COMPUTATIONAL HUMAN GEOGRAPHY: EXPLORING THE GEOCYBERSPACE

Stan Openshaw

WORKING PAPER 94/10

SCHOOL OF GEOGRAPHY • UNIVERSITY OF LEEDS

Introduction

It is observed that the world stage on which we do geography is entering a period of rapid change. An opportunity now presents itself to redress many of the problems of human geography and evolve the subject into a form more suitable for the 21st century. Some of the various fundamental problems that need to be addressed include a mix of the following: an overemphasis on sophisticated and elitist philosophical perspectives about which there is no consensus, a blatant disregard for the basic principles of science, a neglect of the increasingly information rich world about us, a pre-occupation with computer technologies that are incapable of yielding new concepts, and an apparent ignorance of the new computational tools that are potentially so applicable to the subject. It seems that those geographers best able to create new theories are essentially locked into qualitative approaches that rest more heavily on aspatial social science than on traditional geographic concerns, whilst those that are computer based hardly ever get beyond numeric map analysis. The challenge is to build new bridges and evolve a new (new) age of geography in which both the soft and the hard can be brought together within a common framework that is forward and outward rather than backward and inward looking.

The present is an ideal moment to consider the nature of this opportunity because of three profound developments of the greatest geographic significance: first, there is the Geographical Information Systems (GIS) revolution; second the availability of practical Artificial Intelligence (AI) tools that are intrinsically applicable to many areas in geography, and thirdly, the emerging new era of parallel supercomputing. In addition, it is becoming clear that the political, social economic, and cultural context in which we do geography has changed in a number of profoundly fundamental ways as the world moves into the era of the Information Technology (IT) States within the context of the global village (Openshaw, 1993).

The IT state is characterised by: large scale, benign, universal data capture covering most aspects of modern life, the computerisation of virtually all of the management and control systems on which societies and economies depend, the linkage of separate computer systems, the dissolution of technical obstacles to systems integration by the emphasis on open systems, and the increased reliance on computers at all levels and all scales for the continuation of life on earth. The problem is that most geographers have yet to notice many of these changes or appreciate their likely impacts on the subject. As a result there is a vast and growing gap between how we do geography in 1994, thinking that it is still really 1970 something, and how we could do it had it just been invented. It is becoming urgent to close the gap between geography as it is and how it can now be done.

2 The changing worlds' of geography

The world of geography until this century was the world of the map. Geography started as a "science" concerned with establishing "facts" about the earth. Cunningham (1559) quotes Ptolemy (1477) "Geographie is the imitation and description of the face, and picture of the earth ... to furnish topographic information to practical statesmen and scientists". Maps were seen as a source of visual pleasure and as objects to charm the senses (Skelton, 1952; 1958).The maps the early geographers left behind reflected the expansion of knowledge of the physical three dimensional world in which we live. Most non-geographers think that geography is still focused on this map world, a view that their school experiences of geography fostered. The Geographic Information Systems (GIS) revolution of the late 1980s, of course, re-establishes this link between geography and maps! A linkage that not all geographers are now comfortable about.

In the 20th century the domain of human geography has gradually changed from being focused primarily on the earth's surface to more abstract spaces. These include the regional descriptions of Vidal de la Blanche, the geometric transformations of map space of Christaller and Losch, and increasingly, various abstract mathematical and statistical space as geography underwent its quantitative and mathematical revolutions in the 1960s and 1970s. More recently, the view of human geography as a social science, has focused on different types of abstract social, cultural, gender and organisational spaces that have moved the focus of the subject far beyond the simple and visible map domain as attempts are made to understand process and behaviour of people related systems. The danger, of course, is that traditionally the "hard science" based mathematical models are too partial in their representation of these complex behaviour spaces, whilst the soft "social science" approaches are little more than plausible, descriptive, "fairy tales"; something that has been termed the Catherine Cookson approach to geography (Openshaw, 1991). The challenge for the future is how to link them together and indeed move some hard science technology into the social science arena of human geography.

The point to note here is that all these new worlds of 20th century geography ranging from the mathematical to the social can be regarded as different types of conceptual virtual reality. You cannot see or feel any of them directly. You have to imagine they exist and they exist mainly in your mind. They are all highly complex, making scientific understanding difficult. Progress has been slow in understanding their form and function, a task requiring high levels of mathematical or statistical or epistemological skills. Even though they are based mainly on theoretical speculation rather than information yet they are all an early type of what is termed here the geocyberspace.

The cyberspace concept started out in science fiction. Gibson (1986) visualised it as "A graphic representation of data abstracted from the banks of every computer in the human system. Unthinkable complexity. Lines of light ranged into the non-space of the mind, clusters and constellations of data. Like city lights receding" (p7). A decidely non-physical "matrix" with abstract geometries representing the relationships between data systems and the data themselves (Henderson, 1991; p90). Non too helpful definitions but it was always viewed as being much more than this. The idea of a cyberspace is so attractive because it conjures up images of the emerging invisible new world of information flowing along complex networks. It provides a kind of looking glass into a mathematical wonderland able to embrace the entire universe of information, a representation of the ultimate hyperdimensional information display (Sutherland, 1965). Its all about accessing vast quantities of online information and then learning how to live with it. Imagine falling into a database and then looking around. What would you see? You are now totally immersed in the artificial computer generated unreality of multimedia information flows. You look around and see three dimensional projections of infinitely more complex hyperspaces. You are in the cyberspace, an artificially generated, information based world of virtual reality. This is very relevant because it is becoming apparent that the emerging IT State is creating an immense cyberspace as an electronic nonspatial world of computer networks that link computers and databases, characterised by immense complexity, diversity and extra-territoriality. Those who discover first how to colonise and settle this new world, will clearly benefit most (Scientific American, 1991; Benedik, 1992).

The geocyberspace is that version of the cyberspace relevant to geography. A distinguishing feature of this form of artificial reality is the presence of either a map space or map dimensions somewhere amongst the many other non-map dimensions. If the early geocyberspaces were mental constructs, those of the 1990s and beyond are also information and computer based. It is worth noting that this geocyberspace has far more dimensions to it than a 3D virtual reality conceptual world. Also we do not just want to "see it" or "walk around in it", rather unhelpful tasks because of its hyperdimensional complexity. Instead what we desperately need is the technology to describe, model, and analyse the spatial patterns, processes, and relationships that undoubtedly exist in these artificial hyper-worlds that are collectively termed here the geocyberspace. Whilst the geographical explorers of the 19th and early 20th centuries explored the three dimensional realities of earth's surface, in the 21st century it will be the artificial worlds of the geocyberspaces created by the IT State that will be our tramping grounds. The need now is to discover how best to survive in this new world. There have been new geographies before as the subject has developed in response to new and different challenges. The challenge it now faces is far more fundamental to its future than many previous ones. To survive now, a New Geography has to evolve from the existing mix of new and old, good and bad geographies that have developed during the last 50 years. It is aided in this task by three very profound, geography relevant, technological developments.

3 Three key technological developments

3.1 The Geographic Information Systems (GIS)

The GIS revolution is in many countries a mid-to-late 1980s phenomenon. In the UK it dates from the Chorley Report of 1987 (Chorley, 1987) although the ideas are much older. The principal impact of GIS has been the conversion of paper maps into a digital form and the gradual incorporation of geography on many of the world's databases relating to people. The result is that, increasingly, virtually all the available data about people and place on our planet, including many aspects of its physical and human systems, have geographical referencing attached to them; or they will have before long. The IT State is creating via GIS an increasingly spatial data rich world.

The significance of GIS is mainly its possibly transitory role in adding geography to universal data on a global scale. It also, incidentally, re-emphasises the importance of the map dimension in geography and greatly increases the need for geographical analysis. GIS is a global technology and it will become increasingly difficult to think about doing geography at any scale without GIS being there somewhere as a platform on which to do it, or relate it. However, despite its importance, by itself, GIS of relatively little geographical value. The provision of more and more geographic data does not by itself automatically lead to any real increase in geographic knowledge. It is becoming very important to develop new tools that are able to extract and create new geographical knowledge in the form of theories and concepts from the geodata rich environments that GIS is creating for us. Using a GIS to produce multi-colour maps is not enough.

Geographers have still to discover how best to tackle the data rich but theory poor world of GIS. There is an increasingly urgent need to construct Geographical Knowledge creating Systems (GKS) to build on the spatial databases being provided via GIS. It is also important to seek to incorporate rather than ignore the theories and conceptual knowledge we already have when we seek to perform spatial analysis and modelling. We can become instantly smarter in many applications by making more intelligent use of the knowledge we already have rather than engaging in a purely technical exercise that ignores it. For example, improvements in the performance of spatial classifications have over the last 20 years been dependent purely on algorithmic developments; yet even the latest neurocomputing based classifier (Openshaw, 1994a) ignores 50 years or more worth of knowledge of socio-spatial structure. It would be far smarter to seek to incorporate rather than not use such knowledge. If geographers are to build GKS then they need to develop concepts rich rather than concepts poor or concepts free forms of analysis. It is important to explicitly seek to build bridges to and within the existing soft areas and themes of geography, because here there is an immensely rich conceptual resource base that can be used to endow otherwise dumb computer tools with an almost instant degree of intelligence. It short we can seek to become far smarter by developing methods able to use the knowledge we already possess purely by using it and building it into the algorithms and models we develop.

3.2 The Artificial Intelligence (AI) revolution

After a long and sometimes turbulent history lasting over 50 years, a number of practical, useful and geography applicable AI tools now exist. There is a view that AI may well become the next major "fashion" in geography, sometime during the mid to late 1990s. This belief rests on the feeling that much of geography contain applications that are AI suitable, that, some or many of the "hard" problems in geography may well become soluble using AI tools, and the AI community may well view geography as a rich source of opportunities for testing out their technologies. However, it should be clearly understood that the potential for AI in geography is not restricted to the traditional domains of quantitative geography but it extends to encompass many areas previously exempt. Indeed, it is probably in the soft areas of geography where the impact might be greatest as AI significant extends the applicability of computational methods.

AI is an attempt to build "thinking machines" by writing software able to do things that might be regarded as intelligent if they were done by people. In seeking geographical applications for AI we are not so much interested in replicating the behaviour of human experts than in creating super-human levels of intelligence in narrowly focused domains of interest. With this in mind, it is clear that there are three principal areas of AI that are particularly relevant to geography: neurocomputing; genetic algorithms which includes evolutionary programming and artificial life; and computer vision.

Neurocomputing has become a major applied scientific growth area since 1986 with applications being demonstrated in many different disciplines. As yet there are relatively few applications in geography but probably not for long as relevant software is widely diffused. Supervised neural nets provide a basis for modelling spatial processes without worrying about the complexity of the problem. In fact, it is now possible to build computer models for virtually any process or phenomenon for which there is relevant data. The technology is able to handle noisy data, as well as discontinuous and highly nonlinear high dimensional functional spaces and fuzzy relationships. Maybe neurocomputing tools should be kept for those problems for which there is no other alternative or where there is concern that better levels of performance might be obtained. Supervised nets have the potential to revive both traditional modelling activities and, much more significantly, to dramatically extend computational tools and the modelling domain into softer areas. It is also important not to neglect the value of non-supervised neural nets as they provide more flexible, noise tolerant, means of processing spatial data in less assumption critical ways.

Genetic algorithms and evolutionary computing provide the basis for developing new solutions to complex locational optimisation problems as well as building blocks for developing new kinds of spatial analysis and modelling tool; for example, the Artificial Life based pattern hunting creatures of Openshaw (1994d) or the automated modelling system of Openshaw (1988).

The third AI domain relevant to geography is that of computer vision. Many problems of knowledge discovery, modelling, and analysis can be viewed as tasks suitable for computer vision and pattern recognition technology. A geographical concept or theory can be regarded as an abstract pattern or process that is recurrent in a generalised manner without any great attention to precise details. Traditionally, geographers have translated general theories and concepts into highly specific aspatial but testable hypotheses that have in the process lost virtually all of their geographical dimensions. Computer vision tools provide a means of identifying two and three dimensional patterns without being confused by place specific uniqueness. It is now possible to consider handling existing theories as abstract pattern and then look for them still in a highly abstract form in the geocyberspace. (Openshaw, 1994b).

Put your AI spectacles on and many areas of geography will probably never look the same again. AI methods exist that provide a new set of tools available now for both doing and re-doing geography in the 1990s using what will one day be regarded as primitive versions of 21st century technologies.

3.3 A new computer revolution

The computer world has started to change in a most fundamental way as it enters a new era of massively parallel supercomputing. As supercomputers become faster, so new compute intensive approaches have started to revolutionise areas previously dominated by analytical methods developed at a time when computers were extremely slow. As a result Hillis (1992) writes

> "There is a significant technology transition that is taking place in computers, and since it radically changes the costs and capabilities of information processing it is likely to change our lives. The new technology is called massively parallel computing". (p1)

Many sciences are becoming computational as new computer dependent ways of doing leading edge alpha rated science become feasible. Indeed the "grand challenge" areas of science increasingly involve massively computational intensive approaches, particularly in physics, biology, chemistry, engineering, geology, and environmental modelling with particular regard to climate change and global modelling, computational fluid dynamics, modelling of macromolecules, simulation of materials, quantum chromodynamics and genome data management.

Massively parallel supercomputing involves multi-CPU hardware linked by very high speed networks and with shared global memory addressing. The idea is not new and the technology has been around for over 20 years, the point is that recently a certain maturity has set in and the latest systems are starting to provide extremely powerful supercomputing environments that seem likely to reach teraflop speeds in the second half of the 1990s. Whereas the speed of the UK's fastest research supercomputers improved by about one order of magnitude over the last 10 years, it is now likely that the next 5 years could see speed-ups of between three and four orders of magnitude. As a result, by 1999 it is likely that compute speeds will have increased by 10^9 times since the quantitative revolution of the 1960s, 10^8 times since the entropy maximising mathematical modelling revolution of 1970, and 10^6 times since the early days of the GIS revolution.

At a time when teraflop computing machines possessing terabyte memories are expected soon; most geographers are seemingly still astonished by the appearance on their desks of pc's with a performance equivalent to early 1980s mainframes. Their appreciation of the possible is determined by the past and the concept of machines up to a million times faster and bigger in the late 1990s is just beyond belief. Yet this is the new era of parallel supercomputing that is dawning now. There are tremendous new opportunities for innovation in both creating, developing and applying new technologies. We can now start to take major strides forward by becoming more computationally minded about those problems that both need and can exploit the power of massive parallel supercomputers. It is interesting to note, therefore, that many map related modelling and analysis tasks in geography are implicitly parallel and could well benefit from parallelisation. Also there are many hard problems that could not previously be solved, and this partly explains the retreat of many geographers (and other social scientists) into qualitative and more descriptive approaches. A fresh re-look at the problems of geography via a computational paradigm starts to become feasible once computers become fast enough and big enough; that time has either arrived or is now not far off. It is certainly both sufficiently close and certain as to justify the commencement of forward looking research initiatives in this area.

In general then an opportunity now exists to evolve both the soft and hard areas of geography by moving them into the computational era. This is not an attempt to revive quantitative geography. It has nothing to do with any attempt to return to positivism, nor is it merely technology led or using technology for its own sake. It is just about exploiting the new ways of doing geography that are becoming available.

4 Computational Geography

By themselves the three developments (GIS, AI, and parallel supercomputing) may have little impact on geography. However, taken together and located in the context of the emerging IT State and global data capture via satellite and related systems, there are major synergistic effects. They combine as Figure 1 suggests to support a new computational style of geography suitable for the 21st century. Computational geography is, no more than the development of an overtly computational approach to doing geography that seeks to utilise the spatial information riches, AI tools, massively parallel supercomputing and existing knowledge in whatever form it exists

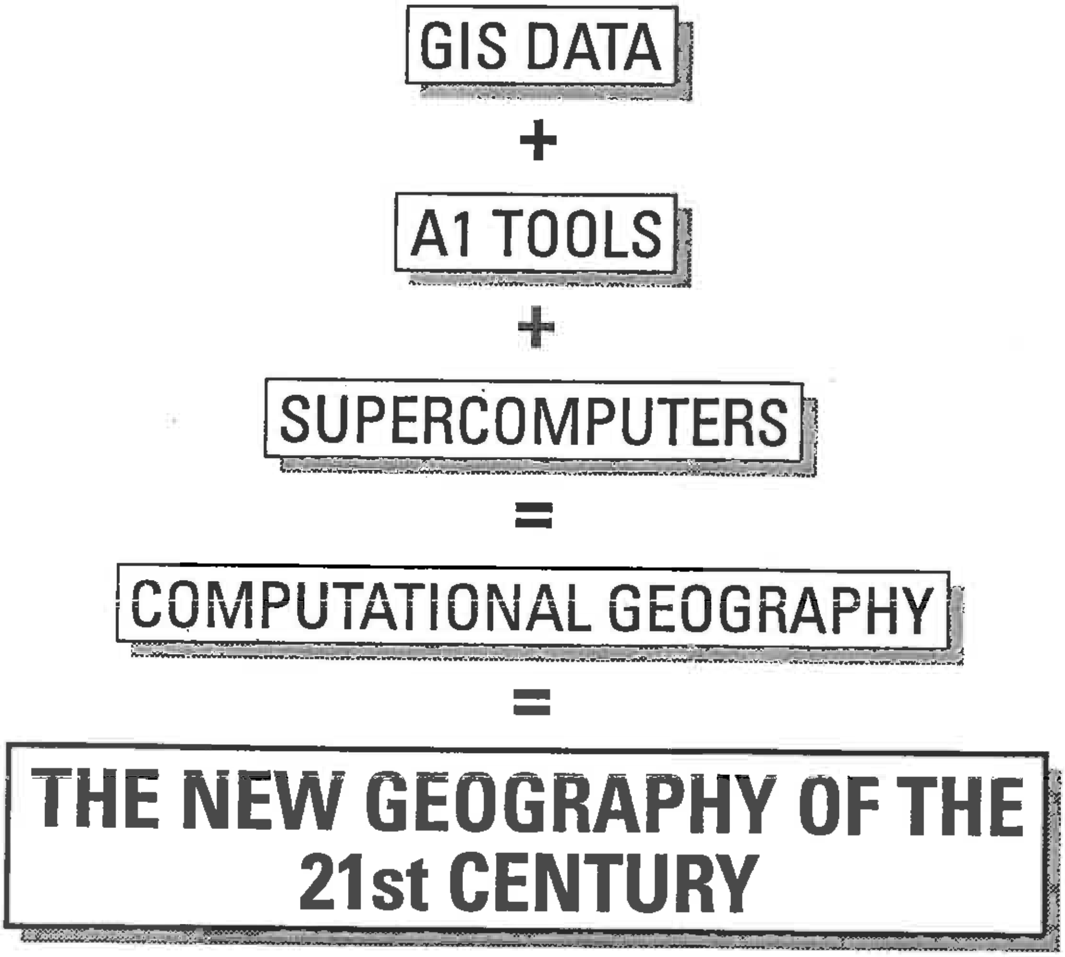

Figure 1. The computational geography Iceberg

(Openshaw, 1994c). It is sufficiently broadly defined, generic in its technology, and flexible in philosophical outlook so as to encompass most, if not all, areas of human and physical geography. We can now start to approach many of the problems of geography (of theory formulation, of analysis, of socio-spatial understanding) with a powerful new set of tools; tools that are near the beginning of their evolution rather than at the end. In this way we can seek to evolve geographical thinking across the entire width of the discipline, more or less simultaneously. Why not create a new grand challenge area for science focused on understanding and modelling the world of people and their systems. Is not this area of Human Systems Modelling (HSM) just as legitimate an area for science as the more traditional areas of chemistry, physics, etc. It is not enough to understand the workings of the world's natural systems (viz ocean, atmosphere, biosphere, climate) whilst ignoring human systems.

Currently we know very little about the scientific nature, behaviour, and processes of change that characterise virtually all the people related systems. Much more is known about the workings of the physical systems than of the operation of even the simplest of our human systems. At a time when over 80% of the population of the UK and EC now live in cities, it is surprising that our best computer models of cities date from over 25 years ago. Considerably more is known about the world's endangered wild life than of the unemployed on Merseyside, of atmospheric dynamics of Mars than the behaviour of commuters in any major cities, of the three dimensional processes of the world's oceanic circulation systems than of the effects of the recession on the local space economy of Bottle, of the quantum behaviour of atoms than the optimal geographical locations and distributions of facilities on which we all depend for critical services. Yet it is still seemingly more acceptable to use high performance computers to search for an even larger prime number than it is to seek to utilise the same hardware to develop an improved scientific understanding of key human systems. Indeed supercomputing is much more likely to be used to investigate seemingly obscure problems in theoretical physics than to search for disease clusters or patterns in crime data or better business applications. Somehow, some of the key concerns of modern life are escaping adequate science because they fall between the 'hard' physical sciences and the 'soft' social "sciences". Nor is it likely that soft social science, that largely non quantitative pseudo-science of philosophers and story tellers, is going to fill the gap. What is needed is a new concept of a spatial or geographic science that seeks to understand and models all the worlds' systems, their interactions, and their dynamics. Nothing less is acceptable. We owe it to the survival of our species to develop a more spatio-holistic approach and we need to start doing it soon.

Perhaps the first steps in this much needed scientific revolution in the social sciences can best be seen in geography, as a blueprint or model for many other subjects. If geography cannot evolve to embrace both its soft and hard areas then what hope is there for the emergence elsewhere of a new and slightly more scientific social science. Indeed the opportunity, right now, is to start this process by moving, probably slowly at first, all areas of the discipline into a new computational era. Computational geography is about doing geography and evolving its technological basis and style to cope with the demands, needs and opportunities of the 1990s. For it to work, this evolution needs partnerships between participants, it needs geographers who previously have been artificially isolated by major epistemological differences to start to think about their commonalties and to work together. In this vision of a new new-geography there is no ready or pre-existing priesthood; there are no experts as yet, no single group of geographers that have all the relevant skills, the start point is a level playing field. Who can yet say whether those who know about parallel programming have any lead over those who understanding something of how society operates. What advantage is their in being a GIS expert without concepts; or a social theorist without spatial information, or an AI expert without any understanding of the problems that it can be used to solve.

Geography is full of hard problems. Many are relevant to the concerns of the 1993 White Paper on science (Cm 2250) with its refocusing of science towards economy and quality of life concerns. The hardness of the problems has traditionally either resulted in a qualitative approach or to a highly specific, narrowly focused, inadequate subsystem modelling. It is easy to become dissatisfied with such imperfect technologies that soon become too complex or too hard to teach or research. No wonder there was a retreat from mathematical urban modelling in the late 1970s; it soon became too difficult and the rate of progress slowed so much as to become imperceptible. A similar fate will soon befall the mathematically simpler soft approaches. Story telling and sociospatial process fables stimulates the need for testable understanding, for prediction, for more "science". Sooner or later, dissatisfaction levels will increase to destroy the appeal of pseudo-science. The world needs more science not less. It also needs management, control, and forecasting systems not just an ability for subjective albeit sophisticated retrospection. As many of the world's key people systems look like spiralling out of control the need is in the future is to look forward not backwards, and to do so in a more scientific, neutral and non partisan manner partly as a means of meeting egalitarian goals.

From a geographical point of view computational geography provides a new paradigm for doing geography, one that is likely to slowly evolve or unfold rather than revolutionise with major radical change over a short time frame. It provides a basis for improving what we already do, it can be used to add a touch of scientific rigor to those areas that currently lack it, to tackle previously impossible problems, and to start to extract value, new concepts, new theories, and make new discoveries from the information rich environments all around us.

5 A research manifesto

There are clearly many different areas and topics of interest within such a broadly defined subject as computational geography. Any prioritisation can only be both personal and subjective. The real need is for brainstorming sessions of flexibly minded geographers to define their own areas of interest. The three outlined here are amongst those than seem likely to become important.

5.1 Smart exploratory geographical analysis and modelling

The world is full of more geographically referenced data than ever before. The expanding geocyberspace is creating more and more opportunities for geographical analysis and there is an increasing expectation that geographers will meet these needs. Indeed, there are a growing number of data sets for which there is a public good imperative for analysis, for example, morbidity and mortality data, disease and cancer data, crime and traffic accidents. The problem at present is that many (perhaps 99%) of these databases are simply not being adequately analysed. There are a number of obstacles: for instance, excessive, misplaced and exaggerated Data Protection concerns; a lack of analysis tradition; and the absence of suitable analysis technology. Another problem is that the conventional analysis process is hypothesis driven. However the newly emerging data riches are occurring in areas and domains that are hypothesis deficient. Existing exploratory analysis tools are dimensionally restricted in the sense that they work best with few variables and simple enough relationships that can be explored by human based graphical means. However, the key feature of the geocyberspace is what might be termed its multidimensional dynamic complexity. There is an urgent need for a new generation of geographical exploration technology that can cope rather than ignore or artificially reduce the complexity so as to allow the simplest of methods to be used.

Parallel supercomputing provides the engine for many new types and styles of geographical analysis tools. The problem of not knowing where to look for localised patterns and relationships can be solved by looking everywhere via brute force or smart search technologies. Much of geographic analysis is a search process which is explicitly parallel and thus ideally suited for massively parallel processing hardware. The Geographical Analysis Machine (GAM) of Openshaw et al (1987), Openshaw and Craft (1991) and the Geographical Correlates Exploration Machine (GCEM) of Openshaw et al (1990) are now feasible on a large scale.

However, the search engines can be greatly improved by the incorporation of AI technology. Smart spatial analysis tools need to be created that can successfully hunt for patterns and relationships (i.e. empirical regularities) in the jungle of the geocyberspace without having to be told, in advance, precisely <u>where</u>, <u>when</u>, or <u>what</u> to look for (Openshaw, 1994d). Traditionally we have coped with complexities mainly by either artificially designing and simplifying it away or by ignoring it. Research designs and study protocols have been applied to the data analysis process with an increasingly stifling effect. When there was little data this was quite acceptable and the impact was small. However, as the supply of data rapidly increases and the number of variables and different levels of data resolution explode, so the unintentional impact rapidly becomes excessive. Increasingly, it seems that data are expected to speak for themselves after having been strangled, and the results obtained may well reflect the constraints of the study protocols rather than the real patterns in the data (Openshaw, 1994f).

The GIS database consists of three types of information: geographical, temporal, and attribute. Analysis and modelling tools are needed that can operate in all three spaces and on any of the seven permutations of them; see Table 1. Each of these spaces is measured in different, non commensurable units; and some may have multiple levels of measurement as well as being multivariate. The world of GIS emphases mainly the geographic and attribute spaces, with the emphasis on the former. This is by no means adequate.

An example may help identify the problems. Suppose you wish to analyse a disease. Conventionally, you would agree protocols that defined the disease of interest, the study region, the time period, any covariates (such as age-sex classes), temporal periods of study, and the spatial scale of the data to be used. The results reflect the interaction of these study design decisions on the unknown patterns in the real world. Uncertainty occurs at all levels. The selection of the disease to be studied disguises considerable uncertainty that the disease categorisation, despite international standardisation, is subjective and other "different" diseases should perhaps also be included on the grounds that the suspected but unknown aetiology (but not the cell morphology) may be similar. The choice of time period can also be absolutely critical to the detection of patterns that may well be localised in space as well as time. The choice of data and uncertainties contained therein is another key design aspect. Why cannot we explore all the data without being so highly specific? Why cannot we seek to identify major empirical regularities and patterns without being initially so prescriptive. Why cannot we develop a spatial exploratory analysis technology able to hunt out the major patterns in our databases wherever, whenever, and whatever they "exist". Data mining and data trawling has traditionally been derided by those with hypotheses to test and lacking in the necessary technology for anything more exploratory. Lets be quite clear, it is these smart look everywhere analysis tools that are now needed. Without them we are geocyberspace blind and will lack the ability to spot virtually anything new that we do not blunder into and even then we may not realise it.

Geocyberspace blindness is not a new disease. Abbott (1884) "Flatland" portrays a two dimensional world which is visited by a sphere from spaceland (a three dimensional world). Much

Table 1. Different data space interactions

geographic space only
geographic space - time space
geographic space - attribute space
geographic space - attribute space - time space
time space only
time space - attribute space
attribute space

Note: the ordering is irrelevant

visualisation in GIS emphasises a two and three dimensional view of a data world which extends to far higher levels of hyperdimensionality. The problem is how to cope with beyond three or four dimensions. New methods and tools are needed to explore these hyperdimensional spaces that cannot be seen or easily visualised. Computer artists think that they can now "see" the fourth spatial dimension. Robbin (1992) write "when the fourth dimension becomes part of our intuition, our understanding will soar" (p73). However, to properly handle the geocyberspace we need to discover how best to "see" patterns and relationships in much higher dimensional spaces. A two or three dimensional reduction strategy will never work well enough!

One possible solution is to use the new science of artificial life to create pattern hunting creatures that can cope and thrive in hyperdimensional and multimetric geocyberspaces. Openshaw (1994f) argues that, in principle, many of the problems of exploratory geographical analysis can be tackled by building intelligent autonomous pattern hunting creatures that can live and function in databases. By following their movement and behaviour via two and three dimensional windows into their higher dimensional and multiple data domain world, it might become possible to visualise, analyse, model, and decode many more of the patterns and relationships that fill the world about us but of which we are blind and know very little.

5.2 Spatial modelling of people systems

Another most urgent need is for models of people systems. It is noted that in the 1960s and 1970s major advances were made in developing mathematical models of various urban and regional systems (Wilson, 1970; 1981; Batty, 1976). However, a most fundamental obstacle was also encountered. The mathematical models worked best with Weaver Type I and Type II systems; they could not cope at all with Weaver Type III systems of organised complexity (Wilson and Bennett, 1985). As a result the best available models of city systems are now at least 25 years old. Furthermore, many other key people related systems remain unmodelled or rely on similarly old technology. It is seemingly much easier to model the worlds' climatic and oceanic systems than to build a good model of even a single city, despite the fact that most countries are now overwhelmingly urban. This modelling neglect is embarrassing and it seems that we know considerably more about the global numbers of eagles, sperm whales, and bison than we do about the number and distribution of people who are unemployed, sick, or deprived in our societies. It also seems that models of global climatic change are being expected to function without any real knowledge or inputs about people related systems. Of course, we are not ignorant about society, it is just that our scientific understanding of how it works now lags far behind a broader and softer social science appreciation of its functioning and organisation. The challenge now is to develop new model building technologies able to incorporate our Catherine Cookson style of descriptive geographic understanding into something more scientifically rigorous and useful.

There is a growing need for good working models of all the key people related spatial systems that matter. The list of excuses for continuing the neglect is now greatly diminished. For instance, there is now a wealth of relevant data, there is better qualitative knowledge and understanding of how the systems work, and there is the basis for a new and much more powerful model building technology that escapes the restrictions of mathematical modelling. In particular, neurocomputing methods can incorporate soft, fuzzy, and qualitative relationships. Computers are also becoming big enough and fast enough to allow microlevel modelling of people's behaviour to occur, and new computer model building methods will soon enable reasonable levels of modelling of chaotic nonlinear systems. Maybe the key to progress in this area is to develop as a matter of considerable urgency practical means of incorporating the great wealth of soft concepts and thus produce an almost instantly more intelligent modelling technology. Massively parallel processing creates the basis for starting again to re-think and re-create useful scientific models of human systems. The task is not easy but maybe, in common with many other leading areas of science, it is possible to compute our way out of the problems. For example, the latest MPP hardware has just about enough power and capacity to start to support large scale spatial modelling. Birkin et al (1994) give an example of a parallel spatial interaction model's being run 2700 faster than the original workstation software. That is good, but this model may now need to be embedded in a system wide global optimisation procedure which may well run it the equivalent of 1 million times. Another example would be the mathematical model breeding machine (Automated Modelling System) of Openshaw (1988). The latest version of this search for good performing models through the universe of all possible models requires 6-8 weeks on a fast workstation (Openshaw and Turton, 1994). Without an MPP implementation it is not a viable technology. With a MPP port it has the potential to create entirely new forms of mathematical model.

5.3 Developing geographic knowledge systems

Another key need is to build computer systems able to create new concepts and discover (or uncover) new knowledge by mining the geocyberspace. It is noted that the quantitative geography of the 1960s style failed to contribute much, or any, new geographical knowledge or theory. The mathematical modelling revolution of the 1970s did likewise little to improve geographic knowledge, although it did provide a much more rigorous scientific basis for existing theory. The GIS revolution of the 1980s will likewise fail to deliver much new knowledge because it is focused on data and lacks the tools necessary to create new knowledge from data.

One problem is that geographers have so far failed to develop an explicitly geographical form of analysis. Statistical methods have been employed in a geographic context to test hypotheses that are only implicitly, or vaguely, geographical. In the process much of the innate geographicalness is lost; for example, a test of spatial autocorrelation is far more precise and limited in scope than the geographer's concept of spatial autocorrelation. Likewise, a test of distance decay effects is more a test of a highly specific distance decay function than of the concept which is function invariant. The statistical methodologies that have to-date been applied in geography are in some senses far too precise to adequately handle many abstract geographic theories and concepts. A much more useful technology is needed if progress is to be made.

One way forward is to adopt a pattern recognition approach. Openshaw (1994b) suggests the use of AI and robotic vision methods to search for abstract geographic patterns that represent concepts. Recurrent spatial patterns that are important may well be regarded as particular instances of more abstract theories and concepts. The secret is to be able to search for general or abstract geographical patterns as entities in their own right, without having to be too specific or precise about their structure and form. For example, is the spatial pattern of Sheffield's residential social structure "similar" to that of Leeds? Note that the pattern of interest here is the general structure of the sociospatial structure after scale, rotation, and translation effects are ignored. This is a much more general and geographically relevant question than can be answered by traditional quantitative geographic methods. AI based pattern recognition methods can probably be adapted to help answer it or else appropriate computer vision tools found that will meet this need.

Once this type of question can be handled then many applications immediately follow; for example, how many different types of urban spatial social patterns exist in Britain? It is interesting that in the 1950s, concepts of this sort were popular in geography. Why not create libraries of scale invariant two and three dimensional geographic patterns as objects that define, contain, and represent geographical knowledge of recurrent phenomenon. It would seem that similar types of pattern recognition paradigm can be applied to many other geographic problems. Intelligent geographical pattern recognition provides one way of processing geographic information to create new geographic knowledge. Figure 2 shows how this process might work. It is, however, likely to be extremely compute intensive albeit in a massively parallel way because of the need to investigate the entire map space at many different levels of spatial resolution.

It is also noted that much geographic analysis is uniquely dumb. For example, a better multivariate classification of a town or city or country is seen as being produced purely by improved algorithms. To some extent this is correct. A neuroclassification procedure will often perform at least as well as a more conventional method, as well as offering improved data error handling characteristics (Openshaw, 1994a). However, it still makes little or no good use of over 50 years worth of research on the socio-spatial characteristics of towns and cities. What is needed are new generations of spatial classifier that take into account knowledge that exists. This knowledge is not of a production rule form commonly found in expert systems, it is much more abstract. For example, a classification of areas should be good from a global database point of view as well as optimal locally within it. Local concepts of spatial autocorrelation need to be incorporated, else it may never provide a good representation of spatial neighbourhood patterns (Openshaw, 1994e). Figure 3 illustrates one way of operationalising intelligence in a spatial classification context.

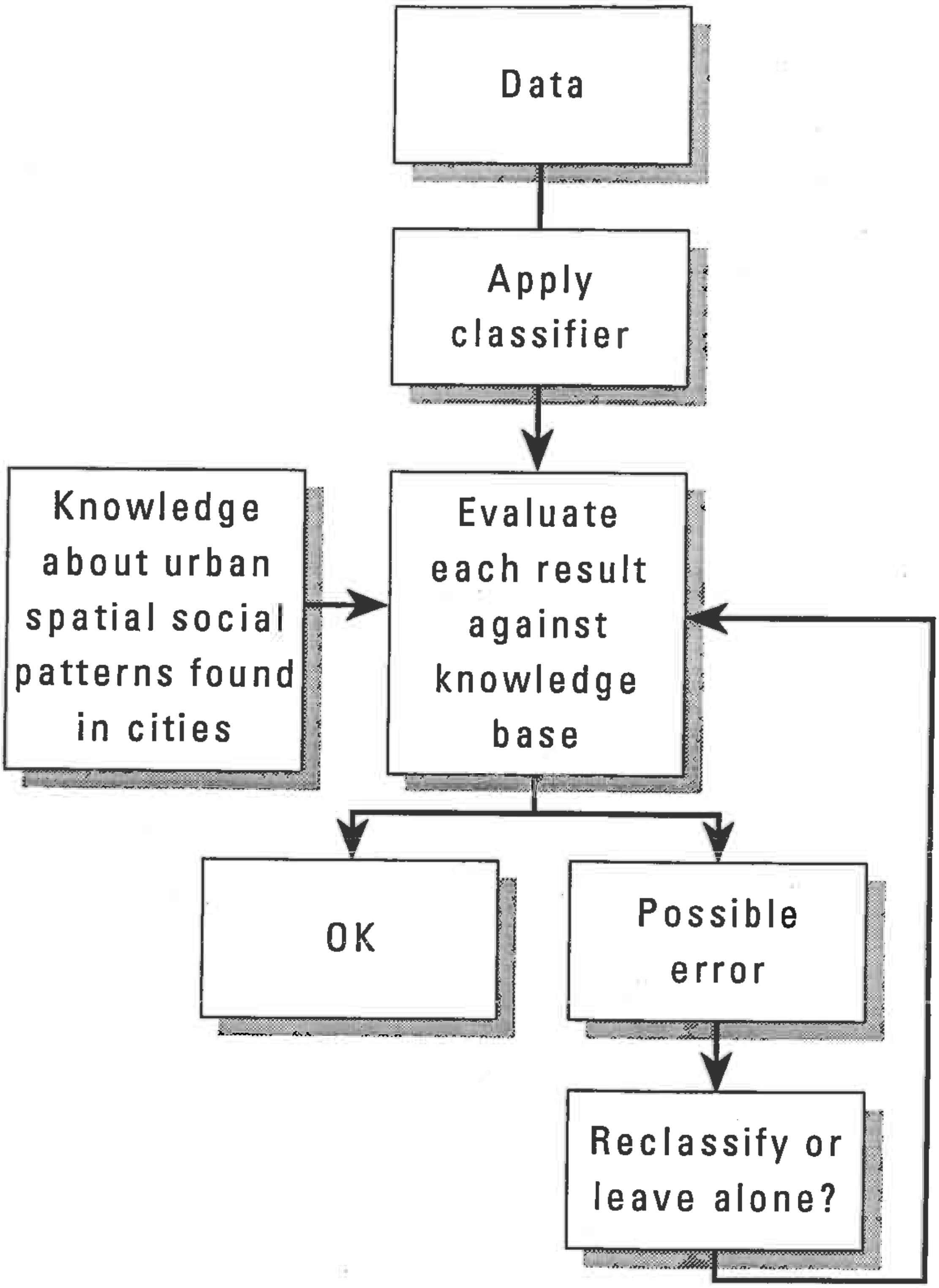

Figure 2. Building an intelligent residential area classification

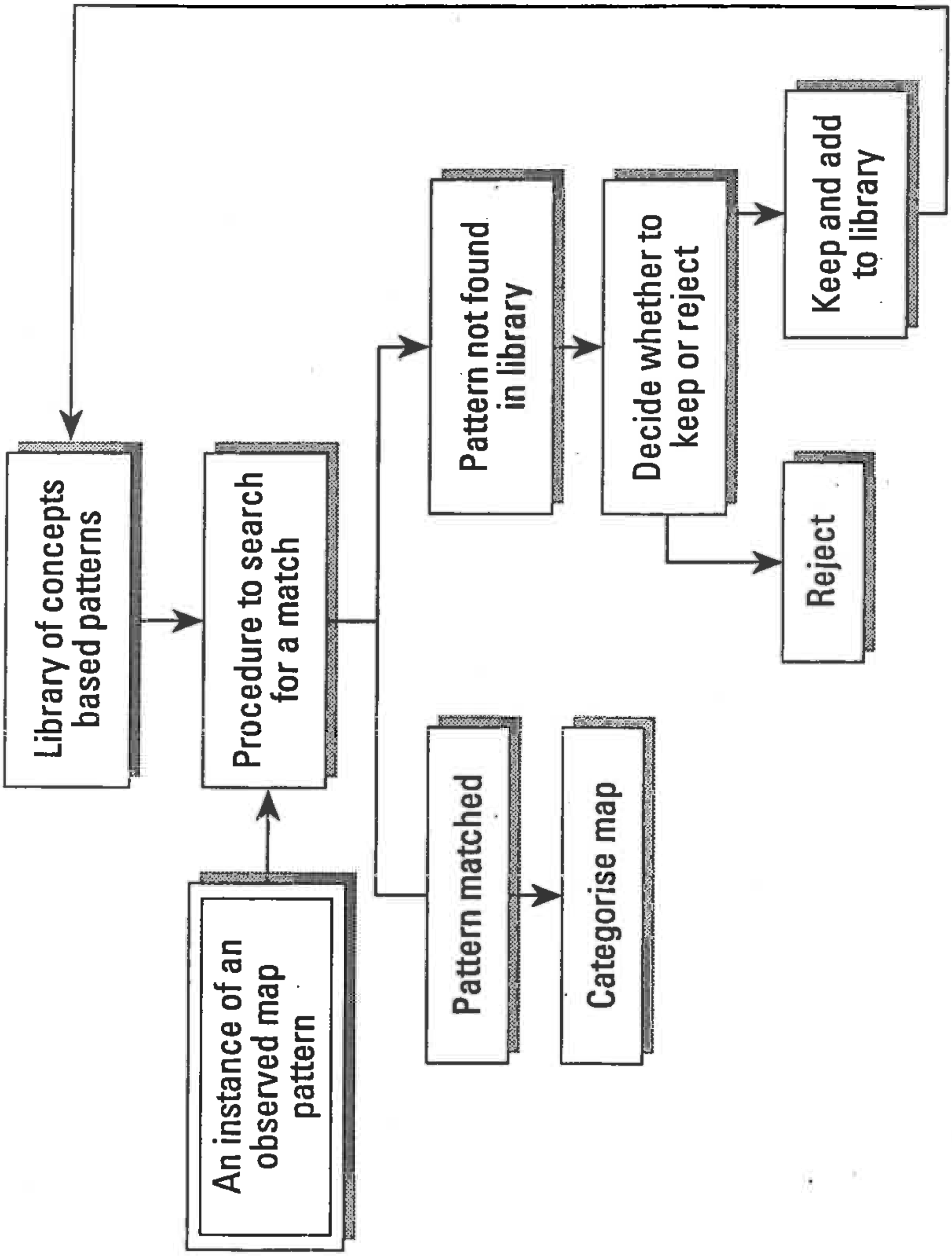

Figure 3. Establishing a library of spatial pattern concepts

6 Conclusions

The 1990s are an exciting time for geographers. It is a time of profound and rapid change in many aspects of the world in which geography is performed. There is now more available geographic data than ever before. There are new ways of doing geography and the context in which geography is performed is changing rapidly. Suddenly there is a real prospect of evolving the subject across its entire width by adopting a computational paradigm that seeks to exploit the opportunities provided by massively parallel supercomputers and by so doing preparing the way for a quantum leap forward sometime in the first decade of the next century. A start has been made but much remains to be done.

References

Abbott, E A, 1884, Flatland: a romance of many dimensions. Blackwell. Oxford.

Batty, M, 1976, Urban Modelling. Cambridge University Press, Cambridge.

Benedikt, M, 1992, Cyberpsace: first steps. MIT Press. Cambridge, Massachusetts.

Birkin, M, Clarke, M, George, F, 1994, 'The use of parallel computers to solve nonlinear spatial optimisation problems'. Environment and Planning A (forthcoming).

Chorley, R, 1987, Handling geographic Information: Report of the Committee of Enquiry. HMSO, London

Gibson, W, 1986, Neuromancer. Grafton. London.

Henderson, J V, 1992, 'Cyberspace representation of Vietnam war trauma', Beyond the Vision, Proceedings of Second Annual Conference on Virtual Reality, Artificial Reality and Cyberspace, Meckler, Westport, p77-103.

Hillis, W D, 1992, 'What is massively parallel computing and why is it important? Daedalus Journal of the American Academy of Arts and Sciences p1-16.

Openshaw, S, 1988, Building an automated modelling system to explore a universe of spatial interaction models. Geographical Analysis.

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

Openshaw, S, 1993, 'The emerging IT State. Proceedings of AURISA '93 Conference, Brisbane.

Openshaw, S, 1994a. Neuroclassification of spatial data, in B Hewitson and R Crane (eds), Neural Nets: applications in Geography, Kluwer Academic, Dordrect, pp. 53-70.

Openshaw, S, 1994b, A concepts rich approach to spatial analysis, theory generation and scientific discovery in GIS using massively parallel computing, in M Worboys (ed), Innovations in GIS. Taylor and Frances, London. pp. 123-138.

Openshaw, S, 1994c, 'Computational human geography: towards a research agenda. Environment and Planning A 26, 499-505.

Openshaw, S, 1994d. Two exploratory space-time attribute pattern analysers relevant to GIS, in S Fotheringham and P Rogerson (eds), Spatial analysis and GIS, Taylor and Frances, London, pp. 83-104.

Openshaw, S, 1994e, 'Developing smart and intelligent target marketing systems'. Journal of Targeting, Measurement and Analysis for Marketing (forthcoming).

Openshaw, S, 1994f, 'Developing automated and smart spatial pattern exploration tools for GIS applications', (The Statistician, forthcoming).

Openshaw, S and Craft, A, 1991. Using the geographical analysis machine to search for evidence of clusters and clustering in childhood leukaemia and non-hodgkin lymphomas in Britain, in G Draper (ed)). The geographical epidemeology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966-83, HMSO, London, pp. 109-122.

Openshaw, S and Turton, 1994. Building new spatial interaction models using genetic programming, Proceedings of the AISB '94.

Openshaw, S, Cross, A, Charlton, 1990. 'Building a prototype geographical correlates exploration machine', Int J. of GIS 3, 297-312.

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

Robbin, T, 1992, Fourfield: Computers Art and the Fourth Dimension. Bulfinch. Boston.

Scientific American, 1991, How to work, play and thrive in cyberspace. Special Issue September 1991.

Skelton, R A, 1952, Decorative printed maps of the 15th to 18th centuries. Spring Books. London.

Skelton, R A, 1958, Explorer's Maps. Spring Books. London.

Sutherland, I E, 1965, 'The ultimate display', Proceedings of IFIP Congress 2.

Wilson, A, (1970). Entropy in Urban and Regional Modelling, Pion, London.

Wilson, A, 1981. Geography and the Environment: System Analytical Methods. Wiley, Chichester.

Wilson, A G, Bennett, R T, 1985. Mathematical methods in human geography and planning Wiley, Chichester.