

CHAPTER 38

THE USE OF GIS IN THE ANALYSIS OF DIVERSE URBAN DATABASES

(D Martin, P Longley and G Higgs)

Computers, Environment and Urban Systems, 1994, 18(1), 55-66



Pergamon

Comput., Environ. and Urban Systems, Vol. 18, No. 1, pp. 55-66, 1994
Copyright © 1994 Elsevier Science Ltd
Printed in the USA. All rights reserved
0198-9715/94 \$6.00 + .00

0198-9715(93)E0011-4

THE USE OF GIS IN THE ANALYSIS OF DIVERSE URBAN DATABASES

David Martin

Department of Geography, University of Southampton

Paul Longley

Department of Geography, University of Bristol

Gary Higgs

*Department of City and Regional Planning, Cardiff, UK, and Department of
Geography, University of Bristol*

ABSTRACT. *Few GIS studies have demonstrated a preparedness to grapple with the messy empirical problems which beset applied research. This paper uses a detailed case study to illustrate some of the problems which are encountered when attempting to integrate widely used urban data sources into a disaggregate GIS. We also anticipate the prospects for information management in the light of apparent changes in the availability of property registers in the UK and the innovation of the UK Ordnance Survey's "Address-Point" product.*

INTRODUCTION: A GEOGRAPHICAL PROBLEM

Since 1990, the UK has undergone a number of major changes in the structure and level of local taxation, moving from a property-based domestic rating system, through different variants of a personal community charge regime (i.e., with and without transitional relief payments and national reductions), to a new council tax, implemented in April 1993. The geographical implications of these changes have been significantly affected by factors such as household size and composition, and property value and condition, which vary markedly at the neighbourhood and street levels (Paddison, 1989), and have been of particular interest to the authors. Our

Requests for reprints should be sent to Dr. David Martin, Department of Geography, University of Southampton, Southampton SO9 5NH, UK.

work is described in previous publications (e.g., Longley, Martin, & Higgs, 1993; Martin, Longley, & Higgs, 1992), and has been made possible by the development of a "loosely coupled" GIS environment incorporating Arc/Info (ESRI, Redlands, CA), a spreadsheet, and custom-written programs (Higgs, Longley, & Martin, 1991). In this paper, it is not our intention to discuss the substantive theme of changes to the system of local taxation, but rather to reflect upon some of the lessons learned regarding the analysis of large municipal databases within GIS, and to assess their implications for local authority database management.

The task of understanding the implications of taxation changes at the local level is a complex geographical problem, and is further complicated by the diversity and incompatibility of the databases concerned, since these often contain no explicit locational referencing. The implementation and administration of successive taxation systems has rested with local authorities, who have a number of extensive but disconnected computer databases relating to persons, properties, and small areas. One of the most frequently cited characteristics of GIS has been their potential as systems for the integration of diverse geographic information (Department of the Environment [DoE], 1987). In practice, however, population-related characteristics are invariably referenced indirectly: via property codes, census zones, administrative divisions or postal geography, and the task of defining a common spatial referencing framework for such data remains one of the main tasks for the full implementation of GIS in the socioeconomic realm (Martin, 1991). The further advance of GIS is dependent upon the application of the technology in applied policy contexts, and yet implementation at the level of local government is confronted with severe data problems of this type (Gault & Peutherer, 1990; Worrall, 1990). Dwellings, households, and individuals are the basic units for many forms of policy, and are important as the units of supply for services. It is suggested here that our Cardiff work with its specific concern for local taxation serves as a useful illustration of the issues which must be tackled when seeking to conduct geographical analysis of households and individuals using pre-existing databases. The socioeconomic, multipurpose, policy-related information dealt with here presents a different set of problems to those conventionally associated with applications in utility and land use management, or in the handling of simple census-type information. We anticipate that our ability to handle such information will be greatly enhanced by the arrival of national address georeferencing in the form of Address-Point from Autumn 1993 onwards.

In the following section, we will outline the characteristics of the datasets involved in our own research, and the principles which guide their incorporation within GIS databases; the third section will review the obstacles to spatial analysis which we have encountered; and the fourth section will consider the future development of studies of this kind.

COMPLEX URBAN DATABASES

Our work has dealt with the "Inner Area" of Cardiff, identified for policy purposes by the City Council, and approximating to the city's 19th/early 20th century urban core. The area contains approximately 47,000 properties and 89,000 electors, and comprises 12 whole or part "communities" (a Welsh administrative division, formally created during the 1970s and broadly equivalent to English "parishes"). The Inner Area and its named community subdivisions are shown in Figure 1. We see integration of a number of geographic databases as a prerequisite to detailed area-wide research and management, and in practice, such databases might be held centrally or at departmental level within local authorities (Campbell, 1993). Regarding our own work on local taxation, the relevant databases have been collected for a variety of purposes and relate to many different urban functions, and most already exist in some form within the City Council. Although geographic in nature, not all the existing databases include any form of explicit spatial referencing.

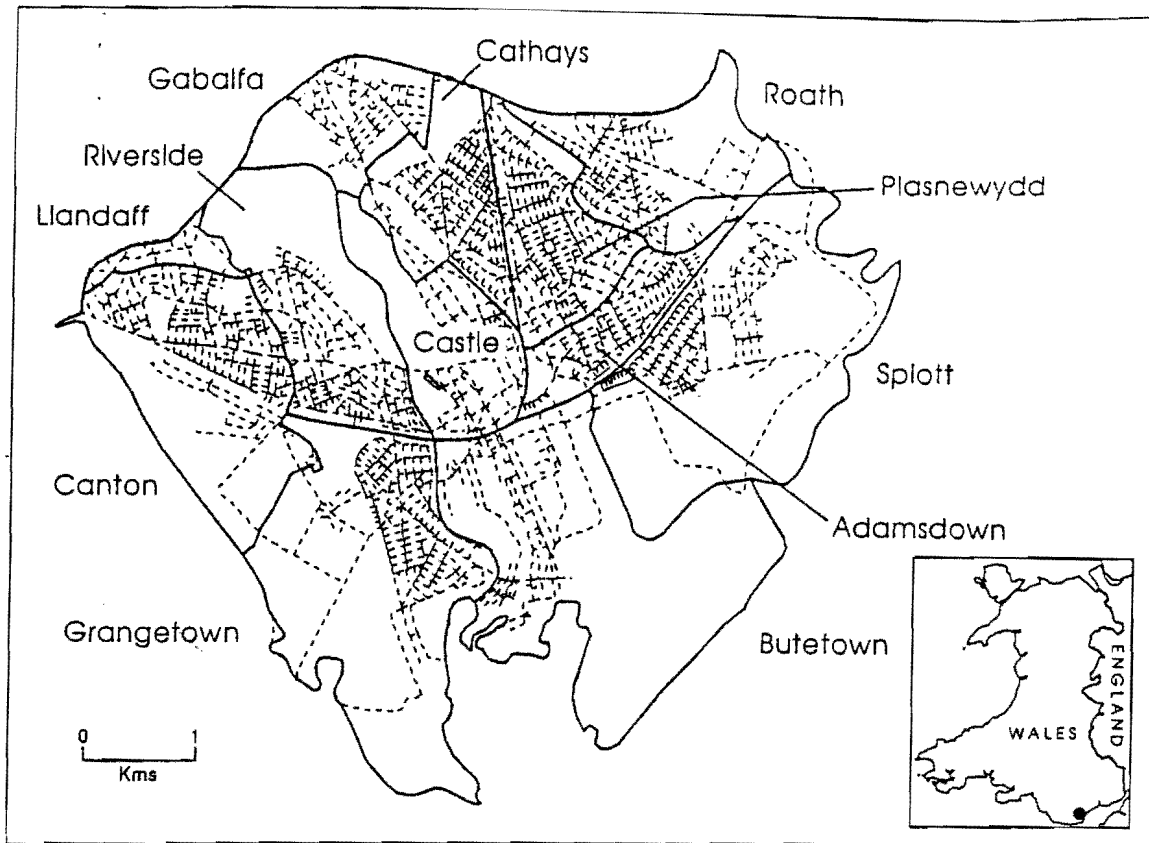


FIGURE 1. Digitized Street Network: Cardiff Inner Area.

Urban GIS research has usually been restricted to the use of census and survey information, but organizations operating in the "real" world are generally dealing with complex geographical registers providing much more disaggregate information, yet without the substantive range of the census. Table 1 provides an overview of the databases which are available in the context of the Cardiff study. These registers, although referring to individual persons or properties, do not usually contain explicit spatial references, and very rarely include information such as the postcode, about which so much has been written elsewhere (DoE, 1987; Raper, Rhind, & Shepherd, 1992). The Cardiff City Council address registers contain unique property reference numbers (UPRNs), which typically comprise codes for community, street, and (sometimes) subdivisions of streets such as large blocks of flats, together with the address number. Additional complications occur where some addresses are subdivided, and central Cardiff contains a large number of houses in multiple occupation (HMOs). For example, the 17-digit UPRN appearing in the council tax register is derived as follows:

- 1-2 billing authority (e.g., 32 for Cardiff);
- 3-5 community (e.g., 229 for Grangetown);
- 6-8 filler;
- 9-12 Street code (or subdivision);
- 13 filler;
- 14-17 property reference.

From a UPRN of this type, it is possible to obtain 7-digit codes (community and street segment), which form the basis of a link to a digitized street network, and a possible method for providing linkage between different registers. In our previous work, all detailed information has been aggregated to the level of the street segment for analysis. We shall now give some

TABLE 1. Primary Databases Available for the Cardiff Study

Database	Basic unit	Georeference	Availability
Domestic rates register	individual properties	Property Ref. No./addresses	available via City Council
Community charge register	individual persons	addresses	confidential, in practice
Electoral roll	individual persons	addresses	available via City Council
Streets	centrelines	segment coordinates	OSCAR/1:1,250 or 1:2,500 maps
Uniform areas (e.g., HCS areas)	small areas	boundary coordinates	own survey/digitizing
House price data	selected properties	street or property	own survey of asking prices
Census data	enumeration district	boundary coordinates/centroid	OPCS via MCC
ED/postcode directory	partial postcode unit	point coordinates	100 m precision from CPD
Council tax register	individual addresses	Property Ref. No./postcoded address	multiple records not generally available
House Condition Survey	sampled individual addresses	Property Ref. No./addresses survey	Confidential
Address-Point/ Pinpoint Address Code	individual address	point coordinates/ address	1 m precision commercially available

brief consideration to each of the primary registers involved in the Cardiff project, many of which are of more general import in the building of urban information systems.

The original (pre-1990) domestic property rating system was based on a domestic rates register, which contains a record for every rateable property, together with details of property type, rateable value, and a unique property reference number. Although the domestic rates system was replaced in Great Britain by the community charge in 1990 (1989 in Scotland), local authorities retain a statutory duty to maintain the domestic rates register for a variety of roles until the year 2000. The rating reference number includes codes which relate to the street and community in which a property falls, but no direct linkage exists to a grid reference or postcode. The rates register does not contain information on any new properties built since the instigation of the community charge in April 1990, and thus it is a data source characterised by increasing obsolescence.¹ For any known rate poundage, it is possible to calculate the rates which would be payable on each domestic property using this register. In the case of the study

¹The Register does continue to be updated, however, with revisions to pre-1990 rateable values which result from appeals heard after that date.

area shown in Figure 1, the Cardiff City domestic rates register suggests that 47,014 hereditaments were located in these streets (some of which were split into separately rated units), of which 45,658 were entirely residential dwellings. This information is in the public domain. An extract of the register is shown in Figure 2.

The electoral register remains the most widely used individual and address frame for social survey research (Hoineville, Jowell, and Associates, 1978), and contains the names and addresses of all registered electors. Its coverage has always been restricted to the adult eligible voting population, and it remains unambiguously within the public domain. However, there is mounting evidence that its coverage has declined significantly and persistently, particularly since the introduction of the community charge. The community charge as implemented in 1990 (in England and Wales) did in fact require the construction of a completely separate register of chargepayers, and the electoral register was generally used only as an address frame in order to aid the identification of liable households. This should have included every eligible adult, although in practice, significant and spatially variable under-registration is believed to have occurred. The community charge register is an individual-level database, but remains confidential at the disaggregate level, and thus cannot be used either in an integrated database, or as a means of modelling and linking other datasets. In our own research, we have found it necessary to use the register of electors as a surrogate public domain approximation used to identify the geography of community charge liability. There are some differences between the criteria for inclusion on the community charge and electoral registers, as well as different (and possibly systematic) biases in the inclusion rates of individuals. The electoral register again contains no direct linkage to grid references or postal geography, but street addresses are provided. In addition, the electoral register information is available as street-by-street counts of electors. Use of the electoral register as a surrogate for the community charge register has been discussed in Martin et al. (1992).

The council tax register shows the banding of all properties in the valuation list for the billing authority area. The register contains a unique 17-figure reference for each property, as well as the address of the property (including its postcode in a number of cases), and the band in which that property falls. A high proportion of the Cardiff council tax register is postcoded. An extract from the register is given as Figure 3. One of the much-vaunted criticisms of the council tax valuation process is that some blocks of properties may have been valued by rapid inspection of the front elevation. If this were the case (and we understand that in Cardiff at least, it was not), and information were used as an input to spatial analysis, this may generate ecological fallacy and modifiable areal unit problems, if it amounts to an assumption of homogeneity of property within street segments (e.g., extensions and improvements may have been

10101200024102	67 HOUSE	24 BUTE STREET	CARDIFF	
10102000021208	161 MAISONETTE	21 CAROLINE STREET	CARDIFF	
10102000039108	161 MAISONETTE	FIRST & SECOND FLOORS	39/40 CAROLINE STREET	CARDIFF
10103000040103	480 MAISONETTE	ST DAVIDS CLERGY HOUSE	REAR 40 CHARLES STREET	CARDIFF
10103000043046	157 MAISONETTE(GF & 1F)	43 CHARLES STREET	CARDIFF	
10103000048054	107 FLAT	THIRD FLOOR	48 CHARLES STREET	CARDIFF
10103000070005	2 CATHEDRAL	ST DAVIDS CATHEDRAL	CHARLES STREET	CARDIFF
10103600024039	223 FLAT	SECOND FLOOR	24 CHURCHILL WAY	CARDIFF
10106450001319	347 FLAT	CARETAKERS FLAT	FOURTH FLOOR PARK HOUSE	GREYFRIARS ROAD
10106800015096	67 FLAT	NINTH FLOOR	ARLBEE HOUSE	GREYFRIARS ROAD
10107200002021	140 CARETAKERS FLAT	2ND FLOOR	2 PEARL HOUSE	GREYFRIARS ROAD
10107400001002	82 HOUSE	PUBLIC BATHS	GUILDFORD CRESCENT	CARDIFF
10109000003056	67 FLAT(THIRD FLOOR)	3/4 HIGH STREET	CARDIFF	
1010900002640A	57 FLAT(FOURTH FLOOR)	26 HIGH STREET	CARDIFF	
10109000028072	60 FLAT(FIFTH FLOOR)	28 HIGH STREET	CARDIFF	
10109000028053	82 FLAT(THIRD FLOOR)	28 HIGH STREET	CARDIFF	
10109000028068	82 FLAT(FOURTH FLOOR)	28 HIGH STREET	CARDIFF	
10109400010002	182 FLAT(FIFTH FLOOR)	OXFORD HOUSE	HILLS STREET	CARDIFF
10112000009307	94 MAISONETTE	1ST 2ND & 3RD FLOORS	9 MILL LANE	CARDIFF
10112600005056	74 FLAT(THIRD FLOOR)	MORGAN ARCADE CHAMBERS	CARDIFF	

FIGURE 2. An Extract from the Pre-1990 Rates Register.

32101000005000010	E	1 BLACKWEIR FARM COTTAGES, BLACKWEIR PARK, NORTH ROAD, CARDIFF, CF1
32101000005000020	D	2 BLACKWEIR FARM COTTAGES, BLACKWEIR PARK, NORTH ROAD, CARDIFF, CF1
32101000005000030	E	3 BLACKWEIR FARM COTTAGES, BLACKWEIR PARK, NORTH ROAD, CARDIFF, CF1
32101000005000040	E	4 BLACKWEIR FARM COTTAGES, BLACKWEIR PARK, NORTH ROAD, CARDIFF, CF1
32101000005000050	G	BLACKWEIR FARM HOUSE, BLACKWEIR PARK, NORTH ROAD, CARDIFF, CF1 3
32101000014000302	D(comp)	CENTRAL BOYS CLUB, BUTE TERRACE, CARDIFF, CF1 2FE
32101000014000500	A	NEW COLLEGE, BUTE TERRACE, CARDIFF, CF1 2FE
32101000020000210	B	FLAT 1ST FLR AT 21, CAROLINE STREET, CARDIFF, CF1 1FG
32101000020000220	B(comp)	22/23, CAROLINE STREET, CARDIFF, CF1 1FG
32101000020000250	C(comp)	KINGS CROSS HOTEL 25, CAROLINE STREET, CARDIFF, CF1 1FF
32101000020000370	B(comp)	37, CAROLINE STREET, CARDIFF, CF1 1FF
32101000020000381	B	FLAT ABOVE SHOP AT 38, CAROLINE STREET, CARDIFF, CF1 1FF
32101000020000391	B	1ST & 2ND FLR AT 39/40, CAROLINE STREET, CARDIFF, CF1 1FF
32101000024000150	C(comp)	GLOBE HOTEL 15, CASTLE STREET, CARDIFF, CF1 2BS
32101000024000300	A(comp)	ANGEL HOTEL, CASTLE STREET, CARDIFF, CF1 2QZ
32101000024000312	B(comp)	CARDIFF CASTLE, CASTLE STREET, CARDIFF, CF1 2BT
32101000030000400	G	ST DAVIDS CLERGY HOUSE REAR OF 40, CHARLES STREET, CARDIFF, CF1 4EE
32101000030000430	A	43, CHARLES STREET, CARDIFF, CF1
32101000034000122	A(comp)	12/13 2ND FLR REAR, CHURCH STREET, CARDIFF, CF1 2BG
32101000034000123	A(comp)	12/13 2ND FLR FRONT, CHURCH STREET, CARDIFF, CF1 2BG

FIGURE 3. An Extract from the Council Tax Register Introduced in April 1993.

carried out either internally and/or to the rear elevation of properties). Direct address matching from registers of this kind is highly problematic, because of the different formats used for the entry of textual address information in the files, and inconsistencies in address definition, punctuation and spelling.

Additional information may also be available for a range of small, usually ad hoc areal units. In the Cardiff context, these include the "House Condition Survey (HCS) areas" used for a major survey of the Inner Area dwelling stock in 1989, and census wards and enumeration districts. Digital boundaries are available for each of these zoning systems, but the labelling schemes used are not compatible between the census and the HCS. Data of this kind include important details about dwelling characteristics, and general socioeconomic conditions.

The datasets outlined here are typical of the important sources of information which will be found in any local authority context, although they represent a more detailed level of information than has typically been discussed in this context before (e.g., Openshaw, 1990). Many applications may be cited, besides the evaluation of taxation regimes considered here, which require integration between such data. A GIS presents perhaps the most appropriate currently available framework for such a task. It should be noted that the use of the postcode system (and postcode-derived products) is increasingly being suggested where there is a requirement to cross-match socioeconomic data (Raper et al., 1992). For the present study, postcode locations are commercially available with grid references to 1 m precision, yet none of the datasets cited above are (fully) postcoded. It has therefore been necessary to develop an alternative method for the construction of a GIS database. Our experience elsewhere suggests that customer referencing systems used by most utilities also comprise hierarchical UPRNs, derived from delivery area and street segment information, but without any explicit requirement for postcoding of records.

Streets, and subdivisions of streets, provide an important intermediate stage in most urban spatial referencing systems. Street segments can be combined to form most levels of zonal hierarchies such as census zones and other small policy areas. However, street segments have figured little in the debates about basic spatial units due to the lack of a central directory or set of definitions, and the difficulty of tracing textual information in street addresses, which may be recorded in many different formats. Digitized street centrelines are often available commercially (e.g., OSCAR [Ordnance Survey, Southampton, UK]) or may already exist in a local authority context through in-house digitising of 1:1,250 and 1:2,500 series maps. These provide a clear link between address-type information and explicit grid references. It is interesting to contrast the relative unimportance of streets to date in the British context with their central

role in the highly successful U.S. street-based census geography using DIME and TIGER (U.S. Bureau of the Census, Washington, DC) (Cooke, 1989).

For purposes of our own research, it has therefore been necessary to develop an alternative method for the construction of a GIS database. Briefly, data on the total revenue paid in the last year of the rating system (1989/90), and under the three years of the community charge system (1990/91, 1991/92 and 1992/93) and the first year of the council tax (1993/4) are held at the street level for 920 streets in the Inner Area. Data are stored in a LOTUS 1-2-3 spreadsheet and linked through a 7-figure code derived from the City UPRNs to the digitized street network in Arc/Info (Figure 1). Aggregation of such data to various administrative units is then possible using the GIS in order to investigate a number of potential policy scenarios (Martin et al., 1992). The data are exported from LOTUS as an ASCII file, and then in the Tables module of Arc/Info, a template is defined and the ASCII file imported into this. The 7-figure street code is the "relational join" between this file and the digitized street network. Using Joinitem in Arc/Info, a series of attribute data has been attached, including:

1. the amount of revenue generated in each of the last 5 years of the various taxation regimes,
2. the number of electors,
3. the number of households,
4. the breakdown of modal council tax banding,
5. the breakdown of modelled modal council tax banding.

Where it was necessary to assign street information into small areas such as the House Condition Survey areas, a line-in-polygon analysis of the entire dataset was performed in order to identify those street segments which crossed polygon boundaries, and the attribute information from these segments was then divided between the polygons concerned. The absence of any directly georeferenced address information during this procedure made it impossible to determine the true proportions in which such information should be split. Similar problems were encountered and solutions invoked with respect to the small HCS area boundaries: we envisage still further problems as we incorporate 1991 Census data into our database, since enumeration district boundaries frequently run down streets and across junctions, placing houses and their occupants into two or more enumeration districts.

In general, a number of new data entities and analytical outputs have been possible by spatial integration of the data which could not have been performed using the existing databases in isolation or by simple cross-matching of lists. Indeed, the difficulty of list-matching for statistical purposes could largely be overcome by spatial manipulation of the data, although the full potential for this will only be achieved with a combination of individual property grid referencing, and unique property reference numbers, as considered below. The linked records have allowed us to evaluate a number of policy scenarios, including the likely impact of the council tax valuations, in advance of the official valuation list (Longley et al., 1993). In order to achieve some of the modelling which we have required, including the development of a methodology for the prediction of council tax bandings on the basis of our price survey information and rates registers, custom programming has been required. This particular process involved the use of a four-stage model for the assignment of each property in the inner areas to a valuation class, using the best of: a direct match with a sampled property; a weighted average of sampled properties in the same street; a weighted average of sampled properties in the same community, providing certain constraints were met; or, finally a regression model based on rateable values and community. The complexity involved in the construction of such a procedure to operate on 47,000 properties was sufficient to prompt us to move outside the GIS, and reimport the banding classifications once they were complete. Output such as Figure 4 has been generated using ArcPlot (ESRI, Redlands, CA), as the result of such a process.

ANALYTICAL DIFFICULTIES

The use of a GIS framework has been instrumental in allowing sufficient integration for the analysis of substantive issues in changing local taxation using the Cardiff datasets. However, our experience is also illustrative of many of the difficulties which beset any attempt to use existing large property databases within GIS in the UK context. Many of the issues which we have encountered here are echoes of those raised by Brusegard and Menger (1989). The issues raised may be considered under two headings: those attributable to data sources, and those associated with analytical operations.

Any attempt to handle property registers on this scale is bound to be beset by problems relating to overlaps and mismatches of various kinds. Each time a new register is included in the base for analysis, the number of properties which can uniquely be traced in every register is further reduced. Two key problems with the Cardiff data have proved to be the addition of relatively large numbers of new residential properties in the Cardiff Bay Development Area, and different definitions of subdivided properties between the rates and council tax registers, making the matching of HMOs problematic.

The lack of grid referencing at the level of individual properties, and the need to assign parts of street segments into different housing condition survey areas has inevitably led to a degree of "fuzziness" in the locational referencing (Openshaw, 1989). This is perhaps surprising in a project which has access to so much property-level information. The data problems we face provide good examples of most of the classes of "obvious" error identified by Burrough (1986), as shown in Table 2.

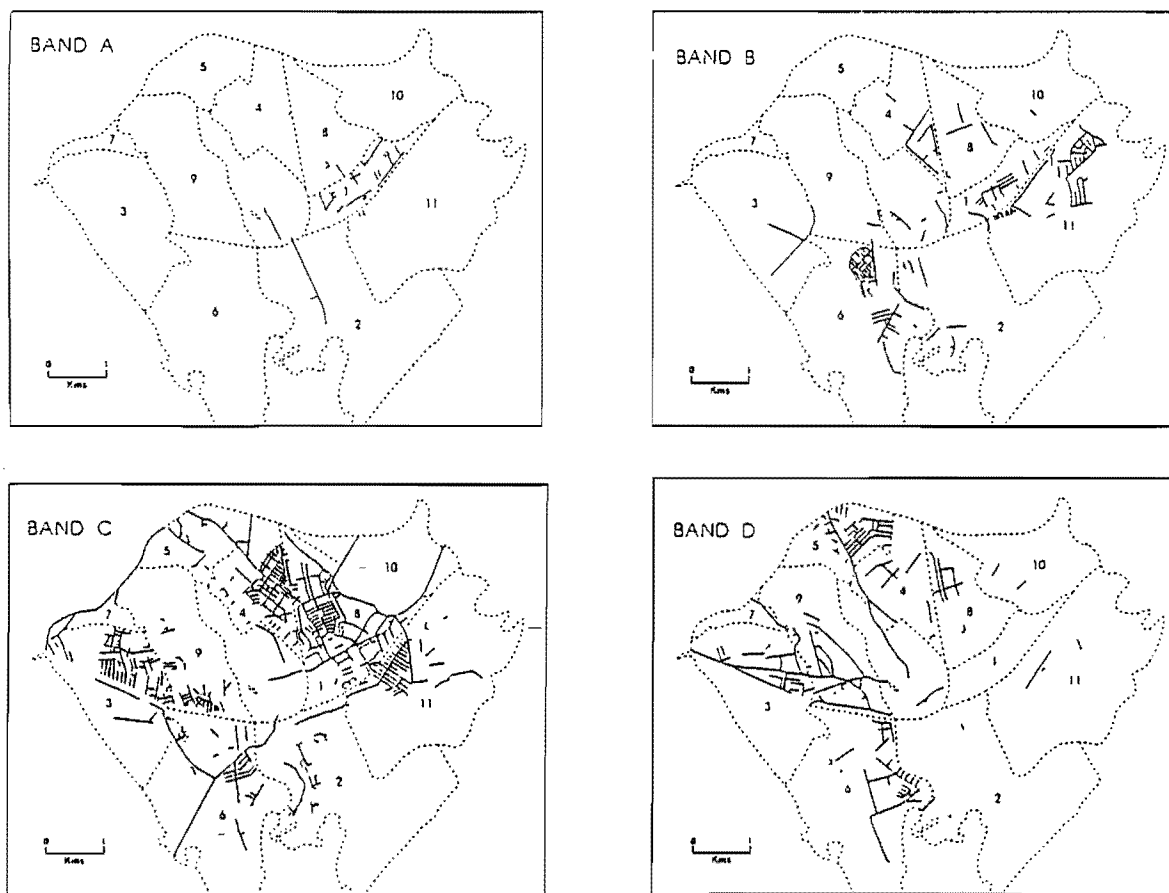


FIGURE 4. Predicted Council Tax Bandings, Bands A-D.

TABLE 2. Examples of Error Sources as Originally Identified by Burrough (1986)

Age of data	Address registers which predate recent demolition and construction
Areal coverage	Difficulty of obtaining price survey information for areas of primarily local authority-owned housing
Density of observations	Again, varying density of house price information, as determined by the property market and availability of estate agents' information
Relevance	Need to use electoral register as a surrogate for the community charge register, which is confidential and hence unavailable
Accessibility	Confidentiality concerns regarding council tax register, and housing condition survey information Cost implications of OS copyright, OSCAR and Pinpoint Address Codes (which we could not obtain)
Positional accuracy	Digitizing error in HCS areas and street centrelines
Accuracy of content	Obvious errors in textual address fields Hard to assess (e.g.), accuracy of rateable values

Some of the inconsistencies identified here are largely political in origin. For example, the community charge register, which was really required, is confidential. However, the electoral register, which is fundamentally very similar, is in the public domain. The difference between the two registers creates a source of error for GIS analysis, although it might be argued that for many areas they contain essentially the same information.

In addition to the (relatively) simple problems inherent in the datasets which we have been using, there is an additional class of problems which are associated with the analytical operations which we have performed. The GIS has been used in order to provide linkages between datasets based on spatial coincidence, where no other mechanism for linkage was available. This has created difficulties which are due entirely to the technique being used, and are unrelated to the substantive research problem. The line-in-polygon analysis for allocation of addresses in street segments provides a unique GIS-oriented approach to the matching of database entities, but is made unreliable due to the resolution of the input data. This arises because digital data were recorded by different digitisers from source material originally recorded at a variety of scales.

An entirely different set of issues relate to the difficulties of data representation, when dealing with address-referenced information. When the basic entities are points (addresses) and lines (street segments), presentation of complex patterns in the data is difficult, and we have resorted to the use of small multiples, as illustrated by Figure 4 (Tufte, 1990). Even here, we have been forced to assign modal property values to street segments for display. The GIS environment offers no adequate tools for the depiction of relationships between variables at the individual property level, and aggregation is almost always necessary simply in order to record results on screen or paper. In some cases, aggregation is also imposed by the requirement for confidentiality, and census-like restrictions on small numbers are necessary. These restrictions are frustrating in a context where there is the opportunity to work with detailed datasets, and to explore complex relationships at the level of the individual household. At various stages of our research, we have not been able to perform the statistical analysis or list matching required directly within the GIS, and it has been necessary to export the information to more general statistical packages or to write customised programs, for example, for the development of a predictive model of council tax bandings using price survey information.

neighbourhood characteristics, and a regression model. It has also proven expedient to export information in order to provide appropriate statistical comparison between rateable values from the rates register, and the council tax bandings from the council tax register.

FUTURE DEVELOPMENTS

In considering the future for research of the type outlined here, it is useful to consider both the specific future of the Cardiff databases, and the national context. We anticipate that the next stages of the Cardiff study will involve the integration of information from a House Condition Survey which was conducted in 1989. This comprises a sample of 20% of Inner Area properties with UPRN and HCS area identifiers, and includes detailed information about property type and condition. The information will provide important insights into the relationships between rateable values, council tax valuations, and housing condition. However, this detailed information will remain confidential at the level of the individual property, and it will therefore be necessary to perform data aggregation in one of the few areas where analysis is technically possible on individual records. An ideal scenario for the evaluation of the real impacts of successive local taxation regimes would be to trace individual properties through the different registers. As has been demonstrated above, the existing data sources have prevented this, and there are problems inherent in any attempt to match the different address-based registers.

It is suggested with increasing frequency that where there is a requirement to cross-match socioeconomic data, the postcode should form the basic spatial unit for such linkage (Raper et al., 1992). However, this review of the Cardiff study demonstrates the poor coverage of postcodes in the major registers, and suggests that the unit postcode may be too coarse for this purpose, although it may have utility for the aggregation and presentation of results.

The only true solution to these issues, given the existing registers, would appear to rest in the existence of a property database containing grid reference or street-segment locations at a higher resolution than any of the possibilities considered so far, and containing unique property references which distinguished subdivisions of addresses. Such a database is being created in the form of Ordnance Survey Address-Point, incorporating individual property referencing to 1 m precision. At this level of resolution, use of a GIS in order to establish spatial coincidence between properties and any other areal unit will be meaningful, as is (arguably) proximity to any particular street segment. Analysis of point patterns, and statistical analysis of individual-level relationships will be possible. Unlike the Pinpoint Address Codes created for parts of the country since 1987 (Gatrell, 1989), Address-Point will offer national coverage and routine updating, including the new OS Address-Point Reference (APR; Ordnance Survey, 1993). Even once Address-Point is available, the task of matching it against the existing registers will still be subject to all the difficulties noted above. It is unlikely that local authorities will have the resources or incentive to match Address-Point with old lists such as the pre-1990 rates register. The common feature of the two address-level grid referencing systems is their considerable commercial cost, a feature reflecting both the enormous input of resources required for their creation, and a symptom of the wider "commodification of information" identified by Goddard and Openshaw (1987, p. 1423). There is one potential resolution level which falls between area aggregation and individual property referencing, and that is the development of universal street-segment encoding, similar to that used in the U.S. TIGER system (Laurini & Thompson, 1992). The UK counterpart of this kind of database might be something akin to a digital Kelly's Directory, which represented every separately identifiable street segment and their house numbering sequence, and could be linked to digital roadline products such as OSCAR.

CONCLUSION

In this paper, we have reviewed our analysis of a large and complex urban database, which is being used in the conduct of an ongoing substantive research theme. The object of the project at the outset was to utilize GIS technology in order to advance our analysis and handle the integration of the databases. In reality, we have only been able to achieve this to a limited extent. Many of our outstanding difficulties relate to data problems which limit the analyses which are possible, or to the lack of appropriate tools within the GIS for tackling our ad hoc statistical and modelling needs. However, the development of more complex spatial analysis using GIS tools is less of a constraint than those issues relating to the data.

These issues are of importance, because of the complex processes of error propagation which occur during analysis of such datasets. Data on properties, households and individuals are a fundamental form of geographic information, and their use is growing. One of the biggest problems with such data sources is (and will continue to be) the proportion of mismatches which occur as successive thematic layers are added to the analysis. Our experience suggests that the solution lies either in national detailed georeferencing, or in a series of widely recognized hybrid units such as street segments, which would provide a form of "fuzzy geography" for the handling of these kinds of information. In order to be of assistance, national georeferencing systems must be sufficiently widely available (in terms of cost, data formats, etc.) to make their integration with existing lists such as the council tax register a viable option. The rates register derives from a time when the technology was not readily available to facilitate its use as a national basis for property referencing. Ironically, that technology is now available, but the two most important current datasets present severe problems to the researcher. There are as yet no plans to make the council tax register publicly available, and the new Address-Point product is likely to prove extremely expensive.

Acknowledgements: A version of this paper was presented at a conference on "Spatial Analysis in ArcInfo," at the University of Lancaster, June 1993. The authors are very grateful to Stephen Gill of Powys County Council, Wales, for his detailed comments on this preliminary version; the usual disclaimers apply. This research was funded by the Economic and Social Research Council Grant No. ROO0234707.

REFERENCES

- Brusegard, D. & Menger, G. (1989). Real data and real problems: Dealing with large spatial databases. In M. Goodchild & S. Gopal (Eds.), *Accuracy of spatial databases* (pp. 177-186). London: Taylor and Francis.
- Burrough, P. A. (1986). *Principles of geographical information systems for land resources assessment*. Oxford: Oxford University Press.
- Campbell, H. (1993). GIS implementation in British local government. In I. Masser & H. J. Onsrud (Eds.), *Diffusion and use of geographic information technologies*. Dordrecht, Netherlands: Kluwer.
- Cooke, D. (1989, July/August). TIGER and the "Post-GIS" ERA. *GIS World*, 53-55.
- Department of the Environment [DoE]. (1987). *Handling geographic information: The report of the Committee of Enquiry chaired by Lord Chorley*. London: Her Majesty's Stationery Office.
- Gatrell, A. C. (1989). On the spatial representation and accuracy of address-based data in the United Kingdom. *International Journal of Geographical Information Systems*, 3(4), 335-348.
- Gault, I., & Peutherer, D. (1990). Developing geographical information systems in local government in the UK: Case studies from Birmingham City Council and St. Helens Regional Council. In L. Woollam (Ed.), *Geographic information systems: Developments and applications* (pp. 109-132). London: Belhaven Press.
- Goddard, J. & Openshaw, S. (1987). Some implications of the commodification of information and the emerging information economy for applied geographical analysis in the United Kingdom. *Environment and Planning A*, 19, 1423-1439.
- Higgs, G., Longley, P., & Martin, D. (1991). An analysis of changing local taxation regimes using a street-level database. *Technical Reports in Geo-Information Systems, Computing and Cartography* (Vol. 34). Cardiff: Wales: South West Regional Research Laboratory.
- Hoineville, G., Lowell, R., and Associates. (1978). *Survey research practice*. London: Heinemann.
- Laurini, P., & Thompson, D. (1991). *Fundamentals of spatial information systems*. London: Academic Press.
- Longley, P., Martin, D., & Higgs, G. (1993). The geographical implications of changing local taxation regimes. *Transactions of the Institute of British Geographers*, 18B, 86-101.

- Martin, D. (1991). *Geographic information systems and their socioeconomic applications*. London: Routledge.
- Martin, D., Longley, P., & Higgs, G. (1992). The geographical incidence of local government revenues: An intraurban case study. *Environment and Planning C*, 10, 253-265.
- Openshaw, S. (1989). Learning to live with errors in spatial databases. In M. Goodchild & S. Gopal (Eds.), *Accuracy of spatial databases* (pp. 263-276). London: Taylor and Francis.
- Openshaw, S. (1990). Spatial referencing for the user in the 1990s. *Mapping Awareness*, 4(2), 24-29.
- Ordnance Survey. (1993). *Address Point demonstration disk 2*. Southampton, UK: Author.
- Paddison, R. (1989). Spatial effects of the Poll Tax: A preliminary analysis. *Public Policy and Administration*, 4(2), 10-21.
- Raper, J., Rhind, D., & Shepherd, J. (1992). *Postcodes: The new geography*. Harlow, UK: Longman.
- Tufte, E. R. (1990). *Envisioning information*. Cheshire, CT: Graphics Press.
- Worrall, I. (1990). GIS: Prospects and challenges. In I. Worrall (Ed.), *Geographic information systems: Developments and applications* (pp. 1-12). London: Belhaven Press.