp.291-300.

[3] Coppock, J. T., 1962. Electronic data processing in geographical research. *Professional Geographer*, Vol.14, pp. 1-4.

[4] Coppock, J. T., 1988. The analogue to digital revolution: A view from an unreconstructed geographer. *The American Cartographer*, Vol.13, No.3, pp. 263-275.

[5] Department of Environment (DoE), 1987. *Handling Geographical Information*. HMSO, London. 208pp.

[6] Goodchild, M. F., 1990. Geographic information systems and cartography. *Cartography*, Vol. 19, No.1, pp. 1-13.

[7] Gore, Al, 1998. *The digital earth: understanding our planet in the 21st century*. www.regis.berkeley.edu/rhome/whatsnew/gore_digearth.html.

[8] Miller, C. and Laflamme, R.,1958. The digital terrain model – theory and applications. *Photogrammetric Engineering*, Vol.24, pp. 433-442.

[9] Morrison, J. L., 1991. The Organisational home for GIS in the scientific professional community. In: Maguire, D., Goodchild, M. and Rhind, D. (eds.), *Geographical Information Systems*, Volume 1: Principles. pp.91-100.

[10] Openshaw, S., 1987. An automated geographical information analysis system (Guest Editorial). *Environment and Planning A*, Vol.19, pp. 431-436.

[11] Openshaw, S., 1998. Some trends and future perspectives for spatial analysis in GIS. *Journal of Geographic Information Science*, (this issue).

[12] Rhind, D., 1988a. Personality as a factor in the development of a discipline: The example of computer-assisted cartography. *The American Cartographer*, Vol.13, No.3, pp. 277-289.

[13] Rhind, D., 1988b. A GIS research agenda. *International Journal of Geographical Information Systems*, Vol.2, No.1, pp.23-28.

[14] Sawyer, J. S., 1960. Geographical output from computers and the production of numerically forecast or analyzed synoptic charts. *Meteorological Magnize*, Vol.89, pp. 187-190.

[15] Tobler, W., 1959. Automation and cartography. *Geographical Review*, Vol.49, pp. 526-534.

[16] Tomlinson, R. F., 1967. *An Introduction to the Geographic Information System of the Canada Land Inventory*. Dept. of Forestry and Rural Development. Ottawa, Canada.

[17] Tomlinson, R. F., 1988. The impact of the transition from analogue to digital cartographic representation. *The American Cartographer*, Vol.13, No.3, pp. 249-261.

[18] Wright, D. J. and Goodchild, M. F., 1997, Data from the Deep: Implications for GIS Community. *International Journal of Geographical Information Science*. Vol.11, No. 6, pp. 523-528.

Some Trends and Future Perspectives for Spatial Analysis in GIS

Stan Openshaw

Centre for Computational Geography
School of Geography, University of Leeds
Leeds LS2 9JT UK

Abstract
This paper reviews the development of GIS and spatial analysis techniques and provides perspectives on spatial analysis in GIS.

I. BACKGROUND

GIS is now a global big business that continues to experience rapid growth. Probably over a million workers now use GIS on a daily basis and well over quarter of a million systems have probably been distributed; indeed ESRI claim that over 100,000 copies of Arc/View have been sold. At the same time new GIS data continues to be created at a rapid rate; e.g. as digital mapping expands to cover the globe and the new generations of space borne ex-spy satellite commercial technologies create vast new remote sensing databases. Additionally, the computerisation of all the world's administrative systems, of retailing, of financial services, of travel, etc are all nearly complete. The IT State is here and nearly all these data bases are capable of being geographically referenced and an increasing number are becoming, or are already, geographic information in the broadest sense. Inexorably, continuously and at an ever increasing speed, this data collection, integration, cross-referencing, and storing process is rapidly expanding to cover more and more aspects of modern life. Within 1 to 10 years (depending on which country is being considered) there will be complete GIS coverage with sub 1 metre accuracy of most physical objects shown on maps or visible on the Earth's surface (e.g. houses, roads, trees, etc) and a large amount of human activity related things (e.g. addresses, movements, spatial patterns of behaviour etc). Commercial business (and governments) in every country are expanding their data warehouses and dart marts, as they realise that their future viability depends on them storing and then using more of the data they are almost "forced" to gather (because of IT) in order to optimise their future profits, competitiveness, or just to improve efficiency. A major driving force is the need to be seen to making good use of the available technology to avoid being negligent and open to legal challenge. For example, in the UK the Crime and Disorder Act (1998) placed a statutory obligation on all Local Authorities to perform crime pattern analysis at a time when virtually none was being performed. This will force a rapid integration of Local Authority corporate GIS systems (if there is one and most probably do not at a District level) with Police databases.

Nearly all of the new data bases being captured and created by developments in IT, probably about 85%, are, or have the prospect of becoming, geographic information. The world really is a GIS world! It is a pity, therefore, that there are some key technologies missing from the currently available GIS tool-kits that may well have a major impact on what GISs are, or could be, used for in the immediate future.

The purpose of creating immensely detailed, comprehensive, accurate and multi-level high resolution GIS databases is to allow users to do useful things with it. You do not create GIS databases merely because you like doing it! Nor do you develop such data bases so that you can ring fence them with copyright and ownership barriers to wider use and access. Nor do you imagine that the principal analysis tool is mapping! Yet it is at this stage that the GIS world's collective imagination seems to desert many of its users. Of course you can analyse point referenced data distributions by drawing maps. You can plan investments in retail networks using potential surfaces and a spreadsheet but there are far more powerful methods around than these. Some think that the answer is a statistical package. This may, sometimes, occasionally be helpful beyond a most basic data description function, but do not rely on it to do anything particularly useful. Imagine a situation in which there is a few hundred observations and a few variables. This presents no problem. You

1082-4006/98/0401~2-5$3.00

can use any one of a dozen different statistical packages, some with linked map and graph displays. Software abounds that is able to help you visualise, plot, and analyse such data. Very picturesque! Now increase the number of observations to 100,000 and, or, the number of variables to 100 or more. How do you now extract value from your investment in the geodata? Well, you could map some of it. You could reduce the 100 variables to 7. You could re-aggregate the data to 20 zones. However, none of these data reduction devices is particularly useful because it negates the theoretical but intangible benefits in having all the data available. There is a danger that you are merely throwing away data so that some pre-GIS legacy technology can be applied, usually misapplied, to spatial data that it was always unsuitable for. What a pity. What a waste!

Some defend this behaviour by arguing that anything else, such as a data driven approach, illustrates the dangers of being data led. So why not ask them what the alternative is? Where are all the hypotheses that would be needed to fuel a deductive approach going to come from when knowledge of the underlying processes that created the data simply does not exist (as yet)? There is no theory (a common criticism of GIS projects) but there really is no theory because data provision has often far outstripped theoretical understanding of the human and physical systems that the data relate to. At the same time many current attempts by human geographers and social scientists to understand how these systems work proceed in a data free manner because of fundamental self-induced philosophical problems associated with the concept of data as "fact", or perhaps due to various often misguided concerns over using GIS technology originally funded, it is alleged, for supposedly "evil" military purposes, or because of the failure of GIS to handle all aspects of place (including those that cannot be measured). However, even if greater scientific attempts were made to improve understanding, progress would still be very slow. Conventional normal science appears not to work well, or at all, when the systems generating the data are complex, chaotic, non-linear, and not well understood. So what are we to do? Well what we have most of, is data. You can revert to drawing maps of data but maps do not take you very far. Mapping data is not analysis. "Just map it" is not sufficient if you want to go beyond data display. At the same time data display is unsafe but because what you see in a data display is not objective and can be manipulated and most probably has been. The solution is surely to look towards exploratory spatial analysis and modelling methods to help you cope with this data rich and theory poor situation. Unfortunately only the most basic of spatial data exploration methods have so far been developed, partly it seems because of a universal fear of being data driven. The real longer term answer has to be geographical data mining but this word has only recently been invented and much of the technology still needs to be created. The success of GIS has created a data rich world that most of its users do not fully use or know how best to exploit.

Indeed, where are the geographical data miners? Where are the vibrant spatial statisticians and quantitative geographers actively enjoying the immense new opportunities that the GIS data riches have created for them? Are they all still fighting inter-disciplinary battles from 20 years ago? Or have they all retired or switched to other themes? One underlying difficulty here is the need to re-think how best to perform spatial analysis. It is gradually becoming apparent that there probably never can be a statistical theory based approach to spatial multivariate analysis relevant to GIS. Clever people have tried but have not yet succeeded. In any case is it relevant? Should we not be re-focusing our attention on the problems of performing appropriate and relevant spatial analysis in GIS environments. Instead more attention has been given by geographers in abstract attempts to understand the social implications of geographical technologies such as GIS than has been given to improving the geographical analysis technologies in GIS; see for examples, Pickles (1995), Curry (1998). A growing research obsession with the inadequacies of the social theoretical framework embedded implicitly or explicitly in GIS and, or its users, is continuing to distract geographical attention from how to improve GIS in relation to its ethical operation and its ability to better represent people and places by trying to demonstrate that you cannot have a GIS and that what exists has failed! Openshaw (1997 ab) offers a different view. Nevertheless, the lack of theoretical attention and the highly empirical nature of GIS activities have caused a lack of interest amongst the very people who have the statistical skills needed to develop the new geographical analysis technologies. Few appear to want to become involved. Why? Maybe the only answer will be to train a new generation of geographers in the limited range of core statistical, computer science and AI skills needed to make progress in this area. The growing interest in GeoComputation is one of the few recent developments that suggest that the future may be more exciting than the last decade or so has proved to be; see Longley et al (1998).

II. A SPATIAL DATA EXPLOSION

Bang! Or is it Boom?

At the present time throughout the world there is a vast explosion of geographically referenced data due to developments in IT, the computerisation of everything, the rapidly falling costs of computer storage, dramatic changes in the price-performance of all aspects of computing, and the global diffusion and success of GIS. There is a vast and rapidly growing geocyberspace of information of all types (Openshaw,1994). The problem is data and there are lots of it! Everyday there are more spatial data than previously. More data, and more data have resulted in the development of data libraries, data warehouses, operational data stores, data dumps, and data archives. However, data by itself is only a raw material. If you do not evolve the data using technologies then having access to more and more data, more bandwidth, and more powerful hardware will not help at all, indeed it will probably make things worse as more and more time and resources are expended on capturing, storing, managing, and archiving data that will never be fully or properly used because of the absence of tools that can extract "value" from it. The prevailing wisdom is that data are now the raw materials of the IT age from which new knowledge creating industries will one day extract new insights, make new scientific discoveries, and justify the immense cost of having spatial data. The problem at present is that these new knowledge creating technologies are somewhat rudimentary. Yet without these developments the end result is potentially disastrous, as a computerised bureaucracy with computer based management systems covering most areas of modern life struggles to store and create more and more data most of which will probably never be properly processed or used. Indeed at present probably at least 95% of all GIS databases are not being exposed to state of the art analysis and modelling technologies and many are hardly used at all.

Spatial Analysis Crime

Openshaw (1993) coined the term "spatial analysis crime" to describe users of GIS who have successfully created databases relating to all kinds of potentially useful information but who then fail to analyse it, or permits others to do, or hide the data behind copyright and cost barriers that effectively prevent analysis even in the public good. It is surely a disgrace that before the public can be assured that emergency services have the fastest possible response, they must first find millions of pounds to buy access to data that they have already paid millions of pounds to have created in the first place! Spatial Analysis Crime is also a consequence of the success of GIS in creating spatial databases and a widespread failure by users to realise that having access to a GIS is not sufficient because there are fundamental gaps in the GIS tool-kits. Spatial analysis crime is also very expensive but the costs of a failure to fully use geodata falls on third parties and is, as yet, not fully recognised. There are probably many lost opportunities for spatial analysis some of which have costs attached. For example, what is the "cost" to society of failing to spot a geographical cluster in a preventable or treatable or pollution related disease?

The EC probably loses billions of ECUs each year because of inefficient spatial planning policies and fund allocation strategies due to the use of unfair, distortive, and inefficient zoning systems used to calculate indicators of need. Most National Governments lose millions of ECUs each year due to poor analysis of spatial information that is used (and not used) in their spending plans. Likewise many companies run at reduced profit levels due to their inefficient use of their geographic information holdings to properly plan the geography of their business or enterprise.

People die each year because no one bothers to properly analyse disease or traffic accident data or crime data or illness data for unusual localised concentrations that may be preventable once known to exist, predictable and thus avoidable. Criminals escape detection because no one as yet bothers to properly analyse the real-time on-line crime data bases that now exist in many countries. International GIS vendors sell, systems they know cannot adequately analyse the data they manage so well. Vendor GIS Slogans such as "just map it" reflect this neglect of more in depth analysis to explore the patterns, relationships, and processes hidden in the mappable data. The vendors often claim (wrongly) that users are not interested in geographical analysis, or if they are, then the vendors (either accidentally or deliberately) confuse the end-users by having spatial analysis modules that equate spatial analysis with little more than polygon manipulation, overlay, and buffering. By way of example note how MapInfo defines spatial analysis: "An operation that examines data with the intent to extract or create new data that fulfils some required condition or conditions. It includes such GIS functions as polygon overlay or buffer generation and concepts of contains, intersects, within or adjacent." (Page 396, MapInfo Professional: User Guide, 1995). Yet this is not a definition of spatial analysis any quantitative geographer or spatial statistician would accept as being useful.

History of GIS (in UK)

Sadly many of these analysis problems are largely unavoidable and reflect the historical origins and the global development path of GIS. In the UK a number of GIS development phases can be identified.

Phase 1:

The pre-Revolution research (before-1985) when GIS was of interest only to researchers.

Phase 2:

Getting the technology accepted via the GIS revolution (1986-1991). The GIS revolution years witnessed an explosion of interest as early software packages introduced GIS technology into AM/FM operational environments. Virtually overnight the paper map making and handling industry went digital or put in place long term investment plans to become digital. A major common concern here was the completion of a national large scale digital mapping database as a necessary first step, although of course this is only really relevant to a minority of GIS users.

Phase 3:

Consolidation and new operational concerns such as standards. Where are the data and how can the national mapping agencies assist? (1992-5). The revolution is now over and serious business planning has replaced the original hype and euphoria. Of particular concern is the cost of digital mappings, with copyright and ownership issues coming to the fore. Sadly the notion of free publicly available geoinformation is regarded as an alien concept as the State's Agencies re-sell their publicly paid for data to the public who originally paid for it. This slows down some key developments but will not in the long term stop them.

Phase 4:

Maturity sets in with serious strategic thinking about the management of data leading to National Geospatial DataFrameworks (NGDF) (1996 onwards). A single universal digital mapping framework is a good idea if it is used by all governmental agencies and all the users, and if it is affordable, up-to-date, flexible, extendible, and as accurate as any of the newer alternatives that may be developed in the next few years. Data libraries are another common manifestation of this phase of GIS.

Phase 5:

In the next millennium another major change in spatial data provision will occur as new space based automated digital mapping replace the terrestrial based manual surveyed databases. It is all a matter of cost, flexibility, access, and convenience. It will also be characterised by web based GIS solutions. However, whereas the means of performing GIS actions are constantly being improved, the nature of the tasks or functions being performed have changed only a little since the late 1980s.

All very interesting but where is the Spatial Analysis and Modelling in all this activity? In what phase of GIS development will it appear in? Sadly, it will not! GIS is all about data but it should also be about data using technologies that go beyond data collection, data ownership, standards mapping, and trivial map based analysis! Yet the whole history of GIS has been dominated by the worries, the concerns and the fears, over digital map data. There is a massive global data-phobic problem that has blinkered the world of GIS. As a result GIS has been very successful at creating large spatial databases but very poor at offering the new analysis and modelling tools designed to allow users to exploit the data that GIS has created. So there is a missing revolution in GIS! A second GIS Revolution is now needed concerned with full and good use of the GIS databases but this is still sometime off. If this is to occur at all then the users must do it themselves by creating their own GIS research agendas; see for example Openshaw and Fischer (1995).

Trends in GIS

The Revolution is over but GIS technology is not yet static. There are some obvious trends in the further development of GIS. The market place is still expanding as many new applications open up in areas far removed from traditional AM/FM concerns. Likewise many less GIS advanced countries appear to be copying the more GIS advanced countries such as the USA/UK by importing their technologies and their consultants, and hence, no doubt, will probably re-discover or re-invent many of the same problems. Meanwhile data access costs is a major short-term obstacle to many new GIS applications outside of the more traditional domains but also many new sources of spatial information are emerging; viz. high resolution satellite imagery and orthophotography which will probably revolutionise digital map making and, before too long, create new and cheaper alternative sources of geoinformation. For example, if you are using raster maps as a backcloth then orthophotographic images might well be a far better, maybe cheaper, alternative than rasterised traditional maps. On the other hand, the optimal solution would be to have and use both!

There is also some evidence of a broadening of the GIS research agenda. Social concerns and ethical considerations need to be given more serious attention in the future; see for example Pickles (1995) and Openshaw (1997a,b). Also major attention is being given to the use of GIS to democratise spatial decision-making by providing Internet access to both data and GIS tools. There will be an explosion in the use of www to broaden public access to GIS data and systems as means of publicity, awareness raising, propaganda, manipulation, and conversion of public attitudes. The numbers of domestic users will continue to rapidly increase as networked systems rapidly expand in the new millennium aided by a move towards more open government and commerce. Maybe the new millennium will also witness what some term GIS_2.

Other trends in GIS are more technical with a continued focus on object orientated GIS although the potential user benefits are probably far away, the appearance of alternative digital mapping sources in competition with national mapping agencies, and the likely re-expression of GIS tool-kit in Java format with the appearance of distributed GIS systems. In short more of the same with major and fast evolutionary improvements to the established technologies. The historic spatial analytical deficiencies will remain due to vendor neglect and an apparent lack of end user demands for it. Maybe the only real hope is to develop separate systems that merely treat GIS as data front ends. Maybe a new spatial analysis industry needs to be created outside of the world of GIS.

III. PERSPECTIVES ON SPATIAL ANALYSIS IN GIS

Trends

It is argued, therefore, that there is a continued lack of the GIS vendors who are treating real spatial analysis seriously and a continued failure of end-users to realise what they maybe missing because of a lack of pre-existing practices or apparent need or awareness of what may now be possible with their data. GIS was so successful essentially because it was an analogue map replacement technology. The cost-benefit analysis was easy to perform. It did not offer much that was not already being done, merely a far cheaper, faster, and more efficient way of doing it. There is probably little in GIS today that could not be done (albeit badly and at huge expense) with a pencil, tracing paper, a map, and a piece of string! Spatial analysis is different because there was often no previously existing technology to replace. It is something new and different that can only be done now because of the existence of GIS and far faster computing. These new needs do not become apparent until you have a fully developed and mature GIS industry and much geodata. This is a new situation for which relatively little suitable pre-existing technology exists. Most quantitative geographical methods were developed in the 1960s and 1970s when spatial data were rare, datasets were very small by modern standards, and computers a million or more times slower. Most of the old methods do not scale well in that much more is involved than merely increasing the size of arrays to handle larger number of cases and variables. It is not so much a matter of adding more computer memory as it is of adding new functionality.

Additionally, hard but critical questions probably need to be quantified before the economic case for spatial analysis can be made; for example, if a detailed geographic analysis of traffic accidents results in changes that save K lives, how much was that worth? The costs of prevention fall on the transport planner but the benefits are evident elsewhere. Maybe the only argument that will succeed in this situation is that of public or political imperative. GIS opens up the prospect of detailed spatial analysis of data previously neglected, but the case for continued or further neglect is now much weaker. Spatial analysis criminals are failing to perform their professional duties and are being negligent. Maybe in the future fear of litigation accompanied by awareness raising may help change the status quo.

GIS vendors are not research orientated which is another reason for their reluctance to do much in this area of GIS. The continued and ongoing heavy focus on all data related matters constantly postpones the recognition of the need for more sophisticated analysis and modelling in GIS environments. All this is very depressing especially for GIS skilled geographers, yet maybe the principal problem is due to their own neglect of quantitative methods suitable for spatial data to the extent that there is now an almost complete absence of suitable Geographical Analysis technology. So in some ways the failure of GIS to offer much real spatial analysis reflects the failure of geographers to develop any and the lack of end-users with a geographical analysis culture. No wonder there is little or no sensible and GIS appropriate spatial analysis technology in place within nearly all the leading proprietary GIS systems. Many end-users are unaware of what is now possible beyond what is provided by leading GIS packages. The world's few remaining spatial analysts have failed to raise levels of awareness of what is now possible by providing significant generic examples and illustrations relevant to end user contexts. As a result vendors can quite accurately argue there is no market, that there are not lots of users demanding spatial analysis functionality, it is all too complex anyway and lies far beyond the digital cartographic origins of GIS. Unless geographers do something to demonstrate the contrary then many of the problems may remain unresolved.

Design objectives

The key questions then are 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 real users?
- Adds value to the GIS investment?
- Promises tangible and significant short-term or instant benefits?
- Safe and reliable?
- Capable of being understood by non-experts?
- Easily applied?

These GISability criteria are very important if practically useful spatial analysis methods suitable for use in GIS are going to be developed (Openshaw and Fischer, 1995). It is not helpful developing sophisticated spatial analysis research tools that users simply cannot understand or use. For instance, many of the recent developments by spatial statisticians in the area of epidemiology cannot be used by non-statisticians. Now it could be argued that the original GIS technology required considerable computational geometry skills, yet this complexity is now totally hidden from the users. The same philosophy needs to be applied to spatial analysis tools. Simple minded non statistical end-users do not require simple methods as distinct from simple to use interfaces to appropriate technology, no matter how complex it may be.

The Spatial Analysis situation in the late 1990s is little improved on that of 10 years previously! Most existing spatial analysis methods are either not useful or usable in GIS environments. The problem is deeply rooted and needs urgent and dramatic action. So throw away many of the old statistical books that assume random sampling of non-spatial data. Junk most of the quantitative geographic methods that assume hardly any data or use inappropriate and unsafe inferential significance tests. Beware of the more heavily duty statistical methods that rely on mathematical proofs which are at the expense of empirical performance and geographical common-sense, and then re-consider the previous GISability criteria.

Additionally, it is also clear that many users have a requirement for both data efficient and user-friendly exploratory spatial data analysis methods capable of being safely used by people who do not have higher degrees in the statistical or spatial sciences. User Friendly Spatial Analysis should seek to provide analysis that users need, that are easy to perform, highly automated making it fast and efficient, readily understood, offer results that are self-evident and can be communicated to non-experts, and are safe and trustworthy. Sadly few methods meet many of these criteria and those that do are not yet generally available.

Map based methods

The problem is essentially that geographers (and others) are still not very good at the geographic analysis of spatial data. Virtually the only widely used uniquely geographical analysis tool is the MAP. Yet it is well known that map based visualisation and analysis is a simple but fundamentally flawed technology. The Modifiable Areal Unit Problem (MAUP) is an added complication. If the interest is in map based methods of spatial analysis then an obvious approach is to improve the map as an analytical device. Instead of accepting zones as fixed objects, manipulate them to optimise something relevant to the analysis being performed (Openshaw and Alvanides, 1998). Use the MAUP to allow the zoning system to function as a visualisation device (Openshaw, 1996). This process is no longer hard, it just requires a degree of imagination, a few clever algorithms, a fast computer, and a GIS. All these components exist in abundance, other than, perhaps, the imagination needed to develop the new geographical analysis systems

It is worth noting that the success of GIS has been accompanied by two other major developments: the appearance of useful Artificial Intelligence tools (Openshaw and Openshaw, 1997) and a most dramatic speed-up in the performance of computers. High performance computers are now an estimated 1000 to 10000 times faster since the GIS revolution, whilst parallel supercomputers offer over 5,000 times the performance of a top end PC (Openshaw and Turton, 1998). Additionally, there is also some evidence of an emergence of a GeoComputational paradigm in geography. GeoComputation is defined as the bringing together of spatial databases, high performance computers, and AI-computational intelligence technologies to solve practical problems. The first International Conference was in Leeds in 1996, in 1997 New Zealand, in 1998 in Bristol. GeoComputation is useful mainly because it offers hope of a new approach to scientific study in GIS, it provides a focus for new research, it emphasises new computationally intelligent technologies, it is applied and it provides a future paradigm for spatial analysis and modelling in GIS. The challenge now is to do it and build the new tools to make good use of GIS databases. One way forwards is to create highly automated geographical analysis machines able to perform basic spatial data exploratory functions.

IV. AN ERA OF GEOGRAPHICAL ANALYSIS MACHINES

The first GAM

In a spatial data rich GIS world perhaps the most basic of all the outstanding needs is for automated spatial pattern description. It is a very fundamental first step that has been widely neglected. If this is not possible then you can probably forget about anything more sophisticated. This is clearly not a problem that humans can do well by eyeballing mapped data especially if the data sets are large and, or, the patterns not so clear cut as to be instantly obvious. A solution is to invent a "machine" to do it for us. So there is a generic need for an exploratory spatial data analysis that is automated, efficient, effective, trustworthy, and easy to use!

The Mark 1 prototype Geographical Analysis Machine (Openshaw, et al, 1987; Openshaw, 1998) was an early attempt at automated exploratory spatial data analysis that was easy to understand. It attempted to answer a very simple question:

> "given some X,Y point referenced data of something interesting WHERE might there be evidence of localised clustering if you do not have the foggiest idea of where to look due to lack of knowledge?"

The first GAM/1 was developed in the mid 1980s. It worked but this technology was never generally available. A GAM is unavoidably very computationally intensive. The term machine was quite appropriate because it really needed a dedicated computing machine. Early runs took over 1 month of CPU time on a large Mainframe (Amdahl 580). Later versions ran on a Cray X-MP, Y-MP, and Cray 2 supercomputers. The GAM/1 was developed to analyse child leukaemia data in Northern England. It easily spotted the suspected Sellafield Cluster but it also found an even stronger major new cluster: the Gateshead cancer cluster in 1986. This is possibly the only instance of a major cancer cluster being found by computer analysis. However, GAM/1 was a mixed blessing! It was praised by many geographers as a major development in useful spatial analysis technology but it was severely criticised by some statisticians (partly due to their ignorance of the geography of the problem). Additionally, the software for GAM was never distributed as ten years ago it was not easily run as most computers were not fast enough.

With time some of the problems went away. The International Agency for Research on Cancer (IARC) commissioned a study in 1989-91 of several clustering methods, many developed by critics of the original GAM; see Alexander and Boyle (1996). Some 50 synthetic cancer data sets were created for which the degree of clustering and locations of clusters were known but kept secret. The data were given to the participants who performed their analyses without any knowledge of the correct results. Nine different methods were investigated. It was anticipated that the statistical methods preferred by the critics of GAM would work best but much to the surprise of Alexander and Boyle GAM was shown to be the best or equivalent best means of TESTING FOR PRESENCE OF CLUSTERING and for FINDING THE LOCATIONS OF CLUSTERS. Alexander and Boyle (1996) concluded:

> "The GAM has potential applications in this area <u>if adequate computer resources are available</u>. At the present time, however, the new, more sophisticated version of the <u>GAM is complex, difficult to understand.</u>" (p157).

That was in 1991 (not 1996). Are these criticisms still valid today?

Reviving GAM/K

GAM/K still runs on the later day version of the Cray X-MP vector supercomputer (the Cray J90). Subsequent modifications to the spatial data retrieval algorithm used in GAM reduced a typical run time to a few minutes on a workstation. GAM/K is now a practical tool that can be readily linked to any GIS and it no longer needed a supercomputer to run it.

GAM has several strengths: it is a totally automated analysis, it can find multiple clusters in a single run, it is fast, the results are largely self-evident, and the method is very easy to explain in plain English without statistical complexity. Its deficiencies relate to the nature of its brute force search, it is dumb, it does not scale to higher dimensional data spaces, it is compute intensive, the mapped results may still mislead, and it is not available in any GIS (yet). A www interfaced version is now available in a platform independent manner. The www GAM is a virtual Geographical Analysis Machine and maybe it establishes a new style of performing geographical analysis. The technology works and it is both user and GIS friendly. It is also GIS independent and will soon be platform independent too. Further information and an evaluation version is available on the author's www home page; see also Openshaw et al (1999).

The other GAMs

The GAM style of exploratory spatial analysis has been, and is being, developed further. MAP EXplorer (MAPEX) is an intelligent search version of GAM/K (Openshaw and Perree, 1996). It uses a Genetic Algorithm to perform a search for clustering and then AVS to create MPEG computer movies of the search process. The hope is expressed that allowing the user to watch the search for clustering provides an understanding of both the method and also of the nature of any patterns that may exist in the data being analysed. Certainly the behaviour of MAPEX for random data is visibly very different from clustered data. Hence by watching computer movies of the search process the user may well develop an artistic feeling as to what any patterns are saying, particularly their strength and distribution. The results are readily communicated (via the MPEG movies which can be viewed on a PC). A library of MAPEX movies for known patterns can be provided in order that users can "calibrate" their artistic pattern detection senses on known results. It seems to work but needs to be tested in operational settings. The AVS "bottleneck" can now be avoided by using Java to create new animation tools. This approach has been developed to allow interactive animation with zoom facilities.

The Space Time Attribute Creature (STAC) described in Openshaw (1995) extends the MAPEX approach to multiple data domains. In many ways STAC is the ultimate geographical database pattern miner. It operates in the three principal data domains (geographic, temporal, attributes) of GIS, each of which can be multivariate and all of which are measured in different units (e.g. what is the temporal equivalent of 1.2 km?). The positions of the STACs during their data pattern hunting safari can be viewed by projecting their geocyberspace co-ordinates onto a map base. The map reigns supreme here because probably nothing else matters in geographical data mining. There are two hard parts to building STACs that work; the mechanisms for performing a smart search and the statistical pattern detection measures. Research is currently underway at Leeds to develop a viable portable technology that can be used in the same way as the GAM; see

http://www.ccg.leeds.ac.uk/smart/intro.html

for further details.

V. CONCLUSIONS

Some trends are quite obvious. Seemingly various national and EU wide GIS data sharing and standardisation strategies will eventually converge and be recognised, probably based on US standards and implicitly enforced by US agents disguised as GIS vendors. There is a strong future trend in GIS for distributed systems running across heterogeneous machines linked over the Internet. Likewise it is possible that many GIS systems will be re-expressed in Java in the next few years mainly because it is a trendy thing to do and platform independence is worthwhile. At the same time the only advances likely in spatial analysis will be those that occur within and emerge from various research environments. Maybe the availability of practical AI and computational intelligence tools will eventually help overcome many of the historic problems. Likewise the emergence of a GeoComputation paradigm may resolve acceptability issues and may even become fashionable in the new millennium. Meanwhile developments in Java and heterogeneous GIS will undoubtedly create new opportunities for distributed spatial analysis machines on the Internet in the near future. Would it not be nice if users could send standard spatial analysis tasks to a particular www site and receive the results back a day or two later. Would it matter much if the analysis was performed on the world's fastest civilian parallel supercomputer located in Tokyo, other than that the results are now available. The concept of "sending data away" for laboratory analysis is not new, maybe the same will soon be true for spatial analysis. This is possible because many spatial analysis needs are generic, data and context independent, and few in number. The vision is, therefore, of what has been called GIS component ware. A number of virtual spatial analysis and modelling machines will soon exist, accessed over the internet, each of which offers and performs only one type of widely applicable analysis. Much of the needed technology exists, it is just not yet being deployed in a suitable manner. A start has been made but much remains to be done. Probably the hardest part is accepting the need for exploratory spatial analysis in GIS rich situations to be performed by non-expert users and then developing good methods that meet these needs that are useful and usable by the end-users of GIS. This task is urgent. The spatial analysis and geographical data mining flood gates are about to open and the internet provides an obvious channel for meeting many of these needs.

REFERENCES

[1] Alexander, F. E. and Boyle, P., *1996* Methods for investigating localised Clustering of Disease, IARC Scientific Publications No 135, Lyon, France

[2] Curry, M R, 1998, *Digital Places: Living with Geographic Information Technologies*, Routledge, New York

[3] Longley, P.A., Brooks, S.M., Mcdonnell, R and Macmillan, B, 1998, *Geocomputation: a primer*, Wiley, Chichester

[4] Openshaw, S., Charlton, M., Wymer, C. and Craft, A.W., 1987, 'A mark I geographical analysis machine for the automated analysis of point data sets', *Int. J. GIS*, 1:335-358.

[5] Openshaw, S., 1994, 'Computational human geography: exploring the geocyberspace', *Leeds Review*, 37, 201-220.

[6] Openshaw, S., 1993, GIS crime and GIS criminality', *Environment and Planning A* 25:451-458.

[7] Openshaw, S., 1995 'Developing automated and smart spatial pattern exploration tools for geographical information systems applications', *The Statistician 44*:3-16.

[8] Openshaw, S. and Fischer, MM., 1995, 'A framework for research on spatial analysis relevant to geo-statistical information systems in Europe', *Geographical Systems*, 2:325-337.

[9] Openshaw, S. and PerrÈe, T., 1996, 'User centred intelligent spatial analysis of point data', in D. Parker (ed.) *Innovations in GIS 3* Taylor and Francis, London 119-134.

[10] Openshaw, S., 1996, Developing GIS relevant zone based spatial analysis methods. In P.Longley, M.Batty (Eds) *Spatial analysis: modelling in a GIS environment*. Cambridge, GeoInformation International: 55-73.

[11] Openshaw, S., 1997a, 'Towards a more computationally minded scientific human geography'. *Environment and Planning A*, 30:317-332.

[12] Openshaw, S., 1997b, 'The truth about ground truth', *Transactions in GIS*, 2:7-24.

[13] Openshaw, S. and Openshaw, C.A., 1997, *Artificial Intelligence in Geography* John Wiley, Chichester.

[14] Openshaw, S., 1998,' Building automated geographical analysis and explanation machines'. In P Longley, P.A., Brooks, S.M., Mcdonnell, R and Macmillan, B (eds.), *Geocomputation: a primer*, Wiley, Chichester. 95-116.

[15] Openshaw, S. and Alvanides, S., 1998, Applying GeoComputation to the analysis of spatial distributions. In P.Longley, M.F.Goodchild, D.J.Maguire, D.W.Rhind (Eds) *GIS: Principles, Techniques, Management and Applications* Wiley, New York, (in press).

[16] Openshaw, S., Turton, I., Macgill, J., 1999, 'Using the Geographical Analysis Machine to analyse census long term limiting illness data', *Geographical and Environmental Modelling*, (in press).

[17] Pickles, J., (ed), 1995, *Ground Truth: The social implications of Geographical Information Systems*. Guildford Press, New York.

[18] Turton, I., Openshaw, S., 1998 'High Performance Computing and geographical developments, issues, and case studies', *Environment and Planning A* 30:1839-1856.