Geographical Information Systems: on modelling and representation

Paul A. Longley

Department of Geography and Centre for Advanced Spatial Analysis, University College London, 1–19 Torrington Place, London WC1E 7HB, UK

I Progress in modelling and representation

Rapid though developments in geographic information (GI) handling technologies continue to be, digital representations of the real world within GIS will almost always necessarily remain partial and incomplete. This point was implicit in my last report (Longley, 2003) with regard to digital data infrastructures that are becoming ever richer and more detailed, yet inevitably remain deficient in detail. Not just with respect to data *per se*, but also with respect to the methods and technologies that are used to assemble them into representations of the real world, there is a sense that the research horizon continues to recede ahead of us, until the realization dawns that attempts to build the perfect representation are almost inevitably doomed to failure. Research effort must therefore refocus upon effective *management* of the uncertainties that are inherent in incomplete representational models of reality (see Miller, 2003). In this report I would like to consider the interplay between networked computer technologies and new approaches to modelling that are helping to further this objective.

For human geographers, it is nearly 40 years since the general notion of the 'model' as a simplification of reality – as opposed to the particular use of the term to describe scaled-down representations fashioned with physical materials – entered common parlance (Chorley and Haggett, 1967). The earliest models adapted a range of physical media to the representation of physical artifacts (as in planners' block models of towns) or abstract relations (as in the embodiment of Weber's model of industrial location in the Varignon Frame): however, as the digital revolution gathered pace, so simulation of mathematical relationships using computers increasingly became the norm. In considering the progress during the intervening decades, it remains helpful to revisit the distinction between *iconic* and *symbolic* models. Iconic models are best thought of as 'scaled-down versions of the real thing' and may have working parts that are similarly scaled down. The town planner's block model or the hydrologist's flume provide specific examples, but so too does the paper map. Symbolic models, by contrast, are based on logical (mathematical and/or statistical) relations and allow the less tangible attributes of the system of interest to be represented in the abstract. Spatial interaction models of retail systems provide good examples. Digital maps are superficially similarto their paper forebears, but the topological relations that they embody, and the abstract use of the concept of scale that they entail identify them as members of the family of symbolic models. The agenda of spatial analysis (Berry and Marble, 1968) can be thought of as driven by symbolic modelling activity.

The wide adoption of GIS has had the particular effect of rendering the traditional procedures of two-dimensional paper map production more or less obsolete, through the replacement of paper (iconic) base mapping with digital (symbolic) framework data (Rhind, 1997). Yet it is also having further and more profound effects, through spatial analysis, by blurring of the conceptual distinction between iconic and symbolic modelling. Today's GIS enables us to engage in symbolic representation not just of abstract systems but also of the physical environments within which symbolic interactions take place. GIS has always been more than digital encoding of paper maps. Developments in modelling and hardware are allowing GIS users to extend the map metaphor into three dimensions and, with the addition of time, into four. As such, GIS is extending geographic representation from the icon of the map to the icon of the building or landscape and is extending symbolic modelling from 2D to 3D. The general extension to 4D (Frank *et al*., 2000) may not be as smooth, but there nevertheless remains much to conjecture.

In general terms, the emerging picture of GIS is of powerful and realistic models embodied in data, networks, software and hardware that enable immediate visual communication to users. A range of developments is taking place upon these foundations. First, digital representations across time and space are being built for a far broader range of spatial and temporal scales than hitherto. Our improving abilities to manage the granularity of time and space, particularly at fine scales, can preserve the unique attributes of 'places' and particular time periods, while retaining the power to generalize about events that occur in different places and times (e.g., Fotheringham *et al*., 2002). Second, the organization of GIS through networks now links the smallest handheld devices to a global network of computing, which is greatly improving our ability to explore, simulate and analyse geographic phenomena at very disaggregate scales (e.g., Li and Maguire, 2003). Third, the immediacy of GIS as a communications medium continues to extend core uses of GIS from applied science to those that improve and promote wider public understanding of generalized spatial systems (Rhind, 1999). These three trends raise a number of methodological and substantive issues, which I would like to develop in the remainder of this report.

II Progress in modelling individuals, forms and processes

Humans and their activities are depicted in GIS as mobile point-referenced 'events' (Martin, 1996), and it has long been recognized throughout social science that the best ways to represent such events are grounded at the level of the individual (see, for example, Hensher and Johnson, 1981, for a discussion of the case of discrete choice modelling, and Golledge and Stimson, 1987 and 1997, for implications for behavioural geography). Only in the very recent history of GIS has our ability to capture and process digital data been up to the task of analysing changes in the locations and states of large numbers of individuals, and in human geography the earliest symbolic models were restricted in scope to coarse zonal aggregations. *Ad hoc* aggregation creates scientific quicksand: the scale and configuration of the basic ‘atoms’ of information – objects located in time and space – can exert critical influence on the outcomes of spatial analysis (Openshaw and Alvanides, 1999) and the inductive procedures of zone design can reinforce, obscure or even negate deductive procedures of inference concerning zonal distributions. A paradox in human (and much of physical) geography has been that aggregation is fundamental to generalization across space, yet the aggregated zones that are created to facilitate such generalization rarely have any validity independent of particular applications – if, indeed, they have any clear validity at all.

These problems are compounded in modelling when the focus is upon interaction, process or the representation of dynamics. Yet progress is clearly being made. Within the applications realm of retail modelling, for example, early symbolic models of spatial interactions were typically framed within limited coarse zonal geographies at the subregional scale (see Foot, 1981, for a pedagogic review); by the mid-1990s, data and computation were up to the task of building richer representations of parts of retail systems; today, retailers’ abilities to predict are constrained principally by the quality of input data, and the drive is towards microscale models of store pitch (Birkin *et al*., 2002). More generally, the combined effects of progress in data and simulation technologies are refocusing spatial analysis at the microscale, as with the development of models of individual pedestrian flows, in which human beings are represented as individual agents and behaviour is analysed through scenarios at scales from the architectural to the neighbourhood. Batty *et al*. (2003), for example, utilize this ‘agent-based modelling’ approach to develop plausible models of individual movement based around sensory attraction to features (in this particular instance, sound systems at the Notting Hill Carnival), within the constraints of built form and street geometry. This is spatial analysis founded upon unique individuals modelled as mobile individual point referenced events, and, as such, the application presents a logical endpoint of the drive towards disaggregation. Similar approaches are being developed using coarser, yet integral, elements in cellular automata research (Torrens and O’Sullivan, 2001), a kind of digital analogue modelling, which can be thought of in conceptual terms as a halfway house between iconic and symbolic forms. Like some agent-based modelling applications, the model foundations here are provided by analogy to some other system: thus the properties of geographic individuals (which may, for example, be individual land parcels of developed space or individual buildings) are deemed to interact according to presepecified rules, and the environment of GIS is used to appraise the dynamics of change in a range of spatially disaggregate scenarios. The nascent field of geosimulation (Benenson and Torrens, 2004) seems set to provide fertile ground for a reinvigorated approach to digital analogue modelling at the widest range of geographic scales.

III Disaggregation, networking and incompleteness

In my last progress report, I observed that improvements in digital mapping products and large-scale land-cover data have not been matched by commensurate developments in the creation of socio-economic digital data infrastructures (Longley, 2003). Here, I would like briefly to broach some of the technology issues that are associated with this developing asymmetry, and to flag some of their implications for GIS research. Fine-grained public-sector data sets, such as national censuses, today make up a smaller real share of current data available to human geographers, given the advent of the global positioning system (GPS) and the increasing activities of those involved in collecting and disseminating 'lifestyles' data (Harris and Longley, 2003).

The maturation of GPS technologies, coupled with the advent of portable GIS devices, allows much more detailed monitoring of the individual activity patterns of willing individuals, as well as tagging of individual characteristics and actions (Li and Maguire, 2003). In important respects, these developments are taking geography backto its roots in the measurement of conditions in the field, and the ability to track the individual potentially reinvigorates the field of behavioural geography (Golledge and Stimson, 1987; Clarke, 1998). There are clear potential applications in the development of experimental designs to investigate cognition although, more generally, there are few individuals who are likely to allow computers to monitor the intricacies of their daily activity patterns – and those that do are unlikely to be representative of populations at large. Thus, most GIS applications that require multi-attribute socio-economic data that are representative of populations, will remain constrained to a more limited range of data sources, aggregated in order to preserve respondent confidentiality.

Yet there are other ways in which underexploited sources of digital data may be assembled across networks in order to develop innovative disaggregate models. The field of geodemographics has become accustomed to fusing records pertaining to individuals with aggregations, as in the use of lifestyles data to freshen up and extend the remit of census-based geodemographic classifiers (Batey and Brown, 1995). Much time in analytical geography has been spent trying to bridge scales, and the geodemographic classification systems of the early 2000s are a successful response to the challenge to devise rich, pertinent indicators of social conditions at local scales for businesses and service planners. In this context, recent geodemographic research (Webber and Longley, 2003) has begun to exploit the very different methods that are traditionally used by government and business to estimate local demand or need and has begun to accommodate a variety of different levels of geography – households, unit postcodes, local communities and wider regions – into the model building effort. This is essentially a development of the multilevel approach (Goldstein, 1995), and begins to merge two contrasting approaches: a business (geodemographic) perspective that summarizes small area information on the basis of *social similarity* into clusters; and a government perspective that summarizes aggregate information for small areas that have been grouped together because of their *locational proximity*. These different approaches are rooted in different traditions, priorities and work practices, but fusion of data assembled and used for each are needed in order to realize synergies of approach. In essence, this hybrid approach can exploit the properties of social similarity and locational proximity, by examining those area effects that are not specified in small area geodemographic classifiers.

This is central to the art of modelling in GIS, since it addresses the need to specify and, hence, generalize about 'place' effects of present or past environments that are either unmeasured or (in practical terms) immeasurable. It is unlikely that the effects of place can ever be wholly encapsulated when specifying individual agents, and it seems axiomatic that we cannot capture the effects of all individuals in every place and all time periods. Even if we could, and even if limitless computer processing power were in our gift, it would be wasteful to do so. Instead, it makes sense to recognize that it isnot just the agents extant today that create and respond to place effects: in the spirit of Thrift's (1981) distinction between the active and reactive traditions to behavioural geography, it seems sensible to use GIS as a repository for models of place effects that encapsulate different event histories and clarify historic trajectories (e.g., Johnston and Griffiths, 1991), and to then appraise the behaviour of agents in this context.

The wider issue is how to develop such models from representations of the real world that are inevitably incomplete. Sometimes, as in the cellular models described by Benenson and Torrens (2004) and the agent-based movement models developed by Batty *et al*. (2003), it remains very challenging to observe and record data on individual decision-making that embody the rules for action and interaction. For example, data on where people originate and where they are destined are notoriously difficult to collect for large crowds and, although advances in laser scanning and remote sensing are able to provide estimates of aggregate volumes, there is no fix other than direct interaction with the objects and subjects of the simulation to generate the data required. Like all historical GIS applications (e.g., Gregory and Southall, 1998), the norm is that representations will be incomplete because relevant information was never collected, is fragmented or has been destroyed.

IV Visual communication and user interaction

Changes in computer architectures over the last decade have enabled not only networking of data collected at remote locations but also remote interrogation of databases from a distance across the same networks. As a consequence, GIS can be used to create an improved sense in remote locations of immediacy and participation: today, the near ubiquity of GIS in developed countries enables more people than ever to use them for accessing representations of different places and different times.

A challenge that this poses is that the sophistication of representations needs to be matched by the quality of guidance available to users. Improved methods of visualiza- tion and user interaction are key to this requirement. Haklay and Tobón (2003) have illustrated how public participation in GIS (PPGIS) can be encouraged through new techniques and procedures that foster clearer perception, improved interpretation, and more thorough interrogation of spatial data. A developing range of techniques for inter- rogating iconic models now exists, many of which allow users to uncover characteris- tics that are superficially hidden by the process of building representations. Twenty-five or more years ago the realization dawned that the audience for modelling papers halved with each successive mathematical equation. The visual medium of GIS is much more intuitive for most users, but the challenge is for intelligibility to at least keep pace with model complexity. 'Advanced' spatial analysis does not necessarily mean that outputs need appear more 'complicated': rather, the experience of ongoing research is that advanced interactive tools of spatial data exploration can be readily assimilated by an increasingly broad user base (see, for example, Kingston *et al*., 2000).

Some of these approaches are already used in fully fledged applications of GIS to Scenario development. The objectives of simultaneously disaggregating events and occurrences and of managing incompleteness of representation place conflicting demands on scenario developers, and these challenges are multiplying as the domainsof plausible GIS applications broaden. For example, in the domain of location-allocation modelling, Shietzelt and Densham's (2003) current research has encountered problems centred on the length of time required to refresh a dynamic simulation, while the improved ability to process data can result in the generation of a mind-bogglingly wide range of scenarios for emergency management under any conceivable system state. In these circumstances, just as selectivity of input data (for example, preprogramming the daily commuting rush) can be used to reduce the processing needed for real-time analysis, so there is also a human-computer interaction need to create a managed scenario base – in order to make the user selection of the most appropriate scenario from meta scenario categories as straightforward as possible.

Until recently, user interaction overwhelmingly entailed one-way traffic – that is, the user interrogated a GIS in order to obtain *information*. However, increasing numbers of users now utilize GIS as a medium for *participation* – that is, sharing ideas about spatial environments. Progress is currently most rapid in urban applications, through use of multimedia techniques to communicate in direct and immediate ways. PPGIS and its extensions into multimedia are increasingly being aided by developments in 3D GIS and CAD that are delivered to the public across networks rather than residing on the desktop – although less complicated representational methods such as panorama photographs are often more effective in representing the focal points of sites of interest. Effective visual communication also requires good networking, and it is in the general area of visualization that extensions to 3D and to the networked and wireless worlds are converging. Batty and Smith (2001) and Kingston *et al*. (2000) illustrate all of this with regard to web-based public participation in urban and environmental planning, respectively.

Extension of GIS to the third dimension relies heavily on developments in image refresh, rendering and graphics software. 3D and panorama imaging are translating more abstract information into a form that many nonexpert users can immediately understand. In this context, the term '3D model' embodies representations that are very different from the symbolic models that used to dominate quantitative geography and spatial analysis. The extension of iconic digital modelling from 2D to 3D is also extending the symbolic modelling of spatial relationships – the third dimension has hitherto been largely ignored in modelling the size, shape, scale and density of urban environments, and restricting measurements of urban land-use characteristics to those prevailing at the (natural or artificial) Earth surface obscures much of the variety of land use and urban function. 3D GIS make it possible to extend GIS functionality to analyse such diversity, while providing a much more complete picture of how cities work.

A final emergent research area concerns the ways in which 3D representations scale to smaller handheld devices. In computational terms, a major issue is how such devices, already used in the field and associated with field data capture and field location of static objects, can be refreshed to contribute to the challenges of navigating urban envi- ronments and wayfinding. This is part and parcel of bringing GIS closer to our direct experience and our need to understand whether and how location-based service (LBS) technologies will become embedded in human behaviour. The historic demarcation in psychology and behavioural geography (Johnston and Sidaway, 2004; Li and Maguire,

2003) between direct and indirect experience blurs when handheld devices are used as an adjunct to reality in the field. Spatial knowledge acquired through interaction with 2D maps ('map reading'), is usually taken to be the most advanced level of spatial knowledge. Yet with the extension of GIS to navigation of 3D environments, the implication is that GIS is as much a direct medium through which spatial knowledge might be acquired as a tool for data input and creation of spatial information bases. This potential will likely be further enhanced as multimedia websites develop to link photogrammetric data with 3D graphics and other forms of multimedia delivered across the web.

V Some prospects: symbolic modelling of the virtual

I have recently been involved in the assembly of a collection of papers written by researchers in one research-orientated spatial analysis laboratory (Longley and Batty 2003). Reflecting on our experience in compiling a collection with a not dissimilar remit the best part of a decade previously (Longley and Batty, 1996), we have dwelt at some length on the profound implications of the shift in computing from the desktop to the network. The additional research issues developed in this report are all part and parcel of this shift, for they engender linkage between different types of software, retrieval and sharing of data and a vastly improved ability to communicate not only among ourselves as researchers but also with the increasingly broad base of non-specialist users. GIS has also clearly become a medium in which physical models of systems – their geographies – are merging with symbolic models of these geographies through rich data and clearly specified spatial relationships. One extension to this line of thinking suggests a need not only to research the use and application of such symbolic models but also to understand the characteristics and objectives of users of them. The characteristics of the emergent 'E-Society' (www.london.edu/e-society) and the ways in which it harnesses networked GIS will be the focus of my next, and final, progress report.

Acknowledgement

This work was funded under an ESRC E-Society Programme grant on 'Digital differentiation: consumption profiles of fracturing digital divides'.

References

Batey, P. and **Brown, P.** 1995: From human ecology to customer targeting: the evolution of geodemographics. In Longley, P.A., Clarke,s G.P., editors, *GIS for business and service planning*, Cambridge: GeoInformation International, 77–103.

Batty, M., Desyllas, J. and **Duxbury, E.** 2003: Safety in numbers? modelling crowds and designing control for the Notting Hill Carnival. *Urban Studies* 40, 1573–90.

Batty, M. and **Smith, A.** 2001: Virtuality and cities: definitions, geographies, designs. In Fisher, P. and Unwin, D., editors, *Virtual reality in geography*, London: Taylor and Francis, 270–91.

Benenson, I. and **Torrens, P.M.** 2004: Editorial. Geosimulation: object-based modeling of urban phenomena. *Computers, Environment andUrban Systems* 28(1/2), 1–8.

Berry, B.J.L. and **Marble, D.F.,** editors 1968: *Spatial analysis: a reader in statistical geography*. Englewood Cliffs, NJ: Prentice Hall.

Birkin, M., Clarke, G.P. and **Clarke, M.** 2002: *Retail geography and intelligent network planning*.Chichester: Wiley.

Chorley, R. and **Haggett, P.,** editors 1967: *Models in geography*. London: Methuen.

Clarke, K. 1998: Visualising different geofutures. In Longley, P.A., Brooks, S.M., McDonnell, R. and Macmillan, B., editors, *Geocomputation: a primer*, Chichester: Wiley, 119–37.

Foot, D. 1981: *Operational urban models*. London: Methuen.

Fotheringham, A.S., Brunsdon, C. and **Charlton, M.** 2002: *Geographically weighted regression: the analysis of spatially varying relationships*. Chichester: Wiley.

Frank, A., Raper, J. and **Cheylan, J.-P.,** editors 2000: *Life and motion of socio-economic units*. London: Taylor and Francis.

Goldstein, H. 1995: *Multilevel statistical models* second edition. London: Arnold.

Golledge, R.G. and **Stimson, R.J.** 1987: *Analytical behavioural geography*. New York: Guilford Press.

–––– 1997: *Spatial behavior: a geographic perspective*. New York: Guilford Press.

Gregory, I. and **Southall, H.** 1998: Putting the past in its place: the Great Britain Historical GIS. In Carver, S., editor, *Integrated information:innovations in GIS 5*, London: Taylor & Francis, 210–21.

Haklay, M. and **Tobón, C.** 2003: Usability engineering and PPGIS: towards a user-centred approach. *International Journal of Geographical Information Science* 17(6), 577–92.

Harris, R.J. and **Longley, P.A.** 2003: Targeting clusters of deprivation within cities. In Clarke, G.P. and Stillwell, J., editors, *Applied GIS and spatial analysis*, Chichester: Wiley, 89–110.

Hensher, D.A. and **Johnson, L.W.** 1981: *Applied discrete choice modelling*. London: CroomHelm.

Johnston, R.J. and **Griffiths, M.** 1991: What's in a place? An approach to the concept of place as illustrated by the British National Union of Mineworkers Strike 1984–85. *Antipode* 23, 185–213.

Johnston, R.J. and **Sidaway, J.** 2004: *Geography and geographers* (sixth edition). London: Arnold.

Kingston, R., Carver, S., Evans, A. and **Turton, I.** 2000: Web-based public participation in geo-graphical information systems: an aid to local environmental decision-making. *Computers, Environment and Urban Systems* 24, 109–25.

Li, C. and **Maguire, D.J.** 2003: The handheld revolution: towards ubiquitous GIS. In Longley P.A. and Batty, M., editors, *Advanced spatial analysis: the CASA book of GIS*, Redlands, CA: ESRI Press, 193–210.

Longley, P.A. 2003: Geographical Information Systems: developments in socio-economic data infrastructures. *Progress in Human Geography* 27, 114–21.

Longley, P.A. and **Batty, M.,** editors 1996: *Spatial analysis: modelling in a GIS environment*. Cambridge: GeoInformation International.

––––, editors 2003: *Advanced spatial analysis: the CASA book of GIS*. Redlands, CA: ESRI Press.

Martin, D.J. 1996: *Geographic information systems: socio-economic applications* (second edition). London: Routledge.

Miller, H. 2003: What about people in geographic information science? *Computers, Environment and Urban Systems* 27, 447–53.

Openshaw, S. and **Alvanides, S.** 1999: Applyinggeocomputation to the analysis of spatial distributions. In Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W., editors, *Geographical information systems: principles, techniques, management and applications*, New York: Wiley, 267–82.

Rhind, D.W., editor 1997: *Framework for the world*. Cambridge: GeoInformation International.

–––– 1999: GIS management issues. In Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W., editors, *Geographical information systems: principles, techniques, management and applications*, New York: Wiley, 583–88.

Shietzelt, T. and **Densham, P.** 2003: Location- allocation in GIS. In Longley, P.A. and Batty, M., editors, *Advanced spatial analysis: the CASA book of GIS*, Redlands, CA: ESRI Press, 345–65.

Thrift, N. 1981: Behavioural geography. In Wrigley, N. and Bennett, R.J., editors, *Quantitative geography: a British view*, London: Routledge and Kegan Paul, 352–65.
Torrens, P.M. and **O'Sullivan, D.** 2001: Cellular automata and urban simulation: where do we go from here? *Environment and Planning B* 28,163–68.
Webber, R. and **Longley, P.A.** 2003: Geo-demographic analysis of similarity and proximity: their roles in the understanding of the geography of need. In Longley, P.A. and Batty, M., editors, *Advanced spatial analysis: the CASA book of GIS*, Redlands, CA: ESRI Press, 233–66.