

# 18

## Applying geocomputation to the analysis of spatial distributions

S OPENSHAW AND S ALVANIDES

Developments in IT and GIS have combined to create an enormously data-rich world. The need now is to develop GIS-relevant spatial analysis tools that will assist endusers in making good use of their spatial information. The problems are: (a) the absence of many appropriate tools; and (b) the new types of analysis required as a consequence of the innovation of GIS. The hope is that a combination of GIS databases, high performance computers, artificial intelligence and a geocomputation paradigm will together provide a generic workable solutions strategy.

### 1 INTRODUCTION

The GIS revolution has created an immense wealth of spatial information in a large number of different application areas. The emphasis in GIS upon database creation and systems building will soon have to be replaced by a new concern for applications using spatial analysis and modelling. The development of GIS can be largely regarded as the computerisation of pre-existing manual procedures and established technologies that were already fairly mature in research terms. Thus no great innovation was required in order to underpin the GIS revolution, although doubtless much did occur as the software systems developed. However, when the focus of attention switches to spatial analysis and modelling then it is a very different story. Frequently there have been no existing useful manual procedures to computerise, and over the last 20 years there has been very little relevant new research focused specifically on the special needs of GIS. Moreover, an increasing number of the emerging analysis tasks are novel and have not previously attracted much or any attention – for example, the exploration of very large spatial datasets for completely unknown patterns and relationships, or the real-time analysis of live spatial databases for emerging patterns, ‘hot spots’ (see Getis, Chapter 16), and anomalies of interest.

The quest for improved analysis of spatial distributions is predicated upon three considerations:

- 1 a general desire to make use of the information resources created by GIS;
- 2 attempts to gain competitive advantages or other benefits from investments in information technology (of which GIS is a component);
- 3 hardware developments that are trivialising the costs of computation and are hence creating new ways of analysing spatial distributions.

One might argue that it is scandalous that so many key databases are not currently being properly analysed – be they pertaining to morbidity, mortality, cancer, or crime incidence – while commercial concerns and government agencies probably waste many millions (if not billions) of dollars or ECUs by poor spatial data management and inefficient locational decisions (Openshaw 1994d). Yet perhaps the users cannot be blamed for not using tools that are unavailable! The problem is essentially a longstanding failure to evolve distinctly geographical-data-appropriate tools and styles of analysis and modelling – albeit with a small number of exceptions. The legacy of statistical methods (Getis, Chapter 16) may not be helpful, at least partly because its inherent limitations need to be properly understood. In essence, no amount of

apparent statistical sophistication should be allowed to hide the fact that much of spatial statistics is very limited in what it can do, and is even more limiting in its view of spatial information and the handling of their special properties. It also needs to be appreciated that the 'post GIS-revolution' world of the late 1990s is quite different from the primeval (by contemporary standards) computing and data-poor environments in which many of the existing spatial analysis and modelling technologies were developed. Are these old legacy technologies and their latter-day offspring still appropriate, or is a new period of basic research and development needed to create the spatial analysis tools likely to be required in the late 1990s and beyond? There is no denial that advances have been made in improving some spatial statistical methods (Getis, Chapter 16), but these advances – such as developing local versions of global statistics – are of limited usefulness in practice.

This chapter attempts to address some of these concerns by focusing on the changes in the computational environment that has occurred during the 1990s. It argues that large-scale computation can now be used as a paradigm for solving some of the major spatial analysis problems that are relevant to GIS. However, if computational power is to be useful, then there also has to be a clear understanding of what the requirements and the user needs are. This leads us to a brief typology of alternative approaches and a brief illustrative case study based on one method in which high performance computing is combined with GIS data and artificial intelligence (AI) tools to develop better ways of engineering zoning systems as a decision support, analysis, modelling, and data management tool.

## 2 ADVANCES IN 'HIGH PERFORMANCE COMPUTING'

There is now considerable excitement in many traditional sciences about developments in supercomputing or high performance computing (HPC). Computation is now regarded as a scientific tool of equal importance to theory and experimentation, since fast computers have stimulated new ways of doing science via large-scale computer-based experimentation, simulation, and numerical approximation. There is equally a case for thinking that a supercomputing-based paradigm is

also relevant to many areas of GIS, but it should be appreciated that this involves much more than attempts to revamp basic GIS functions using parallel computing. HPC is defined as computer hardware based on vector or parallel processors (or some mixture) that offers at least one order of magnitude increase in computing power over that available from a mid 1990s workstation. In fact, as highly parallel processors take over from the earlier vector machines, the performance gain from using leading edge HPC hardware is more usually at least two orders of magnitude. This whole area is now developing at a rapid rate with most HPCs having a two- to three-year life cycle. For example, in 1996 it was possible to buy for £500 000 (US \$800 000) a parallel machine with equivalent performance to one costing about 10 times as much only a few years earlier (see also Longley et al, Chapter 1). The computing world is in the throes of a major technological change; that of highly parallel supercomputing (Hillis 1992). A highly or massively parallel processor (MPP) is a computing system with multiple central processing units (CPUs) that can work concurrently on a single task. This idea is not new but it was only in the mid 1990s that the technology matured sufficiently for multiple CPUs to become the dominant future HPC machine architecture (see also Batty, Chapter 21). A nice feature of MPPs is that both processing capacity and memory are scaleable – if you want more computer power, then simply add more processors. If spatial analysis tools and models are also scaleable then running them on more processors reduces computer wall-clock times in a linear way (Turton and Openshaw 1996).

Openshaw (1994b) has suggested that by 1999 it is quite likely that available HPC hardware will be  $10^9$  times faster (and bigger in memory) than what was common during the 'Geography Quantitative Revolution' years of the 1960s (when many of the current so-called spatial statistical methods were developed: see Getis, Chapter 16);  $10^8$  times faster than hardware available during the mathematical modelling revolution of the early 1970s (on which virtually all of the so-called 'intelligent' model based spatial decision support systems employed in today's GISs were based: Birkin et al 1996);  $10^6$  times since the GIS revolution of the mid 1980s (a time of considerable neglect of quantitative geography), and at least a further  $10^2$  times faster than what in 1994 was Europe's fastest civilian supercomputer – the Edinburgh Cray T3D.

A widespread problem is that many potential users appear to have failed to appreciate what these developments in HPC mean. For instance, the Edinburgh Cray T3D has 512 processors and has a peak theoretical performance of 76.8 gigaflops – but what does that mean? One way of answering this question is to create a benchmark code that can be run on the widest possible range of computer hardware, ranging from a PC, UNIX workstations, vector supercomputers, and massively parallel machines. The widely used scientific benchmark codes measure machine performance in terms of simple matrix algebra problems, but it is not clear whether this is relevant in a GIS context. Openshaw and Schmidt (1997) have developed a social science benchmark code based on a scaleable spatial interaction model which can be run on virtually any serial and parallel processor. Table 1 provides a preliminary assessment of the performance of some current HPC hardware in terms of processing speed relative to the performance of a 486 66 MHz PC. At the time of writing, the best performance for small problem sizes was the SGI Onyx (a workstation with multiple CPUs), followed by the vector processor (the Fujitsu VPX240) which was about an order of magnitude more expensive. However, once problem sizes increase then soon there is no alternative to the massively parallel Cray T3D with speed gains of about 1335 times for a 10 000 by 10 000 zone matrix (equivalent to the UK ward level journey-to-work or migration data from the 1991 Census). This run took 2.4 seconds of wall-clock time (compared with 18 hours on a workstation), while the even larger 25 000 by 25 000 benchmark required 13 seconds. Note that HPC is not just about speed but also memory. The larger memory sizes required in these latter two runs reflect problems that have previously simply been uncomputable. It is interesting that during the 1990s machine speeds have been doubling every 1.5 to 2 years, and that this is expected to continue for at least another 10 years. One way of explaining what these changes in HPC hardware mean is to ask: how would you approach the spatial analysis challenges presented by GIS if that workstation on your desk were about 5000 times faster and bigger?

The criticism that few GIS end-users will ever be able to afford HPC hardware is irrelevant for two reasons. Firstly, what is possible using, for example, a mid 1990s national HPC research centre machine will within five years be affordable and ‘do-able’ using many workstations – even earlier if

**Table 1 Relative performance of a selection of available HPC hardware on a social science benchmark code in relation to 486 PC.**

Hardware	Problem size: numbers of origin and destinations					
	Number of processors	100 by 100	500 by 500	1000 by 1000	10 000 by 10 000	25 000 by 25 000
<b>Massively parallel</b>						
Cray T3D	64	88				
	128		241	258		
	256			545	665	
	512				1335	1598
<b>Parallel</b>						
SGI Onyx	4	218	221	192	np	np
SGI Power Challenge	4	51	66	63	np	np
<b>Vector supercomputer</b>						
VPX240	1	162	195	196	np	np
Cray J90	8	8	35	39	np	np
<b>Workstations</b>						
SGI Indy	1	10	10	9	np	np
HP9000	1	14	12	10	np	np
Sun Ultra 2	1	18	17	16	np	np
<b>Personal computers</b>						
133MHz Pentium	1	3	4	4	np	np

*Note:* Benchmark problem sizes greater than 1000 by 1000 are not possible on a 486 PC. The times are estimated using linear interpolation which provides a good statistical fit to a range of smaller-sized problems.

‘workstation farms’ are used. Note that five years is probably also the lead time for the research and development cycle of new spatial analysis tools. Second, it is possible that the need for highly specialised but generic analytical functions will be met through the development of *embedded systems*. Embedded systems are unifunctional hardware that typically employ multiple CPUs. They are currently mainly used in signal processing. However, there is no reason why they cannot be programmed to perform specialised spatial analysis functions which may need large amounts of processing power. Such a system could appear to the GIS user as a call to a subroutine, except that the subroutine is in fact located somewhere else on the Internet and is not software but an integrated hardware and software

system (see Maguire, Chapter 25). This form of heterogeneous distributed GIS is possible now and is one way of handling highly specialist but generic needs for spatial analysis functionality. The problem at present is that there is not yet a single example of such a system in operation, and not many ideas about the nature of the spatial analysis technology that should be run on them. On the other hand, the good news is that the languages and software tools needed to develop portable and future-proofed parallel applications are now quite well developed. Very significant here is the recent international standardisation of both a Highly Parallel FORTRAN (HPF) compiler and of the message passing interface (MPI).

### 3 A GEOCOMPUTATION PARADIGM FOR GIS

The rise of scaleable parallel hardware dramatically increases the opportunities within GIS for large-scale spatial analysis, using new approaches that seek to solve some of the traditional problems by switching to a more computationally intensive paradigm (see Openshaw 1997 for a review). There are now new ways of approaching spatial analysis using what has been termed a geocomputational paradigm. Geocomputation is itself a relatively new term, defined as the adoption of a large-scale computationally intensive approach to the problems of physical and human geography in particular, and the geosciences in general. Geocomputation is a paradigm that is clearly relevant to GIS, but also goes far beyond it. Spatial data manipulation on parallel supercomputing may well involve a return to flat data held in massive memory spaces, rather than recursive relational and hierarchically structured databases held on disk (see Worboys, Chapter 26). It also involves the development and application of new computational techniques and algorithms that are dependent upon, and can take particular advantage of, supercomputing. The motivating factors are threefold:

- 1 developments in HPC stimulating the adoption of a computational paradigm to problem solving, analysis, and modelling;
- 2 the need to create new ways of handling and analysing the increasingly large amounts of spatial information about the world stored in GIS;
- 3 the increased availability of AI tools and Computational Intelligent methods (Bezdek

1994) that already exist and are readily applicable to many areas of GIS (Openshaw and Openshaw 1997). Geocomputation also involves a fundamental change of style with the replacement of computational minimising technologies that reflect an era of hand calculation by a highly computationally intensive one. It also brings with it some grand ambitions about the potential usefulness that may well result from the fusion of virtually unlimited computing power with smart AI-based technologies that have the potential to open up entirely new perspectives on the ways in which we do geography and perform GIS applications (see Openshaw 1994a, 1995). This new emphasis on geocomputation is an unashamedly applied, problem-solving approach. The challenge is to create new tools which are able to suggest or discover new knowledge and new theories from the increasingly spatial data-rich world in which we live.

### 4 WHAT SORT OF HPC-POWERED GIS-RELEVANT SPATIAL ANALYSIS TOOLS ARE NEEDED?

A longstanding difficulty with GIS-Relevant Spatial Analysis (GRSA) is the lack of any consensus as to what it means, what its requirements are, what its users want now, and what its users would want if only they knew it were possible (or if the methods existed to stimulate demand). The situation has not improved much over the last five years (see Openshaw 1991). Far too often GRSA is equated with whatever old or new statistical technology a researcher happens to be familiar with, or with what a largely unskilled enduser thinks is required based on knowledge of what a proprietary GIS vendor provides. Yet GIS offers far more than a source of data that can be run with pre-GIS methods! 'Yes', that can be done, and 'yes' it can be useful; but GIS presents a much deeper challenge to spatial analysts. The question 'what kind of spatial analysis do researchers and academics want in GIS?' has to be tempered by the feasibility constraint of 'what kinds of spatial analysis can be implemented in or with GIS?' and the sensibility constraint 'what types of spatial analysis is it sensible to provide for GIS and its user community?'. Another set of general design constraints reflect other very important but hitherto neglected considerations, such as who are the likely

users, what it is (in generic terms) that they want, and what sort of analysis technology they can handle given fairly low levels of statistical knowledge and training in the spatial sciences. Table 2 summarises many of the principal design questions. It is noted that the abilities of users are very important and the future viability of whatever technologies are proposed will ultimately depend on the extent to which the methods can be safely packaged for use by non-experts. There is a considerable mismatch between the criteria identified in Table 2 and the capabilities of existing spatial statistical and spatial modelling tools: for example, statistical packages are really of potentially very limited use outside of a research organisation and, in any case, they lack the power to cope with most of the analysis problems created by GIS.

**Table 2 Basic design questions.**

What kinds of spatial analysis are:

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

A most important challenge is to identify and develop generic spatial analysis tools which are appropriate for use with spatial data in GIS environments. Table 3 shows the ten basic 'GISability' criteria that spatial analysis methods should ideally attempt to meet (Openshaw 1994e). It is important to recognise that GIS creates its own spatial analysis needs and that these needs make it a special and different sub-field of spatial statistics. It is within this context that spatial analysis tools need to be regenerated, rediscovered, or created anew. The present is a good time to tackle these problems.

The debate as to whether these methods should be accessed from within or without a GIS package is irrelevant. There is no reason to insist on only one

**Table 3 Openshaw's 10 basic 'GISability' criteria.**

1. Can handle large N values
2. Study region invariant
3. Sensitive to the nature of spatial data
4. Mappable results
5. Generic analysis
6. Useful, and valuable
7. Interfacing problems invariant
8. Ease of use and understandable
9. Safe technology
10. Applied rather than research-only technology

Source: Openshaw 1994f

form of integration or interfacing, leaving aside the obvious point that to be a GRSA tool the spatial analysis operation has to be called from, and ultimately end up within, a GIS environment. In an era of heterogeneous distributed computing there is no longer any need for all of the systems to be on the same machine (see Coleman, Chapter 22). Equally, the extent to which methods are perceived as having to be run within a GIS environment is often overplayed, since the special properties that a GIS can offer spatial analysis amount to little more than spatial data and consistently defined contiguity lists. Thus much system complexity can be avoided by the simple expedient of separating out the different components needed by the analysis process and developing a high-level system to call a GIS here, a model or analysis tool there, or a map drawer when one is needed.

These GRSA criteria can be converted into a series of researchable topics that would appear to have considerable relevance in the late 1990s. Participants at a workshop on 'New Tools for Spatial Analysis', held in Lisbon in 1993, were asked to think about the research themes that might be the most useful in the spatial analysis area. A summary of the suggestions that emerged after several hours of discussion and debate spread over a three-day period is given in Table 4. At present there is no funding to develop any of these themes, and this is one of the reasons why spatial analysis relevant to GIS is so backward.

**Table 4 A research agenda for spatial analysis in the mid 1990s.**

Theme 1: Toolkits for spatial analysis in GIS
Theme 2: Methods and tools for handling uncertainty in spatial data
Theme 3: Methods of automated and exploratory spatial analysis
Theme 4: Data driven modelling and data mining of GIS databases
Theme 5: Statistical aspects of model evaluation and choice
Theme 6: Confidentiality of spatial data and zone design
Theme 7: Impediments to the development and use of spatial analysis methods

Source: Adapted from Openshaw and Fischer 1995

## 5 A TYPOLOGY OF SPATIAL ANALYSIS TECHNOLOGIES

The developments in HPC environments and the increasing interest in geocomputation as a paradigm relevant to GIS offer a useful perspective on the current state of GRSA methods. A threefold typology is suggested:

**1 Type One** methods are based on computationally limited technology. Most conventional statistical methods are of this type. It is true that some require supercomputers to invert rank  $N$  matrices, where  $N$  is the number of spatial observations which can be very large. However, this is still computationally limited technology if, for example, there are only a small number of spatial origins, or if only a few variables can be handled.

**2 Type Two** methods are computationally intensive, but in a dumb manner. The early uses of supercomputers in spatial analysis resulted in the development of 'brute force'-based exploratory pattern and relationship detectors. For example, they include the Geographical Analysis Machines (GAM) of Openshaw et al (1987), and the GAM/K version of Openshaw and Craft (1991) who used Cray 1, 2, and Cray X-MP and Y-MP vector supercomputers to power a large-scale exploratory search. These methods were originally criticised by some spatial statisticians (Besag and Newell 1991) who suggested simpler variations. Subsequent testing appears to have demonstrated the superiority of the GAM/K

variant, as documented in Openshaw 1997b, although the results were five years late in being published! The Geographical Correlates Exploration Machine (GCEM) of Openshaw et al (1990) is another type of 'brute force' search for localised patterns and geographical associations. Since 1990, both methods can be run on workstations. A few variations on the GAM theme in particular have been suggested by other researchers – for example, Fotheringham and Zhan (1996) describe a procedure almost identical to part of the GAM. Note though that the quality of the results does depend on the purpose of the experiment or investigation. For example, if the objective is to test hypotheses then typically several million Monte Carlo significance tests with repeat replication may be required in order to handle multiple testing problems. The process is highly parallel but probably still needs the next generation of machine. However, if the objective is to use a GAM style of approach as a descriptive tool, indicating areas where to look or perform more detailed work, then this can be avoided. The advantages of this style of approach are essentially those of automating an exploratory spatial analysis search function as well as obviating the need to have prior knowledge about where to look for localised spatial patterning. Type Two methods were used in the first spatial analyses to recognise the importance of searching for localised patterning rather than global patterns.

**3 Type Three** methods are computationally intensive but also computationally intelligent. The difficulty with the Type Two methods is that as problem sizes increase (e.g. as a consequence of improved data resolution) and as the dimensionality of the data increases from two spatial dimensions to multiple data domains, then this technology breaks down. The answer is to switch to a smart search strategy. Openshaw and Cross (1991) described the use of a genetic algorithm to move hypercircles around a multidimensional map in the search for crime clustering. This technology has been developed further into suggestions for exploratory analysis that can operate in space, in time, and in multivariate data domains. The resulting database exploring creatures (termed Space Time Attribute Creatures – STACs) are described in Openshaw (1994e, 1995). Interpretation of the outcome of analysis can be aided by using computer animation

to follow the search behaviour of the STACs as they go on a data pattern hunting safari. This has been developed into a prototype system called MAPEX (Map Explorer: Openshaw and Perrée, 1996). The hope is that the inexperienced or unskilled endusers can visualise and discover what is happening in their databases by viewing a library of computer movies illustrating different amounts and types of geographical patterning. The resulting user-centred spatial analysis system uses computationally intensive methods in order to ensure that the results that the users see have been processed to remove or at least reduce the effects of multiple testing and other potential complications. The users are assumed not to be interested in  $p$ -values and Type One or Type Two errors, but merely need to know only where the patterns are strongest and whether they can be trusted. The computer animation provides a useful communication tool. The expansion of this technology into the multiple data domains occupied by STACs and its integration into a standard GIS environment is currently underway. The zone design problem of Openshaw (1978) is another example of a Type Three problem which is of practical significance. Its wider application has been delayed until both digital boundaries were routinely available and HPC hardware speeds had increased sufficiently to make zone design a practical proposition.

The challenge for the geocomputational and HPC future is how to evolve more Type Three methods, which can handle rather than ignore the challenges of performing more intelligent spatial analysis, using computational and AI technologies. Spatial analysis needs to become more intelligent and less reliant on the skills of the operator, and this can only be achieved in the long term by a movement towards Type Three methods. The ultimate aim is to develop an intelligent partnership between user and machine, a relationship which currently lacks balance. Many of the statistical and computational components needed to create these systems exist. What has been lacking is a sufficient intensity of understanding of the geography of the problems and of the opportunities provided by GIS in an HPC era. A start has been made but much work still needs to be done.

## 6 SPATIAL ANALYSIS INVOLVING COMPUTATIONAL ZONE DESIGN

Geographers have been slow to appreciate the importance of spatial representation in their

attempts to describe, analyse and visualise patterns in socioeconomic data. The effects of scale and/or aggregation of zones upon the nature of the mappings that are produced are well known (Openshaw 1984; Fotheringham and Wong 1991). Yet this very significant source of variation is usually ignored because of the absence of methods in existing GIS packages that are able to handle it. Once this mattered much less, inasmuch as users had no real choice since they were constrained to use a small number of fixed zone based aggregations. GIS has removed this restriction and as the provision of digital map data improves, so users are increasingly exposed to the full range of possible zoning systems. Openshaw (1996: 66) explains the dilemma as follows: 'Unfortunately, allowing users to choose their own zonal representations, a task that GIS trivialises, merely emphasises the importance of the MAUP. The user modifiable areal unit problem (UMAUP) has many more degrees of freedom than the classical MAUP and thus an even greater propensity to generate an even wider range of results than before.' The challenge is to discover how to turn this seemingly impossible problem into a useful tool for geographical analysis. What it means is that the same microdata can be given a very large number of broadly equivalent but different spatial representations. As a consequence, it is no longer possible to 'trust' any display or analysis of zone-based spatially aggregated data that just happens to have been generated for a zoning system. Users have to start seriously worrying about the nature of the spatial representations contained in the zoning systems they use. It is argued that this problem mainly affects socioeconomic information where, because of confidentiality constraints and lingering data restrictions, attention is often limited to the display and analysis of data that has been spatially aggregated one or more times.

The only alternative to an 'as is' spatial representation is to develop zone design as a spatial engineering tool. The zone design problem can be formulated as a non-linear constrained integer combinatorial optimisation problem. Openshaw (1977) defines this so-called automated zone design problem (AZP) as optimise  $F(X)$  where  $X$  is a zoning system containing an aggregation of  $N$  original zones into  $M$  regions ( $M < N$ ) subject to the members of each region being internally connected and all  $N$  zones being assigned to a region. The  $F(X)$  can be virtually any function that can be computed from the  $M$  region data: for instance, it could be a simple

statistic or a mathematical model that is fitted to the data. Typically,  $F(X)$  is non-linear, it could possess multiple suboptima, it need not be globally convex, and it is probably discontinuous (because of the contiguity and coverage constraints on  $X$ ).

Additionally, the user may wish to impose extra constraints either on the regions created by  $X$  (e.g. shape or size) and/or on the global data generated by the set of  $M$  regions (e.g. normality or spatial autocorrelation properties). This is achieved by converting  $F(X)$  into a penalty function. The Powell-Fletcher penalty function method has been found to be very effective (Fletcher 1987). This involves optimising the new function:

$$\Phi(X, \sigma, \theta) = F(X) + 0.5 \sum_i^N \sigma_i (C_i(X) - \theta_i)^2$$

where

$\Phi(X, \sigma, \theta)$  is a penalty function that is dependent on  $X$ ,  $\sigma_i$  and  $\theta_i$ ;

$\sigma_i$  and  $\theta_i$  are a series of parameters that are estimated to ensure gradual satisfaction of the constraints;

$C_i(X)$  is a constraint violation which is some function of the zoning system;

$F(X)$  is defined as previously.

This AZP was solved for small problems in the mid 1970s (Openshaw 1977, 1978). The technology was revived in the early 1990s when the increased availability of digital map data and the GIS revolution highlighted the importance of the problem. Larger datasets required better algorithms for optimising these functions. Openshaw and Rao (1995) compared three different methods and suggested that simulated annealing was the best choice. Unfortunately, they also found that simulated annealing took about 100 times longer to run than the other methods, while the use of additional constraints would have added another factor of 30 or so. Attempts were made to speed-up the simulated annealing approach by switching to parallel supercomputers. The immediate difficulty was the need for a fully parallel simulated annealing algorithm relevant to AZP types of problems. After considerable effort a hybrid simulated annealer with multiple adaptive temperatures controlled by a genetic algorithm was developed and was shown to work very effectively (Openshaw and Schmidt 1996).

It is now possible to use zone design to re-engineer all types of zoning systems with five principal areas of application:

- 1 to demonstrate the MAUP by seeking minimum and maximum function value zoning systems – this also helps prove the lack of simple minded objectivity in GIS;
- 2 to design zoning systems with particular properties that are believed to be beneficial for certain applications – for example, electoral re-districting or to minimise data confidentiality risks;
- 3 as a spatial analysis tool – for example, the zoning system acts as a pattern detector tuned to spot particular patterns;
- 4 as a visualisation tool – for example, to make visible the interaction between a model and the data it represents;
- 5 as a planning aid – for example, to define areas of maximum but equal accessibility or regions which are comparable because they share common properties.

These five application areas have been illustrated using the ARC/INFO-based Zone Design System (ZDES: Openshaw and Rao 1995; Alvanides 1995). The system attempts to routinise zone design using a number of generic zone design functions, and has been designed as a portable add-on to ARC/INFO.

## 7 ZONE DESIGN ANALYSIS OF SPATIAL DISTRIBUTIONS

This section provides an example of a geocomputation approach to zone design that demonstrates some of the potential capabilities of ZDES as a spatial analytical tool. Consider a problem that involves the analysis of non-white population data for England and Wales. This is currently of interest to some telecoms companies seeking to establish retail networks offering cut-price long-distance telephone calls. The 1991 Census data for persons born outside the UK may provide one surrogate indicator of this potential market demand for cheap long-distance calls. Figure 1(a) shows the 54 county zones that cover England and Wales while Figure 1(b) displays a choropleth map of the ethnic population. The key underlying geographical question is whether Figure 1(b) provides a 'meaningful' spatial representation of the Census data. Certainly, the visual patterns appear to identify

some areas of apparent concentrations. Yet distortions introduced by differences in county sizes may blur some of the patterns and diminish local concentrations by averaging them out. The map is the outcome of highly complex distortion of the data, caused by its aggregation into counties. Additionally, the use of counties as the object of study introduces an arbitrary geography, since there is no reason to suppose that it has any relevance whatsoever to the factors governing the distribution of ethnic communities in the UK. Indeed the principal attraction of the county zone is the convenience and ease of using a standard geography! Different zoning systems may be expected to offer different levels of data distortion, and some may tell a different story as a consequence. It is interesting, therefore, to explore some of the alternative patterns of the same 1991 Census data aggregated from the underlying 9522 census wards as an exercise in spatial analysis by zone design.

First consider what happens if an attempt is made to create a new set of 54 different ‘county’ regions that have approximately equal ethnic population counts, and then remap the ethnicity rates (see Figures 2(a) and 2(b)). The constant shading of Figure 2(b) shows the effects of the equal size function – the three small light zones are formed by contiguity islands in the data. One problem is the very intricate zonal boundary patterns formed by the ZDES optimiser as it attempted to create regions of equal ethnic population size in Figure 2(a). In fact, the more efficient the zoning system optimiser becomes so the more intricate are the resulting boundaries because of the imposed constraint of equal population sizes. The simulated annealer used here was very successful in optimising the function, but produced extremely crenulated and irregularly-shaped regions. Perhaps this also says something about the fine-scale spatial distribution of the ethnic population at the ward level – that is, the pattern is concentrated but discontinuous hence the need to link widely separated areas together in order to meet the population size restriction. It is a matter for further research as to whether there is an optimal spatial scale (or level of aggregation) at which the zones suddenly become more regular. This would be an interesting question to try to answer. Another problem here is that there are potentially many different zoning systems that will yield areas of nearly, or approximately, equal population size. It may also be necessary to introduce shape constraints.

A more useful spatial analytic function would be some measure of the spatial dispersion of the population around a set of region centroids. This is broadly equivalent to a type of large location–allocation problem. The objective is to minimise the global sum of the population weighted distances to each of the ward centroids. This function is expressed as follows:

$$\text{minimise} \quad \sum_j^M \sum_{i \in j}^N P_i D_{ij}$$

where

$P_i$  is the population value for ward  $i$ ;

$D_{ij}$  is the distance from ward  $i$  to the population centroid of region  $j$  of which  $i$  is a member. (Note that this centroid depends on the current membership of region  $j$ .)

The restriction on the second summation indicates that the summation only occurs for wards that belong to region  $j$ . This is a way of partitioning the zoning system of  $N$  Census wards into  $M$  regions.

Figure 3 shows the results of minimising this local spatial dispersion function. This map is the outcome of a contest between different parts of the UK, as the ZDES algorithm seeks to trade-off population gains in some parts of the country against losses in others. The resultant zoning system is thus a visualisation of the tension (or interaction) between the objective function, the Census ward zonation and aggregation effects. Those parts of the UK with the largest ethnic populations have relatively small regions. The map in Figure 3(b) seems to offer a more sensitive view of the distribution of ethnic populations compared with Figure 1(b), and it highlights some of the areas where aggregation effects have seemingly removed some of the patterns. This approach is developed further in Figure 4(a), which shows the results for the same objective function but subject to constraints that the total population of all the regions should be at least 75 per cent of the average. This involves solving a penalty function version of ZDES. On this application the simulated annealer needed a Cray T3D to produce the results within a convenient time – that is, one hour. The patterns in Figure 4(b) are considerably more disjointed than in Figure 3(b) with more intricate boundary resolution in the principal areas of ethnic concentration.

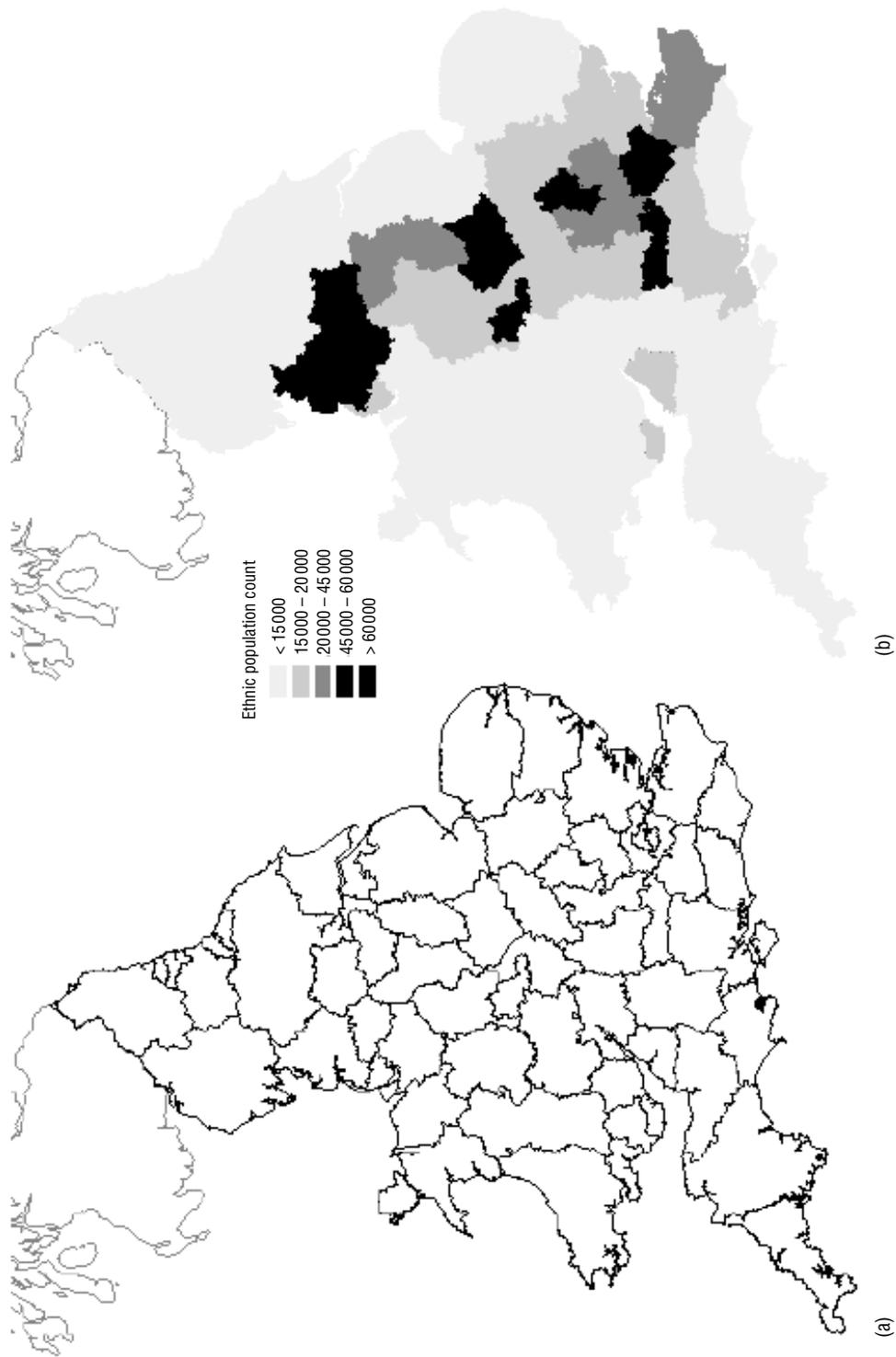


Fig 1. (a) The 54 counties in England and Wales; (b) distribution of ethnic population for counties in England and Wales.

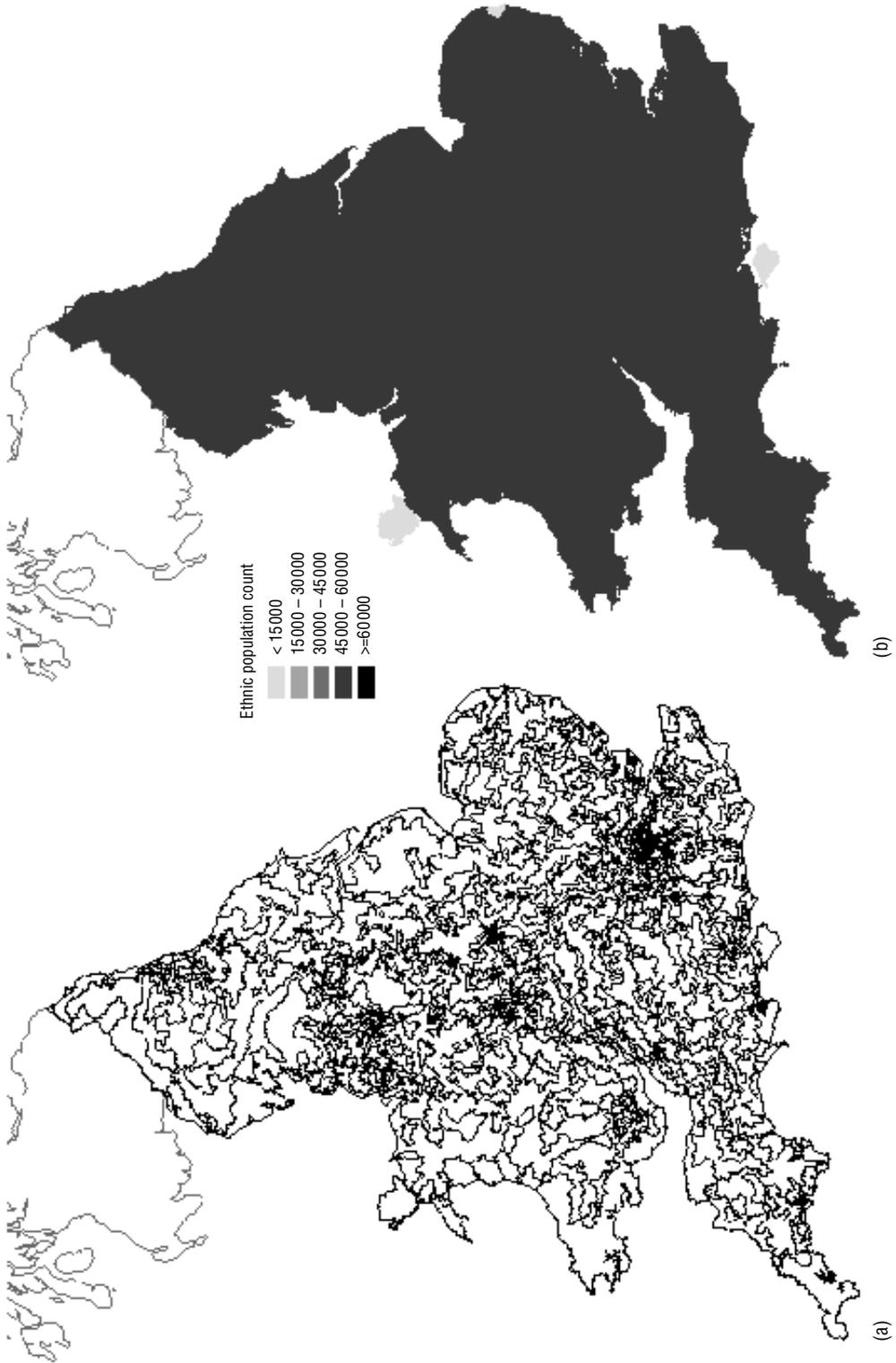


Fig 2. (a) Equal ethnic population regions in England and Wales; (b) distribution of ethnic population in England and Wales for equal population regions.

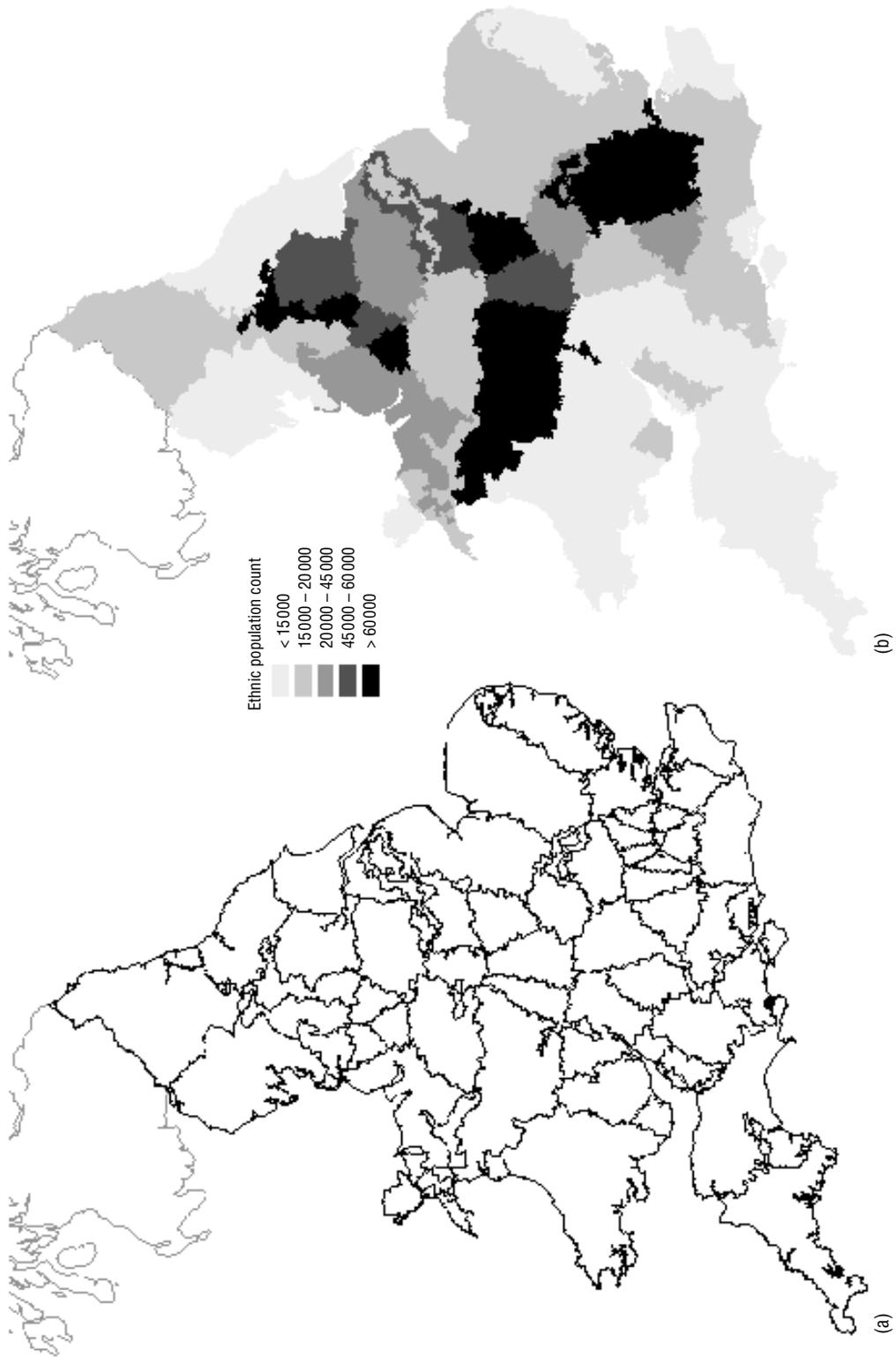


Fig 3. (a) Regions in England and Wales that minimise population weighted distances; (b) distribution of ethnic population for accessibility regions.

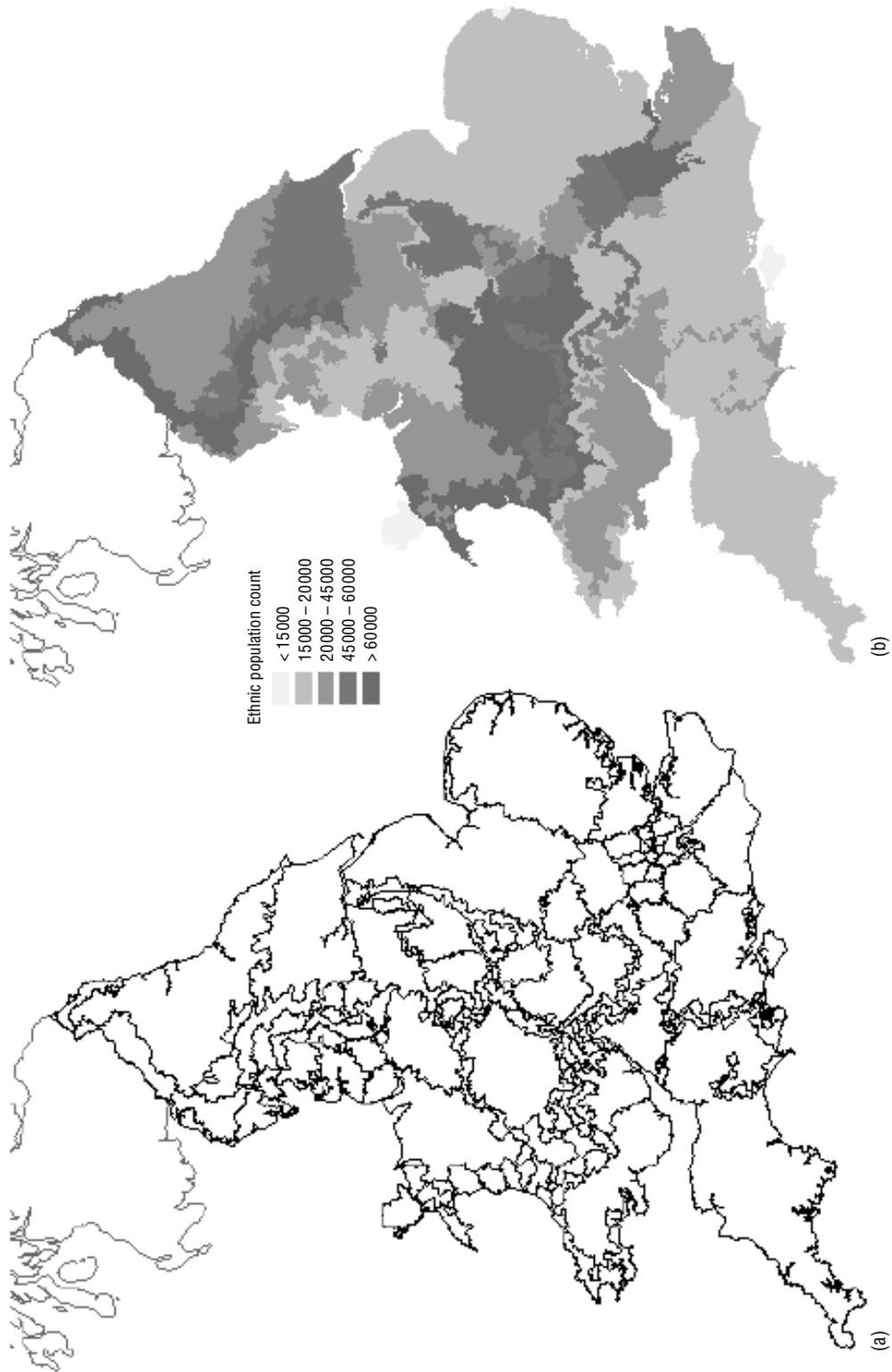


Fig 4. (a) Regions in England and Wales that minimise constrained population weighted distances; (b) distribution of ethnic population for constrained accessibility regions.

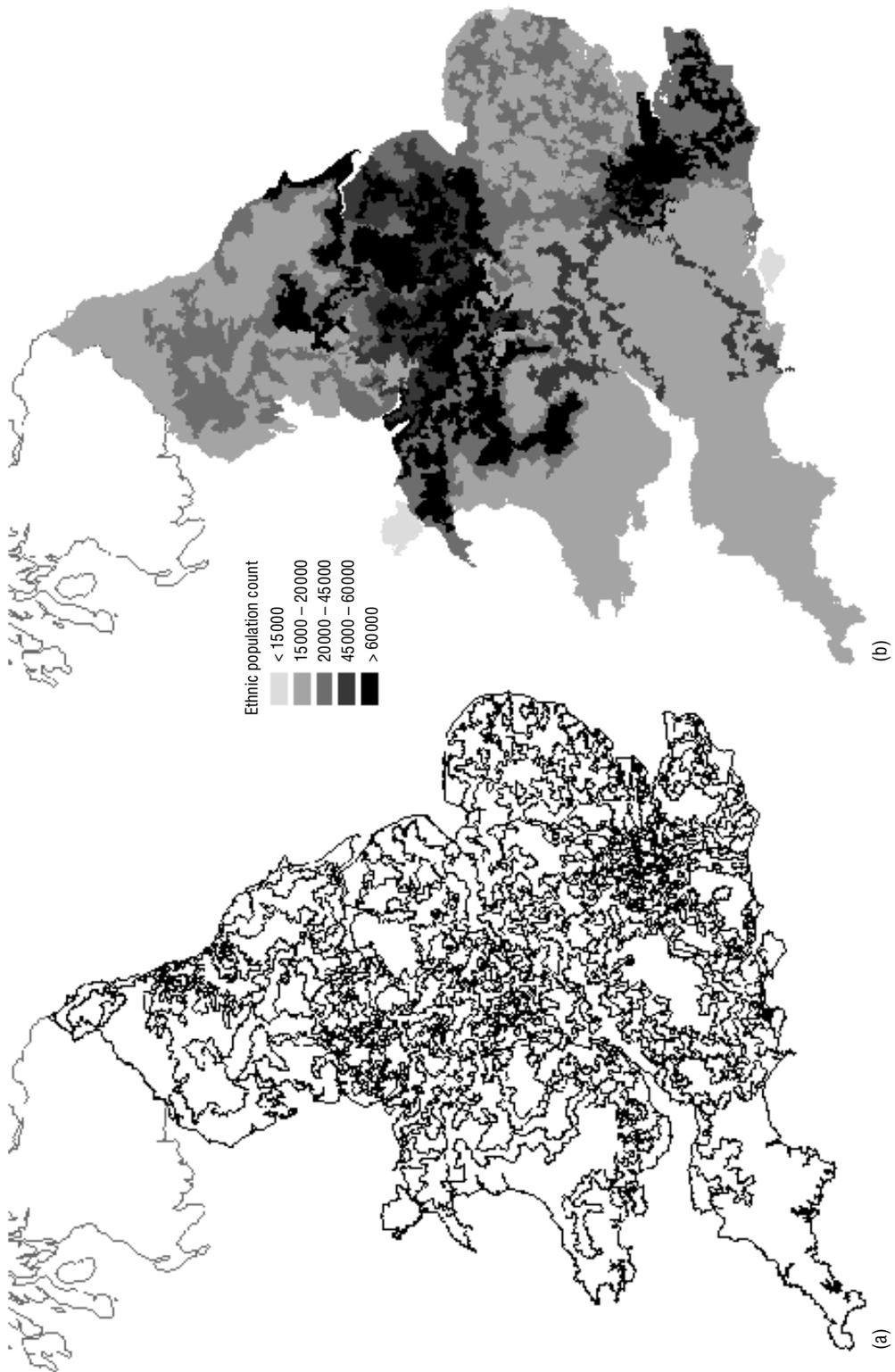


Fig 5. (a) Regions in England and Wales that minimise a constrained spatial interaction model; (b) distribution of ethnic population for constrained spatial interaction regions.

A final demonstration involves converting the previous objective function into a measure of population potential as a form of crude spatial interaction model. The function is as follows:

$$\text{minimise} \quad \sum_j^M \sum_{i \in j}^N P_i D_{ij}^{-2}$$

subject to minimum size constraints. The resulting zonation is shown in Figure 5(a). The region boundaries are once again very crenulated, especially in areas of high ethnic populations. The resulting population distribution is mapped in Figure 5(b). Once again the underlying impression is that of local area concentrations surrounded by gaps. This may be a reflection of areas of localised negative spatial autocorrelation. The deficiency of the administrative scheme, represented here by the county geography, is demonstrated by a series of descriptive statistics. Table 5 shows some statistics concerning the ethnic population per zone for each of the different models illustrated earlier. The administrative county geography scores worse than any other scheme with zones extending beyond 700 000 ethnic residents and an extreme standard deviation figure (113 750). The equal ethnic population geography demonstrates the best statistical properties for a representative administration scheme, with a maximum ethnic population very close to the mean ethnic population value (54 650 residents). However, the boundaries of the zones are rather obscure for any policy making exercise and the same problem occurs with a constrained interaction geography, as we saw in Figures 3(a) and 5(a) respectively.

A trade-off between a comprehensible geography and satisfactory statistical properties is the constrained accessibility model illustrated in Figure 4(a). In this case the weighted distance minimisation function works as a shape constrained thus producing relatively compact zones, while the population constraint function restricts the maximum population of zones to

**Table 5** Statistical properties of different geographies for ethnic population.

<i>Geography</i>	<i>Maximum</i>	<i>Mean</i>	<i>Standard deviation</i>
Administrative counties	703 700	46 850	113 750
Equal ethnic population	58 650	54 650	13 200
Population accessibility	320 900	54 650	74 700
Constrained accessibility	176 200	55 700	39 850
Constrained interaction	273 200	55 700	51 500

176 200 and retains the standard deviation low enough for the zones to be comparable. Possibly the 'best' zoning systems that would represent this type of socioeconomic data would be very different from the conventional administrative zoning scheme and this forms the subject of ongoing research (Openshaw and Alvanides 1997b).

## 8 CONCLUSIONS

The case for developing a wide range of GRSA tools is very compelling. GIS is creating an immensely data-rich environment. The technology for spatial data capture, management, and handling has far outstripped the available tools for its analysis. The hope now is that it may be possible to compute our way out of the data swamp by developing new generations of intelligent spatial analysis tools which are better able to cope with the conflicting requirements of large volumes of data, geographical reasonableness, and the endusers. The need is undeniable and there are various ways forward involving the use of zoning systems as data, pattern, and model visualisers; and the development of smart GIS database explorers. The technological basis for these exists: HPC, AI toolkits, computational statistics, large spatial databases, and well-developed GIS. Powerful zone design algorithms have been developed. Artificial life-based 'creatures' or 'agents' can be created that are able to move around space-time-attribute GIS databases under their own control in an endless search for patterns and relationships of possible interest. Computer animation provides the basis for users 'watching' what is happening in highly complex geo-cyberspaces projected on to a 2-dimensional map (Openshaw and Perrée 1996). The basic methods will probably run on a workstation, but as the complexity of the data domain increases or as greater use is made of computational statistics in order to improve performance, so they will need HPC. There is a suggestion that the evolution of new types of spatial analysis technology is about to start.

The challenge then is to solve the principal outstanding spatial analysis problems by developing various geocomputational approaches, to demonstrate they work on a range of generic problems, and then to make them available either within or without current GIS software. This task is becoming increasingly urgent and requires geographers (in particular) to use their understanding of the geography

of the problems to create a new approach to developing geographical analysis methods. A start is being made (here as well as elsewhere in this book) but much still remains to be done.

## References

- Alvanides S 1995 'The investigation of a Zone Design System for reconstructing census geographies'. Unpublished dissertation 3031 available from School of Geography, University of Leeds
- Besag J, Newell J 1991 The detection of clusters in rare diseases. *Journal of the Royal Statistical Society A* 154: 143–55
- Bezdek J C 1994 What is computational intelligence? In Zurada J M, Marks R J, Robinson C J (eds) *Computational Intelligence: Imitating Life*. New York, IEEE: 1–12
- Birkin M, Clarke G P, Clarke M, Wilson A G (eds) 1996 *Intelligent GIS: location decisions and strategic planning*. Cambridge (UK), GeoInformation International
- Fletcher R 1987 *Practical Methods of Optimisation*. Chichester, John Wiley & Sons
- Fotheringham A S, Wong D W S 1991 The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23: 1025–44
- Fotheringham A S, Zhan F B 1996 A comparison of three exploratory methods for cluster detection in spatial point patterns. *Geographical Analysis* 28: 200–18
- Hillis W D 1992 What is Massively Parallel Computing and Why is it important? In Metropolis N, Carlo Rota G (eds) *A new era in computation*. Cambridge (USA), MIT Press: 1–15
- Openshaw S 1977 A geographical solution to scale and aggregation problems in region-building, partitioning, and spatial modelling. *Transactions of the Institute of British Geographers* 2: 459–72
- Openshaw S 1978 An optimal zoning approach to the study of spatially aggregated data. In Masser I, Brown P J B (eds) *Spatial representation and spatial interaction*. Boston (USA), Martinus Nijhoff: 95–113
- Openshaw S 1984 *The modifiable areal unit problem*. Concepts and Techniques in Modern Geography 38. Norwich, Geo-Books
- Openshaw S 1991c Developing appropriate spatial analysis methods for GIS. In Maguire D, Goodchild M F, Rhind D (eds) *Geographical information systems: principles and applications*. Harlow, Longman/New York, John Wiley & Sons Inc. Vol. 1: 389–402
- Openshaw S 1994a A concepts-rich approach to spatial analysis, theory generation, and scientific discovery in GIS using massively parallel computing. In Worboys M F (ed) *Innovations in GIS 1*. London, Taylor and Francis: 123–38
- Openshaw S 1994b Computational human geography: towards a research agenda. *Environment and Planning A* 26: 499–505
- Openshaw S 1994d GIS crime and spatial analysis. *Proceedings of GIS and Public Policy Conference*. Ulster, Ulster Business School: 22–35
- Openshaw S 1994e Two exploratory space-time-attribute pattern analysers relevant to GIS. In Fotheringham A S, Rogerson P (eds) *Spatial Analysis and GIS*. London, Taylor and Francis: 83–104
- Openshaw S 1994f What is GISable spatial analysis? In *New Tools for Spatial Analysis*. Luxembourg, Eurostat: 157–62
- Openshaw S 1995 Developing automated and smart spatial pattern exploration tools for geographical information systems applications. *The Statistician* 44: 3–16
- Openshaw S 1996a Developing GIS relevant zone based spatial analysis methods. In Longley P, Batty M (eds) *Spatial analysis: modelling in a GIS environment*. Cambridge (UK), GeoInformation International: 55–73
- Openshaw S 1997 Supercomputing in Geographical Research. *Proceedings of Conference on IT in the Humanities*. London, British Association
- Openshaw S, Alvanides S 1997 Designing zoning systems for representation of socioeconomic data. In Frank I, Raper J, Cheylan J (eds) *Time and motion of socioeconomic units*. GISDATA series. London, Taylor and Francis
- Openshaw S, Charlton M, Wymer C, Craft A 1987 A mark I Geographical Analysis Machine for the automated analysis of point datasets. *International Journal of Geographical Information Systems* 1: 335–58
- Openshaw S, Craft A 1991 Using Geographical Analysis Machines to search for evidence of clusters and clustering in childhood leukaemia and non-Hodgkin Lymphomas in Britain. In Draper G (ed) *The geographical epidemiology of childhood leukaemia and non-Hodgkin lymphomas in Great Britain, 1966–83*. London, Her Majesty's Stationery Office: 89–103
- Openshaw S, Cross A 1991 Crime pattern analysis: the development of Arc/Crime. *Proceedings of AGI Annual Conference*. Birmingham (UK)
- Openshaw S, Cross A, Charlton M 1990 Building a prototype geographical correlates exploration machine. *International Journal of Geographical Information Systems* 3: 297–312
- Openshaw S, Fischer M M 1995 A framework for research on spatial analysis relevant to geostatistical information systems in Europe. *Geographical Systems* 2: 325–37
- Openshaw S, Openshaw C A 1997 *Artificial intelligence in geography*. Chichester, John Wiley & Sons
- Openshaw S, Perrée T 1996 User-centred intelligent spatial analysis of point data. In Parker D (ed.) *Innovations in GIS 3*. London, Taylor and Francis: 119–34
- Openshaw S, Rao L 1995 Algorithms for re-engineering 1991 Census geography. *Environment and Planning A* 27: 425–46
- Openshaw S, Schmidt J 1996 Parallel simulated annealing and genetic algorithms for re-engineering zoning systems. *Geographical Systems* 3: 201–20
- Openshaw S, Schmidt J 1997 A Social Science Benchmark (SSB/1) Code for serial, vector, and parallel supercomputers. *International Journal of Geographical and Environmental Modelling*
- Turton I, Openshaw S 1996 Modelling and optimising flows using parallel spatial interaction models. In Bougé L, Fraigniaud P, Mignotte A, Roberts Y (eds) *Euro-Par 96 Parallel Processing Vol. 2 Lecture Notes in Computer Science* (1124). Berlin, Springer: 270–5