WSG Diskussionspapiere

WSG 44/95

A Framework for Research on Spatial Analysis Relevant to Geo-Statistical Informations Systems in Europe

Manfred M. Fischer and Stan Openshaw

Institut für Wirtschafts- und Sozialgeographie

Wirtschaftsuniversität Wien

Department of Economic and Social Geography

Vienna University of Economics and Business Administration

WSG Discussion Papers

Abteilung für Theoretische und Angewandte Wirtschafts- und Sozialgeographie
Institut für Wirtschafts- und Sozialgeographie
Wirtschaftsuniversität Wien

Vorstand: o.Univ.Prof. Dr. Manfred M. Fischer
A - 1090 Wien, Augasse 2-6, Tel. (0222) 313 36 - 4836

Redaktion: Mag. Petra Staufer

WSG 44/95

A Framework for Research on Spatial Analysis Relevant to Geo-Statistical Informations Systems in Europe

Manfred M. Fischer and Stan Openshaw

WSG-Discussion Paper 44

February 1995

Gedruckt mit Unterstützung
des Bundesministerium
für Wissenschaft und Forschung
in Wien

WSG Discussion Papers are interim reports presenting work in progress and papers which have been submitted for publication elsewhere.

ISBN 3 85037 049 6

Abstract
The paper emphasises the importance of a research programme focused on developing and making widely available GIS relevant spatial analysis technology. It outlines generic criteria able to discriminate between GIS-relevant and GIS-irrelevant spatial analysis tools and outlines a list of six researchable spatial analysis themes. It is argued that presently there is an opportunity to develop a EU based spatial analysis research programme and then install the technology in the World's GIS.

Keywords
Spatial Analysis, GIS, AI, research programme

INTRODUCTION

There are several arguments why a spatial perspective in data analysis might be considered helpful (Goodchild et al. 1992). First, space provides a simple, but very useful framework for handling large amounts of data. Second, the spatial perspective permits easy access to information on the relative location of objects and events. Third, it allows spatial data to be linked, in a process formalized in GIS as an overlay. Finally, the distance between objects and events reveals is often an important factor in interactions between them, in both environmental and socio-economic applications. As a result the ability of geographic information systems (GIS) to handle and analyse spatial data is usually seen as a major useful characteristic that distinguishes GIS from other types of information systems; for example, those SQL types of system developed to serve the needs of business data processing, or computer aided design systems or systems whose primary objective is map production rather than the full range of GIS functionality.

Geographic Information Systems incorporate many state-of-the-art tools such as relational database management, powerful graphics algorithms, elementary spatial operations, polygon overlay with logical operations, interpolation, zoning and simplified network analysis. However, there are some important areas of weakness. In particular what is termed spatial analysis and modelling is often no more than map data manipulation such as polygon overlay and buffering. Indeed the lack of much sophisticated analytical and modelling functionality is widely recognised as a major deficiency of current systems (see, for example, Openshaw 1990, 1991a, 1991b, Fischer and Nijkamp 1992a, b, Anselin and Getis 1992, Fotheringham and Rogerson 1993). This deficiency continues even though there is some broad agreement that the future success of the GIS technology will depend to a some arguably large extent on incorporating more powerful analytical and modelling capabilities that increasingly focus on the use of geographic information rather than its capture, storage, and manipulation. Indeed in the post GIS revolution 1990s the emphasis will increasingly be focused on the use of geographic information. The current neglect of this aspect is causing some concern (Openshaw, 1994a). The need for GIS relevant and powerful spatial analysis tools has never been greater.

the EU, the European Science Foundation (ESF) GISDATA Research Programme has targeted spatial analysis as one of about 16 key areas for attention. There would seem to have been a massive amount of research performed on spatial analysis in the last 5 years, but in reality there has been very little explicit activity and few new methods have emerged.

This is perhaps not too surprising because none of the well funded research focused on spatial analysis per se, rather than on a much broader agenda. For instance, the RRL Initiative's total investment in spatial analysis research probably amounted to about 2 or 3 person years of effort. The NCGIA in its I-14 initiative had a budget of about £35,000, spent mainly raising awareness of the problems and need for spatial analysis tools, and nothing on research. The ESF Exploratory Spatial Analysis and Modelling initiative was limited to one three day workshop, it is not directly sponsoring research, although like the NCGIA, there will be a book to further raise awareness of the problems and the need for some real research. The urgency now is to build on these researches, develop useful tools, and deliver them in a usable form to the end users. However, a major difficulty at present is that there is no organised spatial analysis research programme either on a national or international scale at present, or indeed probably anywhere in the world dedicated to this task. Moreover, the subject does not appear on any of the active research agendas; for instance, in the UK, the Analysis of Large and Complex Databases initiative (ALCD) has no spatial analysis component. What research is being done, is being done by individuals, in an uncoordinated manner, via a plethora of different and unrelated projects few of which are explicitly concerned with spatial analysis.

As a result it is quite understandable why the GIS vendors have so far failed to put much spatial analysis functionality into their systems. They are clearly aware of this neglect but currently still do not appear to believe there is any significant user demand for it, or many seem confused as to what methods are appropriate, or are concerned that they do not have the high level of statistical skills they seem to believe is a fundamental pre-requisite to support it. They also seem to hold the view, at present, that if users want to perform statistical analysis, then all they need do is export their data to a statistical package or link a statistical package (or a statistics friendly programming language) into their GIS. Clearly, this is better than nothing but it does little to address the need for useful spatial analysis methods available for end-users in applied GIS environments rather than academic contexts.

What is special about spatial data?

It is helpful to understand that spatial data is not the same as non-spatial data; for example, a sample survey. There are a number of important differences; see Table 1. It might also be noted that cross-national spatial data series usually have additional problem features; due to the existence of many different data sources. This can cause non-random, spatial structured, differences in data reliability; and these harmonisation derived uncertainties need to be handled rather than ignored. GIS technologies do little or nothing to help. It is already fairly easy to use a GIS to mix multiple data sets regardless of their source levels of resolution and accuracy. Errors and uncertainties added by

GIS RELEVANT SPATIAL ANALYSIS INITIATIVES

The origins of spatial analysis lie in the development of quantitative geography and regional science in the early 1960s. This involved the use of quantitative (mainly aspatial statistical) procedures and techniques to analyse patterns of points, lines, areas and surfaces depicted on analogue maps or defined by coordinates in 2- or 3-dimensional space. Later more emphasis was placed on the special features of geographical space, on building mathematical models of spatial processes and on studying the spatio-temporal evolution of complex spatial systems. Despite a very large number of rather diverse contributions in this area - two main themes can be identified (see Fischer and Scholten 1993):

(1) statistical spatial data analysis focusing on exploring geographical patterns and relationships, and
(2) spatial process modelling in the context of spatial decision support systems.

The emphasis here is on exploring the nature of the spatial data analysis task considered relevant to GIS. Spatial data analysis methods are important and becoming even more so as the supply of spatial information has developed from a trickle 20 years ago into a flood. Indeed now that Geographical Information Systems is an established technology the supply of spatial data is rapidly increasing in virtually every area in which it is relevant; this includes both terrestrial and space systems. As a result more and more users of spatial information are finding themselves having to function in increasingly data rich environments. Indeed there is already in some areas evidence of considerable data overload as faster, bigger, and better statistical information systems provide ready access to more and more spatial data from multiple sources that are now being linked by GIS. An increasingly common response is to simply not analyse data unless there is a strong imperative for analysis but, even when there is a strong imperative for analysis, there are often few relevant, available, and appropriate analysis tools. Information management, and storage technologies are far outstripping our tools for spatial analysis, resulting in data-drenched users having to operate in a "data swamp" without adequate tools. Yet where the tools exist, they are often not in a form that is widely diffused and usable. This is partly a matter of finding ways to help people cope with and use the current technology via education and awareness, but there is also an urgent need to provide the tools and systems that they can use, which are relevant, which will yield reliable results, and which are understandable. Of course none of these general problems are unique to any particular country although the nature of official European data statistics often creates some novel additional difficulties, that are only recently becoming apparent.

There has been many workshops and conferences on spatial analysis and GIS in the last five years and the topic has appeared as a major theme in at least three different GIS Research Programmes. In the UK, the Economic and Social Research Council (ESRC) Regional Research Laboratory Initiative (1987-91) targeted spatial analysis as one of several priority areas within a broad spectrum of GIS projects. In the USA, the National Centre for Geographic Information and Analysis (NCGIA) (1987 onwards) had a spatial analysis initiative (I-14, their 14th different research initiative in 4 years). In geoprocessing are simply lost. Somehow the end-user has to either ignore them or try and estimate what they might be. This is poor science.

Table 1 Major Characteristics of Spatial Data

Many cases/objects/points
Many variables
Large data volumes
Spatial autocorrelated values
Spatial data error of various types
Non conformity with standard statistical distributions
Data precision can be spatially structured
Errors need not be random
Not samples in usual sense
Surrogate variables abound
Nonlinearity is the norm
A high degree of complexity
Modifiable areal unit, scale and aggregation effects
Mixtures of measurement
Small number problems can be important

Some of the special features of spatial data render classical statistical methods unsafe unless they have been modified to handle the problems. However, producing modified classical statistical tools is still not sufficient because the analysis needs in the GIS era are not the same as that of a statistician interested in the analysis of a random sample of households. Many of the spatial analysis concerns are related to applications far removed from classical statistical methods such as t-tests, correlation and regression. Moreover, many of the end-users of GIS are not statisticians or trained spatial scientists. Yet it is estimated that by 2000AD there could be over one million people working with spatial information systems of which the vast majority will not be experts in statistical or geographical analysis. It is important, therefore, that some fresh effort be devoted to developing applied spatial analysis tools that are appropriate to the needs of GIS.

What types of spatial analysis are GIS relevant?

The principal challenge is to identify and develop generic spatial analysis tools which are appropriate for use with spatial data in GIS environments. Table 2 shows Openshaw's 10 basic GIS ability criteria that spatial analysis methods should ideally attempt to meet. It is important to recognise that GIS creates its own spatial analysis needs, it is not simply a source of data that can be analysed using more classical statistical methods. These needs make it a special and different subfield of spatial statistics. It is within this context that spatial analysis tools need to be either regenerated, or rediscovered, or newly created. The present time is a good moment in which to tackle these problems. There is a unique window of opportunity to develop in Europe globally useful and relevant spatial analysis technology that can subsequently be installed in the world's GISs, which are incidentally, nearly all made in the USA. Many of the world's leading and most experienced applied spatial analysts are located in the EU. Here is an opportunity to sell made in Europe spatial analysis technology relevant to the global user-base of GIS.

Table 2 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. Results are mappable
5. Generic analysis procedures
6. Useful, and valuable
7. Interfacing problems are irrelevant
8. Ease of use and understandability are important
9. Safe technology
10. Applied technology

RESEARCH CONSTRAINTS

In formulating a list of researchable spatial analysis themes relevant to the EU the following organisational and political constraints are understood to exist:

1. the research should be of a form that cannot be done at the national level and address issues of pan-EU significance;
2. it should address key user problems, but not exclusively so;
3. there is no restriction on the nature of the technology being proposed so it could include AI or purely statistical methods;
4. commercial collaboration is important; and
5. the themes should attract interest from, and be doable by, several different research groups in two or three different EU countries, preferably one of which should be a recent entrant.

The seven research themes discussed in the rest of the paper attempt to meet most of these constraints. The "constraints" also demonstrate why it might be considered "hard" to do EU research at present, as distinct from networking. It is also important to take a strategic and long term view by defining a broad framework at least for the rest of the 1990s.

A LIST OF RESEARCHABLE SPATIAL ANALYSIS THEMES

Participants at the DOSES workshop on 'New Tools for Spatial Analysis', Lisbon, November 18-20 1993, were asked to think about what research themes might be most useful in the spatial analysis area. The following suggestions emerged after several hours of discussion debate spread over a three day period. A summary of the themes is given in Table 3.

Table 3 Spatial analysis research themes

1. Tool-kit(s) for spatial analysis in GIS
2. Methods and tools for handling uncertainty in spatial data
3. New methods of spatial analysis
4. New styles of modelling and statistical aspects of model evaluation and choice
5. Confidentiality of spatial data and zone design
6. Impediments to the development and use of spatial analysis methods

Theme 1: Tool-kits for spatial analysis in GIS

There is an increasingly urgent and very obvious need to package and hence commodify the available spatial analysis methods which currently exist mainly as research tools. Despite the problems of analysing spatial data there are a number of existing methods that can be usefully applied. The Spatial Analysis Tool-kit version 1 proposal lists a number of these methods; see Table 4; although it is possible that due to no funding SAT/1 will never exist in a full and final form. Indeed, it is impossible to build a comprehensive spatial analysis tool-kit without extensive collaboration between a network of different groups of researchers (who have specialist techniques) and GIS vendors (who know how to interface them). No single country's spatial analysis experts probably cover more than a quarter of the available methods that might be usefully included; indeed, the SAT/1 contents shown in Table 4 has had to involve four different universities. It represents only about 20% of what might be expected to be in a useful SAT, but just about exhausts what the UK quantitative geographic community could provide from the shelf of available and relevant technology.

It is also apparent that, whilst it is fairly obvious to academics that users will need methods that can spot patterns, find relationships, detect changes, model spatial phenomena, and simplify complex data, it is not obvious as to what the users do really need. Some market research of what tools might be needed by users in different application areas is urgently required. What precisely are the user's demands from spatial analysis? What does spatial analysis mean to different groups of people? Is there need for single tool-kit full of generic tools or many different application specific tool-kits? How should they be packaged? It is also likely that what users want will reflect what they think can be done, and that, therefore, a certain amount of toing-and-froing as well as awareness raising will be needed. Convincing demonstrations and case studies will also be extremely useful to overcome organisational and motivational problems in deploying the technology.

Table 4 : Proposed SAT/1 contents as an example subset of available spatial analysis tools

Diggles point density method	
A simple GAM	Besag-Newell method
Cartograms	
Kernel estimated surfaces	
Stones method	
Getis-Ord g statistics	
Zone design methods	
Regionalisation and classification	
Zone ranking	
Spatial regression modelling	

Theme 2: Methods and tools for handling uncertainty in spatial data

All data are wrong but, apart from sampling error, no tools exist in a usable form to help people cope with the analysis of uncertain spatial data. The uncertainty can result from various sources: data collection, operations performed on the data including geoprocessing and harmonisation, missing data, and wrong data. The NCGIA I-1 initiative raised awareness of many of the problems, (see Goodchild and Gopal, 1989) but provided no solutions.
There are various aspects that are important here:

1. raise awareness of errors and uncertainty amongst the user communities;
2. provide methods able to estimate levels of data uncertainty in spatial information sources stored without it;
3. develop methods to estimate the effects of error propagation that impact on the results of spatial analysis, and decision support; and
4. create new tools to help users cope with uncertain information.

The problem is complicated because whether spatial data errors and data uncertainty matters is context dependent. Data need not be 100% accurate before it can be safely used. Some data that may be highly inaccurate may still be quite sufficient for one application but not another. Equally, the problem can be reversed by asking how accurate (or what level of uncertainty) is adequate for a given application. If we could find out, then considerable money might be saved by gathering data of a quality sufficient for the intended purposes.

This discussion also raises the question as to how best to analyse data which are either known to be poor or even wrong. Instances of the analysis of wrong data are not rare; for example, in computing unemployment rate for 1989 using census economically active data for 1981. How should users cope with this and related problems? Best practice needs to be defined and basic standards created.

Theme 3: New methods of spatial analysis

This is another very important area where urgent research and development is needed if good value is to be extracted from many spatial information systems. Most of the available analysis tools pre-date GIS and come from an era when data sets were small, computers were very slow and small in memory capacity, and there were almost as many skilled statisticians as there were data sets to be analysed. Moreover, analysis was often fairly basic conducted at a slow, even leisurely, pace. In the 1990s the world of spatial information has changed in a very fundamental way and new tools are now desperately needed to cope with both old and new problems of analysis.

A number of themes can be identified. They include the following:

1. Computer tools for exploring GIS databases which contain spatial, temporal, and multivariate attribute information. Methods for this task range from highly interactive map based exploratory systems to automated data sievers which seek to identify potential database anomalies for subsequent in depth study, to smart search tools that automate the exploration process (see Openshaw (1992, 1994b, c). In all cases the methods should be designed to provide systems which complement rather than replace human skills and abilities.

2. Tools to explore the effects of spatial planning and consequences of scenarios. It should be possible to develop sensitivity testers that try to identify the "faults" or "weaknesses" in spatial planning by employing search strategies that try to make them fail. This would be the GIS equivalent to fault-tree analysis and related tools used in engineering reliability studies.

3. Geographical analysis performed in a highly visual manner. Picture icons that perform specified spatial analysis and modelling functions can be developed to simplify use problems. A spatial analyst equivalent to a graphical user interface (GUI) based on state of the art hypertext technologies could provide a different non paper based approach to information display.

4. Application of computer vision, multi-media, animation, and pattern recognition methods to a spatial analysis domain (see Dorling and Openshaw, 1991, Openshaw et al (1994d). This is potentially a major research area. Scientific visualisation is well established. However, here the system being visualised is not a model but the dynamical aspects of a spatial system as represented by a spatial database. Unfortunately the task is not easy. The simple animation of maps in a time sequence often confuses because there is too much information. Instead the data

being animated in a spatial multi-media context needs to be smoothed or filtered or even converted into idealised representations prior to its animation. The benefits of visualisation are considerable; for example, decision makers could be presented with movies of the change process in their region of interest. It is an excellent communication tool. It may also be an insightful spatial analysis tool if appropriate picture processing can be applied to the data to render visible the principle processes.

5. Virtual reality representations of the geocyberspaces that are being created. The cyberspace concept of a world of information flows can be extended into a geographic domain. The geocyberspace is seen by some as being the artificial reality that will become the tramping ground of the 21st century geographer (Openshaw, 1994e). GIS technologies can only handle part of these new information spaces but its extension into space-time will, if developed, handle the remainder. It is important, therefore, to develop technologies that attempt to understand the data and develop views of the information or database world in which attempts are made by our clumsy spatial analysis tools to find patterns and relationships. Can we look inside spatial databases and walk around in them via virtual reality concepts? Currently we are so blind to many aspects of the data and the data flows being analysed. We cope by being highly selective and subjective and in the process probably fail to see and find many of the patterns and relationships of potential interest other than those we blindly stumble over by chance!

Theme 4: New styles of data driven modelling such as neurocomputing

Neurocomputing is just one new methodological approach that is attracting interest. Neurocomputing is predominantly concerned with a non-programmed adaptive information systems, called artificial neural networks. These models attempt to achieve good performance via networks of simple computing elements, highly interconnected and running in parallel. (Hewitson and Crane, 1994). Neurocomputing methods are particularly appealing partly because of their biological inspiration and mainly because they extend the capabilities of computer analysis into applications previously found too hard to handle. In a spatial analysis context it is the abilities of supervised neural nets to provide useful models of virtually any dynamic process, no matter how nonlinear or chaotic it may be, for which there is adequate and representative data. Other types of unsupervised neural net architectures are also extremely useful as potential spatial data classifiers. In both cases they are well suited to handle noisy data and provide the basis for a superior level of data analysis technology. (Openshaw, 1992).

Fischer and Gopal (1994) have demonstrated the feasibility of the two-layered feedforward neural network approach with sigmoidal processing units to model interregional telecommunication traffic in Austria and evaluated its performance in comparison with the classical regression approach of the gravity type. Similarly Openshaw (1993, 1993a) compares neural net spatial interaction models with conventional entropy maximising models. The application of this neural network approach may be viewed as a three-stage process. The first stage refers to the identification of an appropriate network from the family of multi-layered feedforward networks. The second stage involves the estimation of the network parameters of the selected neural network model, and the final stage to the testing of the prediction performance. The analysis shows that the neural network model may well outperform the classical approach. The task now is to mold these powerful tools into a GIS relevant tool-kit. Maybe the technology should be saved for the hard problems that more conventional methods cannot handle, rather than replicate what current methods can do. Substitution is not by itself sufficient, unless there is also some value added. There is a risk with neurocomputing of both over-enthusiasm and also deliberate complexity added to preclude its use; for example, by converting what is essentially a simple technology into a form that only a few experts can understand. In a GIS context, it is not the neurocomputing tool-kits that are needed but new generations of spatial analysis tool that have neurotechnology embedded somewhere within them.

Theme 5: Statistical aspects of model evaluation and choice

There is an urgent need to discover how best to evaluate, compare, and then choose the best method or model from a list of alternatives. The increasing availability of AI and neurocomputing substitutes for more conventional statistical technologies makes this task increasingly urgent. Many of the new methods are "brittle" and have application specific levels of performance. We need to use the tools we have as challenges to the new tools we are still developing. As automation of the analysis process increases so we need to ensure that safety, reliability, and robustness are important factors. (Openshaw, 1992a).

It is important to be confident that the best obtainable results have been achieved in any particular application. One way of achieving this goal is to build analysis machines that investigate all reasonable models , and compare their levels of performance and reliability (Openshaw 1994f, g). The old models or methods are used as benchmarks against which to compare the new ones. The questions is how to find the best, or model. What statistical methods are best employed for doing this. Are there benefits in combining results, what cross model validation methods are most useful, and how can unacceptable levels of performance be identified.

Theme 6: Confidentiality of spatial data and zone design

Most statistical data are reported for zones of some kind. It is well known that different zoning systems (eg NUTS1, or NUTS2) will result in different representations of the same data (Openshaw, 1978, 1984). This zoning system scale and aggregational dependency influences all the results of spatial data analysis. Moreover when there are systematic national differences in the nature of the zones, it may well benefit or disadvantage some EU countries more than others. For example, a zone small enough and homogenous enough to represent a high unemployment area will have a high unemployment rate. However, if the equivalent zone in another country or region is much larger than the high unemployment area and also includes some low employment areas, then the unemployment rate can become average. A further problem is that when data for different sizes of zone are compared there might well also be problems due to size related differences in spatial data accuracy. Statisticians refer to this as the modifiable areal unit problem and it is endemic whenever spatial data for zones (or regions or areas of any kind) are described, analysed, and modelled in anyway. There are two components to it: a scale problem (i.e. how many zones is best) and an aggregation component (i.e. if we want 20 zones which is the best aggregation). Typically, N zones can be aggregated into M bigger zones in zillions of ways; for example, if N=1000 and M=10, then there are about 10^{1240} different possible zoning systems. GIS gives the user the ability to change the nature of the zones of interest; in particular, they can design new zoning systems that have known properties. It also allows the suppliers of sensitive personal data to create sets of spatial reporting areas that are simultaneously safe from a data confidentiality perspective and meaningful from a data content or data quality point of view, see Openshaw and Rao (1995).

Viewed from a pre-GIS historical perspective, the zones used to report statistical data are usually large. They had to be, mainly for operational data collection reasons. Typically, these zones would reflect administrative areas. For example a data set such as the Census is usually collected at a much smaller scale; indeed the census enumeration district concept in the UK is over 100 years old. Until 1966, the smallest census area for which data could be obtained in the UK was the ward/parish; there were about 15,000 wards/parishes compared with 150,000 eds in 1991. Currently, the smallest census reporting area is the enumeration district (a region of space that typically contains about 400 people but with a wide size variation). The census ed was not designed as a census reporting area but as a convenient spatial unit for organising the task of collecting census forms. In the GIS era, there is no need anymore for the data collection level to be the same as the data reporting level since the individual data can be aggregated to whatever zoning system is considered appropriate or best. Census eds may well be good areas for equalising census enumerator workloads but they are not necessarily good units for data analysis (too much variation in social heterogeneity and size, they ignore social divides, and provide an insensitive partitioning of space).

Three types of zones can be recognised:

1. zones useful for organising data collection in an efficient manner;
2. zones for reporting data in a safe form; and
3. zones engineered for analysis and policy purposes.

Historically zone types 1 and 2 are identical but the type 1 zones should really be as small as possible and the type 3 zones carefully designed to minimise data damage on the resulting applications.
It is clearly desirable that in view of changing European Data Protection legislation that tools are able to design zones which can report spatial data in a confidentiality safe form but without doing unacceptable damage to the data (Openshaw 1994h). It is also desirable that the geographical objects

that the zones represent in one country are the same or comparable to what they represent in another otherwise there may be unfair biases in favour or against some countries and thus no safe basis for spatial planning and cross country analysis. Research is needed urgently to provide users with the design tools that will allow them to effectively and efficiently handle the freedom, that GIS provides for flexible aggregations of spatial data. Until recently, zones were fixed and rigid geographical objects. GIS has changed this situation and there is need for useful tools to handle the resulting uncertainty in zonal choice and thus in analysis results. It is also important to use zoning systems that consist of comparable objects and of developing new methods for comparing area objects when they are not.

The following research topics are suggested:

1. investigate the extent of zoning system effects in existing data and policy analyses; how robust has the spatial analysis been, could it be improved by changing or re-engineering zoning systems;
2. develop reconstructed standard zoning systems for the EU which attempt to even out or remove the existing biases and anomalies in current ad hoc definitions;
3. to provide users with a set of tools for flexible zone design that will allow them to replace administrative areas, which often do not correspond to socioeconomic reality, by new sets of areas which do or at least are less disastrous and with a minimal level of ecological fallacy error;
4. investigate and determine which characteristics of zoning systems would best suit particular applications; and
5. investigate how best to define zones that preserve confidentiality of personal and other sensitive data in an even manner across various countries.

Theme 7: Impediments to the development and use of spatial analysis methods

A further area of research concerns improving awareness, training, and dissemination of useful spatial analysis methods.

There are three principal areas of need:
1. training and exchange of information about new methods at the practitioner level;
2. investigation of organisational factors affecting the uptake of spatial analysis methods, with a view towards standardisation of the methods used at both EU and national levels; and
3. legalistic aspects of spatial analysis, for example, what litigational sensitive factors are there, how does the situation in the EU compare with the USA, see Openshaw (1993b).

CONCLUSIONS

GIS as a tool for handling spatial statistical information is here to stay. Vast resources have been invested and continue to be spent on the creation of spatial databases of one kind or another. There is now an increasingly urgent need to develop relevant applied spatial analysis methods that can make the most of the investment in spatial data collection. The nature of the EU's geo-statistical databases put considerable emphasis on certain spatial analysis problems; particularly those relating to zone design and spatial representation issues. The practical need to apply EU spatial planning and spatial policy instruments across the EU over a wide range of different spatial data sets emphasises the importance of doing so safely and optimally. However these concerns and needs are most clearly seen in a European rather than a North American context. Europe also has research groups with the necessary skills. What it lacks at present is a coordinated and well funded basic and applied research programme for solving the problems that exist in Europe but which are also relevant to the rest of the world. Hopefully, this paper will help stimulate and inform this debate.

Acknowledgement: *The first author acknowledges a grant from the Austrian Fonds zur Förderung Wissenschaftlicher Forschung (P09972-TEC).*

References

Anselin, L. and Getis, A. (1992): Spatial statistical analysis and geographic information systems, in M M Fischer and P Nijkamp (eds.): **Geographic Information Systems, Spatial Modelling, and Policy Evaluation**, Berlin et al.: Springer-Verlag, pp. 35-49.

Dorling, D., Openshaw, S., (1992) 'Using computer animation to visualise space-time patterns', **Environment and Planning B** 639-650

Fischer, M.M. (1993) From conventional to knowledge based geographical information systems, **Proceedings, 16th Urban Data Management Symposium,** Vienna: ADV Handelsgesellschaft m.b.H. pp 17-25.

Fischer, M. M. (1994): Expert systems and artificial neural networks for spatial analysis and modelling: Essential components for knowledge-based geographical information systems, **Geographical Systems 1** (in press).

Fischer, M. M. and Gopal, S. (1994): Artificial neural networks. A new approach to modelling interregional telecommunication flows, **Journal of Regional Science 34** (in press).

Fischer, M.M. and Nijkamp, P. (eds.) (1992a): **Geographic Information Systems, Spatial Modelling, and Policy Evaluation.** Berlin Springer-Verlag

Fischer, M. M. and Nijkamp, P. (1992b): Geographic information systems and spatial analysis, **Annals of Regional Science 26**, pp. 3-17.

Fischer, M.M. and Scholten, H.J. (1993): Geographic information systems, spatial data analysis and spatial modelling. Problems and possibilities, **Position Paper prepared for the GISDATA Specialist Meeting on GIS and Spatial Analysis**, Amsterdam, December 1-5, 1993

Flowerdew, R. and Green, M. (1991) Data integration: Statistical methods for transferring data between zonal systems, in I Masser, and M Blakemore, (eds): **Handling Geographical Information: Methodology and Potential Applications**, London, Longman pp.38-54

Fotheringham, A.S. and Rogerson, P.A. (1993): GIS and spatial analytical problems, International **Journal of Geographical Information Systems 7**, pp. 3-19.

Goodchild, M., Haining, R., Wise, S., et al. (1992): Integrating GIS and spatial data analysis: Problems and possibilities, **International Journal of Geographical Information Systems 6**, pp. 407-423.

Hecht-Nielsen, R. (1990): **Neurocomputing**. Reading (MA): Addison-Wesley.

Hewitson, B.C., Crane, R.G., (1994) **Neural Nets: Applications in Geography** Kluwer Academic, Boston

Kanal, L.N. (1993): On pattern, categories, and alternate realities, **Pattern Recognition Letters 14**, pp. 241-255.

Openshaw, S. (1990) Spatial analysis and geographical information systems: A review of progress and possibilities, in H J Scholten, and J C H Stillwell, (eds) **Geographical Information Systems for Urban and Regional Planning,** Kluwer, London pp. 153-163.

Openshaw, S. (1991a) Developing appropriate spatial analysis methods for GIS, in D J Maguire, M F Goodchild, and D.W. Rhind, (eds.) **Geographical Information Systems, Volume 1: Principles,** Essex: Longman. pp.389-402

Openshaw, S. (1991b) A spatial analysis research agenda in I. Masser, and M. Blakemore, (eds): **Handling Geographical Information: Methodology and Potential Applications**, Essex: Longman. pp.18-37

Openshaw, S. (1992): Some suggestions concerning the development of artificial intelligence tools for spatial modelling and analysis in GIS, in M M Fischer and P Nijkamp (eds.): **Geographic Information Systems, Spatial Modelling, and Policy Evaluation,** Berlin et al.: Springer-Verlag. pp.17-33

Openshaw, S., (1992a), ' A review of the opportunities and problems in applying neurocomputing methods to marketing applications', **J. of Targeting, Measurement and Analysis for Marketing** 1, 170-186

Openshaw, S. (1993): What is GISable Spatial Analysis? . . .

Openshaw, S., (1993a), 'Modelling spatial interaction using a neural net', in M M Fischer and P Nijkampt (eds) **GIS, Spatial Modelling and Policy** Springer-Verlag, Berlin, 147-164

Openshaw, S., (1994a), 'GIS crime and spatial analysis', Proceedings of GIS and Public Policy Conference, Ulster Business School, Ulster

Openshaw, S., (1994b), 'Two exploratory space-time attribute pattern analysers relevant to GIS, in S fotheringham and P Rogerson (eds) **GIS and Spatial Analysis** Taylor and Francis, London 83-104

Openshaw, S., (1994c), 'Developing automated and smart spatial pattern exploration tools for geographical information systems applications', **The Statistician** 43, (forthcoming)

Openshaw, S., Waugh, D., Cross H, (1994d), 'Some ideas about the use of map animation as a spatial and analysis tool, in H M Hearnshaw and D Unwin (eds) **Visualisation in GIS** Wiley, London 131-138

Openshaw, (1994e), 'Computational human geography: exploring the geocyberspace', **Leeds Review** 37 (forthcoming)

Openshaw, S., (1994f), 'Developing smart and intelligent target marketing systems, part 1', **Journal of Targeting Measurement and Analysis for Marketing** 2, 289-301

Openshaw, S., (1994g), 'Developing smart and intelligent marketing systems, part 2' **Journal of Targeting, Measurement, and Analysis for Marketing** 3, 31-38

Openshaw, S., (1994h), 'Social costs and benefits of the census', **Proceedings of XVth International Conference of the Data Protection and Privacy Commissioners** p89-87, Data Protection Registrar, Manchester