November, 2000

GIS and Spatial Analysis in Urban and Regional Research

Derek Bond

Available at: https://works.bepress.com/derek_bond/11/
GIS and Spatial Analysis in Urban and Regional Research

Derek Bond

Abstract

The early optimism surrounding GIS has now faded. Few GIS implementations have lived up to expectation. The most successful have been those at the operational level where their main role has been to replace paper maps with digital ones. However, as the cost of GIS implementation falls and spatial data becomes more readily available there is a need to consider the future of GIS and spatial analysis. This is particularly important in the field of urban and regional research especially where considerable investment in GIS has already been made. This paper looks at the major issues that face regional and urban research in the next stage of development of GIS and spatial analysis. The main conclusion of the paper is that regional and urban statistics offices and researchers are well placed to benefit from future developments in GIS and spatial analysis.

Derek Bond

School of Business, Retail and Financial Services
University of Ulster at Coleraine
Cromore Rd.
Coleraine, Co. Derry
United Kingdom
BT52 1SA
e-mail: d.bond@ulst.ac.uk

C&R – ISSN 1568-167X
Realising the Dream

Modern statistical techniques and theories have, generally, been developed to assist in the reduction of large amounts of primary and secondary data into seemingly useful and manageable information. Until recently this has meant that statistics have been primarily concerned with data reduction. That is, trying to produce summary figures and tests which describe something about the underlying data generation processes.

With the increasing performance and the falling costs of information and communication technology (ICT) this reliance on mainly parametric based data reduction techniques has been challenged. It is becoming increasing cost-effective for end-users to handle the large-scale datasets themselves. One of the main areas where this trend is most pronounced is in the handling and analysis of spatially referenced data using Geographic Information Systems (GIS). However, many of the original dreams and expectations of what GIS could offer do not seem to have been realised. In particular their role as an essential tool for policy and decision-makers (see for example Department of Environment (1987)) do not seem to have been met. Reviews of the current topics in the GIS literature together with the research findings of studies on the implementation of systems (for example Hendriks (1998) and Campbell and Masser (1995)) show that on the whole GIS is being used mainly in support of operational level tasks. In many cases they do little more than replacing paper maps and records with digital equivalence. The use of GIS and spatial analysis at the strategic and policy related levels are currently limited. Various perspectives on this have been put forward covering aspects such as availability of tools to management issues (for example Worrall and Bond (1996) argue that the main issues are organisational, Hendriks (2000) similarly takes a 'learning organisation' perspective.)

Spatial analysis and GIS requires large investment in staff and time to provide adequate data in formats that can readily be used for detailed spatial data analysis. Many regional and urban statistics offices and research institutes have committed considerable resources to this task. It is therefore important that the returns on such investments are realised. In this paper and the other papers in this volume of Cities and Regions various issues associated with spatial analysis and GIS in urban and regional research will be critically examined. The aim is to show that while there might be a concern with the general direction that GIS and spatial analysis research are moving those concerned with urban and regional research are attempting to address issues related to ensuring that the investment in spatial data and related software is realised. This is done by considering research in the various aspects of spatial analysis and GIS. These range from the availability of spatial data, the software needed to handle the data to methods for handling the data.

SPATIAL DATA

Fundamental to the use of GIS and spatial analysis is the availability of spatial data. This data is normally of two types: topographic and attribute. In the early days of GIS there was a lack of readily available topographic data and it was a large resource commitment to produce it. Now digital topographic data is readily available both in raster and vector formats, though issues of copyright and access still exist. While topographic data is now readily available few national spatial infrastructures exist
(c.f. Masser (1999) for discussion of the issues surrounding the development of national spatial infrastructures). The availability of topographic data has reduced the overheads associated with developing a GIS and has led to an increase in demand for more spatially dis-aggregated and referenced attribute data to use with such systems. This demand and a lack of national and international spatial data infrastructures have various implications for those involved in regional and urban research and analysis. The most common problems found in the literature are those of attaching consistent spatial references to attribute data and of the requirement for fine detail small area statistics which do not compromise on confidentiality. Caught in the middle of these problems are Urban and Regional Statistics Offices. Most of these offices are both data users and providers. While involved in some primary data collection, often as a by-product of the administrative processes of their administrations, they often add value to data collected by National Statistical Institutes (NSIs) and other bodies. Often these secondary data sets do not contain spatial references that are consistent with the topographic data or other attribute data used. The issue of attaching spatial identifiers is a complex one (for example, see Visvalingam (1991) for discussion) and is linked to that of statistical confidentiality.

Statistical confidentiality is often quoted as a major issue hindering the development of GIS. The problem is that in the ideal world a GIS would contain the basic records of any attribute dataset with consistent spatial references attached. In most countries this is impossible as statistical disclosure rules do not permit the identification of individuals or single organisations (see for example Marsh et al. (1992) for discussion on confidentiality surrounding the 1991 United Kingdom's Census of Population and EuroStat (1992) for general discussion of various issues surrounding statistical confidentiality). The issues of statistical confidentiality and spatial tagging are addressed in this volume by Wendy Treadwell.

While spatial tagging and statistical confidentiality are problems that need to be addressed for the successful implementation of GIS in an organisation GIS also provides tools with which to consider these issues. For example, GIS software is often used to assist in the construction of statistically 'sound' spatial reporting regions. However, as Coombes (1998) concluded, while discussing the construction of 'Travel To Work Areas' for England and Wales: '... the methods and software which is needed for defining new sets of areas ... are still too specialised to be found in many GIS packages'.

The problems facing urban and regional statistics offices are two faced. They are often trying to combine data from various data sources to provide high quality insights into the functioning of their areas whilst being seen as the first port of call by researchers and other end users looking for detailed secondary data on the area. Much innovative research has been done by these offices to integrate spatial data sets and provide consistent data and information to their end users (see, for example Tammilehto-Luode and Backer (1999) for a description of a Scandinavian exercise to provide small area statistics using GIS).
SHARING SPATIAL DATA

In an increasingly globalised world most socio-economic activities do not stop at international borders so there is a need to share data across international boundaries. However, the problems of sharing data are many and range from the issues of compatibility of formats and comparability of coverage to issues of commercial exploitation. At the urban and regional level many of these issues are more acute than at the national level. For example when comparing small area data minor definitional differences can result in major problems.

In Worrall and Bond (1996) it was argued that the commercial or quasi-commercial nature of much of the attribute data required for the successful development of GIS is a major handicap in the success of major GIS implementations. This can be a particular problem in small urban and regional statistics offices that are mainly concerned with adding value to secondary data. The secondary data used is often at a resolution and detail that is of use in the commercial sector for such things as marketing, utilities management etc. and is main source of non-public revenue for the public body collecting the data. With the increasing requirement in many countries that their public sectors move to cost recovery models this has a serious implication for urban and regional statistics offices.

Despite such problems many regional and urban statistics offices have managed not only to obtain various spatially referenced data sets but also to share the data with other offices on an international scale allowing for comparative statistics to be produced. (see for example City of Helsinki (1999) and City of Nuremberg (1995) for examples of how city offices have tried to produce comparable statistics).

ACCESSING SPATIAL DATA

Advances in information and communication technologies have led both to an increased demand for spatially reference data and also to its availability, subject to the commercial and confidentiality problems discussed above. Whilst being part of the catalyst for this increased demand for data GIS has become to be seen as part of the solution to the problem of developing effective access to large amount of spatially referenced data. These very large spatial datasets are becoming available through such outlets as digital libraries (Smith (1996)), data archives (Lievesley (1992)) and data warehouses (Jarke et al. (2000)). Users are faced with the problem of finding and correctly using relevant data from such sources. This can be an intimidating task.

In recent years there has been a growth in interest in the use of metadata to help search such databases. The problem is that there are no agreed standards for metadata though in the social science community attempts are being made to standardise on the 'Dublin Core' (Lee-Smiltzer (2000)). Various projects, such as NESSTAR/FASTER (www.faster-data.org) in the European academic community, are attempting to provide user-friendly interfaces to attribute data using metadata-based interfaces. The problem facing researchers and other users of the data sets is lack of detailed knowledge of the information that can be derived from the data contained in these complex databases.
The visual element of GIS has always played a large part in its popular appeal and research has shown (Smelcer & Carmel (1997)) that, provided the task is relatively simple, a map can provide an ideal starting point for information discovery. This has led to growth in GIS based interactive information seeking software. The aim of this software is to replace traditional text-based information retrieval strategies with ones based on visualisation and knowledge discovery in databases (Fabrikant (2000)). Knowledge Discovery in Databases (KDD) has been defined as ‘the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data’ (Fayyad et al. (1996) p.6). KDD is closely linked with data mining and uses the computational power of modern IT to explore data sets rather than relying on data reduction techniques. A simple example of the use of visualisation and KDD is the Casweb Interface to the United Kingdom's 1991 Census of Population (http://jimay.mcc.ac.uk/casweb/). In this volume the paper by Hans Voss et al. looks at, amongst other things, the issues surrounding the implementation of map interfaces to complex data and the role of KDD.

SPATIAL DATA ANALYSIS

It is often argued (c.f. Csillag and Kabos (1999)) that developments in methods of spatial analysis have failed to keep up with those in GIS thus hindering its uptake. While there may be some truth in this most of the discussion is couched in the traditional Fisher-Neyman paradigm and emphasis is put on the need for parametric estimation and model building rather than on KDD and Geographic Visualisations (GVIs) (c.f. Maceachren et al. (1999)). The development in the use of these techniques has come from two distinct fields. GVIs has been developed by mathematical geographers whereas KDD has grown from work on non-spatial data mining in mathematics.

Traditional spatial analysis has also developed and adopted a more Exploratory Data Analysis (EDA) approach. The paper in this volume by Peter Brown et al. is a good example, using geodemographic profiling, of the current policy interface between GIS and spatial analysis. The problem with approaches such as this is the reliance on the results of one data reduction exercise. GVIs and KDD allow the user to more flexibility explore the larger complex datasets. The paper by Hans Voss et al. in this volume looks at the issues of visualisation and KDD in spatial datasets and the paper by Maribel Santos and Luis Amarel gives an example of how spatial reasoning might be applied in KDD.

GIS AND SPATIAL ANALYSIS SOFTWARE

To many end users the talk of GVIs and KDD in spatial datasets seem a distant dream as traditional GIS software, other than simple raster based mapping packages have had the reputation of being extremely expensive (Pinals (1998)). However, as the papers by Patrick Gerland et al. and Dang Van Due in this volume show there is now a wealth of low cost and free software available for end-users. Similarly the paper by Maribel Santos and Luis Amarel shows how KDD might be done on spatial datasets using standard software and the paper by Natali Andrienko et al describes ready available visualisation and KDD software.

In the first of these papers Patrick Gerland et al. describe the GIS software developed by the United Nations for population research which includes software for capturing, storing, retrieving and analysing spatial data - all of which is readily available on the web (http://www.unido.org/popin/softproj/index.htm) and are windows based for ease of use. In the second of these papers Dang Van Duc outlines the web based MapOnline software which has developed out of the UN's PopMap software. The paper by Maribel Santos and Luis Amarel gives an example of how the SPSS add in 'Clementine' might be used for spatial KDD. Clementine is a ready available datelining programme and again has an easy to use window interface. The last of these papers by Natali Andrienko et al. discusses the use of the Java based Descartes software (http://allanon.gmd.de/and/java/iris/). This software like MapOnline is a web based spatial analysis tool allowing for KDD.
These four papers show that:

• to use GIS and spatial analysis effectively there is no longer the need to invest heavily in expensive cartographic based GIS software;
• large amounts of data do not have to be stored locally; and
• easy to use interfaces for even complex operations such as GV'i's and KDD are now becoming readily available

**SUMMARY**

This paper began by considering the generally rather pessimistic view of the cost effectiveness of investment in GIS and spatial analysis. Through examples the paper has shown that the 'state of health' of GIS and spatial analysis in urban and regional research is good. Indeed from considering the information contained in the papers in this volume one should conclude that the, often considerable, investment by city and regional statistical offices in compiling and maintaining large-scale spatially referenced datasets, places them and associated researchers in an extremely strong position to benefit from the current developments in the area.
References


