Progress in Human Geography 22,2 (1998) pp. 283–292

Trends in quantitative methods II: stressing the computational

A. Stewart Fotheringham
Department of Geography, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK

I Introduction

In the first of my three reports on quantitative methods (Fotheringham, 1997), I described the recent trend in spatial analysis concerned with *local* as opposed to more traditional *global* analyses. The movement towards local statistics is actually part of a larger trend in quantitative analysis concerned with exploiting the massive increases in both computational power and available spatial data which have taken place over the last decade (Openshaw, 1994; 1995). Recently the term 'Geocomputation' has been coined to describe techniques, primarily quantitative, within geography that have been developed to take advantage of these increases in computer power and data (Openshaw and Abrahart, 1996; Openshaw and Openshaw, 1997; Openshaw *et al.*, 1998).

The term 'computation' carries two definitions. In the broader sense it refers to the use of a computer and therefore any type of analysis, be it quantitative or otherwise, could be described as 'computational' if it were undertaken on a computer. In the narrower and perhaps more prevalent use, computation refers to the act of counting, calculating, reckoning or estimating – all terms that invoke quantitative analysis. The term 'geocomputation' therefore refers to the computer-assisted quantitative analysis of spatial data. However, I shall further restrict use of the term 'geocomputation' to refer to quantitative spatial analysis *in which the computer plays a pivotal role*. This definition is meant to exclude fairly routine analyses of spatial data with standard statistical packages (for instance, running a regression program in SAS). I will try to demonstrate in the examples below some of the ways in which modern quantitative geography is being extended through the innovative use of computer resources. Under my definition of geocomputational analysis, the use of the computer *drives* the form of analysis undertaken rather than being just a convenient vehicle for the application of techniques developed independently of computers. Geocomputational techniques are therefore those which have been developed *with the computer in mind* and which exploit the large increases in computer power that have, and still are, being achieved.

 0309–1325(98)PH197PR

A simple example serves to distinguish between the two types of computer usage. Consider a correlation coefficient being calculated for two sets of spatially varying data – variable x and variable y. To assess the significance of the resulting correlation coefficient one could apply the standard formula for a t-statistic calculating the standard error of the correlation coefficient from a theoretical distribution – the procedure used in all standard statistical computer packages. I would not term this procedure as geocomputational because the computer is simply used to speed up a calculation developed well before computers were in use. An alternative, geocomputational, technique would be to derive an estimate of the standard error of the correlation coefficient by experimental methods. One such method would be to permute, independently and randomly, the x and y variables across the spatial zones and calculate a correlation coefficient for each permutation. With a sufficiently large number of such correlation coefficients (there is no reason why millions could not be computed but thousands or even hundreds are generally sufficient), an experimental distribution can be produced which allows statistical inferences to be made on the observed correlation coefficient. The probability of the observed value being produced by a random process is given by the proportion of experimentally derived correlation coefficients which have a value in excess of the observed value. In this case, computational power is used to replace a theoretical distribution; the advantage being the avoidance of the assumptions underlying the theoretical distribution which may not be met, particularly with spatial data. More details on experimental significance tests are given by Diaconis and Efron (1983), Efron and Gong (1983) and Mooney and Duval (1993).

II Geocomputation and philosophical debates

While I remain sceptical of claims regarding the one true philosophical approach (which usually excludes the use of quantitative methods in any form) to understanding parts of human geography, there are none the less some interesting philosophical issues that underpin geocomputational analysis. It is hoped those who espouse one of the myriad of anti-naturalism philosophies (Graham, 1997) will not fail to recognize that much of what geocomputational analysis has to offer is in fact the opposite of the 'geography is physics' approach which is sometimes mistaken as a credo of quantitative geography. A premise of much geocomputational analysis is that processes and relationships are *not* necessarily the same all over and that it is of interest to describe such spatial variations as an aid to better understanding of spatial processes. Some of the geocomputational techniques described below have been designed explicitly to investigate *local* rather than *global* relationships. Admittedly, the focus on local exceptions and anomalies can be used as a means of improving global models and so would not escape anti-naturalist critiques (which only apply to human geography). Equally, though, local models can be used with the a priori assumption that there are intrinsically different types of behaviour over space and that no global model exists.

Within quantitative geography, geocomputational analysis is at the centre of two related debates. In the first, that concerning the use of *confirmatory* versus *exploratory* techniques, geocomputation is often seen as being in the latter camp. Confirmatory techniques emphasize hypothesis testing and the calibration of exogenously derived models of spatial processes while exploratory techniques are data, and often computer, intensive and are used to *suggest* hypotheses from an analysis of the data. The latter can be used to uncover interesting patterns in the data or to highlight exceptions and are therefore often viewed as a precursor to more formal modelling and analysis. However, computationally intensive calibration techniques exist, such as neural networks (see below), which could in some instances be thought of as a confirmatory technique.

The second debate is that between those who support deductive reasoning and those who believe in inductive techniques. The main role of geocomputational techniques, as described below, is through induction: promoting and displaying aspects of the data to suggest new hypotheses concerning the spatial processes which have produced the data. However, rather controversially, geocomputational techniques have also been applied to model building (*inter alia* Openshaw, 1983; 1988; Diplock, 1996) as a means of by-passing traditional deductive logic. In such applications, a large number of combinations of variables and functional forms is examined to uncover the model form which produces the most accurate fit to the data. Such applications have not found widespread favour (*inter alia* Veneris, 1984) because the data clearly drive the form of the resultant model and there is no guarantee that anything resembling the same model will result from a different data set. Even from the same data set, many different models could be produced which fit the data reasonably well and slight alterations in the goodness-of-fit criterion used to drive model selection can then produce very different models. In the model-building applications that have been undertaken some rather strange forms of models have been produced (Openshaw, 1983).

In the remainder of this article four applications of geocomputational analysis are described. The first example concerns GIS-based spatial analysis; the second describes computational issues surrounding the modifiable areal unit problem; the third discusses computational issues in geographically weighted regression; and the fourth covers what has become known as 'artificial intelligence' – computer-intensive mimicking of the way the brain works.

III Example 1: GIS-based spatial analysis

Geographic information systems (GIS) provide potentially very powerful tools for geocomputational analysis. They allow the storage, manipulation and mapping of large volumes of spatial data. Over the last two decades various linkages have been established between GIS software and statistical analysis packages in order to facilitate geocomputational analysis. As early as 1973 the Chicago Area Transportation Study used an IBM mainframe interactive graphics package called INTRANS to display and evaluate planning data generated by transportation models (Harper and Manheim, 1990). Initial attempts to integrate analytical packages with GIS fall into what Anselin *et al*. (1993) refer to as one-directional or static integration, where the results of one operation are fed into the other with no feedback. More advanced integration between GIS and spatial analysis involves bi-directional connections where there is two-way interaction between the GIS and spatial analytical routines (such as where data are derived from a GIS to calculate a statistic and then the statistic is imported back into the GIS for mapping). An example of such integration is SpaceStat (Anselin, 1990) – a software package for the statistical analysis of spatial data that can be hooked on to a variety of GIS packages. The most advanced coupling between GIS and spatial analysis comes through dynamic integration where movement between the two is continuous; an example being the brushing of data in a scatterplot in one window and the automatic referencing of those points on a map in another window. Examples of this type of integration are provided by the ESDA Arc/Info package developed by Xia and Fotheringham (1993) and the Spatial Analysis Module (SAM) described in Ding and Fotheringham (1992). In the latter software the user can open up to five linked windows for analysis and at least one window can contain a map of the study area. Observations brushed in any window will automatically be highlighted in the other four windows.

As a result of these initiatives and a multitude of calls for the development of more sophisticated spatial analytical tools to be incorporated within GIS (*inter alia* Goodchild, 1987; Rhind, 1988; Burrough, 1990; Fotheringham and Charlton, 1994), GIS vendors are now adding such tools to systems that previously only performed basic query, buffer and overlay routines. For instance, GIS-Plus has a suite of transportation-related models; SPANS has, among other things, a retail analysis package; SAS have developed their own GIS; and S-PLUS, an advanced statistical package for both exploratory and confirmatory analysis, can now be run through AML commands under Arc/Info which gives the user access to over 1400 statistical functions.

In terms of new areas of geocomputation it is true that simply providing the means for greater linkages between mapping and analysis does not guarantee greater insights into spatial processes. In many cases it could be argued that it is not essential to integrate the statistical software with a GIS. However, for exploratory analysis, the ability to move seamlessly between the analytical and the mapping software and the ability to interrogate the mapped data produce a reasonably high probability of producing insights that would otherwise be missed if spatial data were not analysed within a GIS. Consequently, GIS-based computational analysis will continue to grow and is likely to become the dominant means of geocomputation. The development of new ways of interacting with, and displaying, both spatial data and the results of spatial analyses provides a very fertile and exciting area into which geocomputational analysis will continue to expand (Fotheringham and Rogerson, 1993; 1994; Fischer *et al*., 1996).

IV Example 2: the modifiable areal unit problem

Spatial analysis frequently involves the use of areal data, a common example being the analysis of census data reported at various spatial scales (such as enumeration districts and wards in the UK and block groups and census tracts in the USA). One of the most stubborn problems related to the use of areal data is sometimes referred to as the zone definition problem or the modifiable areal unit problem (MAUP) and which relates to the sensitivity of analytical results to the definition of the spatial units for which data are reported (Openshaw, 1984; Fotheringham and Wong, 1991; Waller and Turnbull, 1993; Green and Flowerdew, 1996). The implications of this problem are potentially severe: if the conclusions reached from an analysis of aggregate spatial data reported for one set of zones differ from those reached when data are reported for a different arrangement of zones, then how reliable can any one analysis be as a means of uncovering knowledge on spatial processes?

There have been recent attempts to provide analytical ‘solutions’ to the MAUP such as those by Arbia (1989), Wrigley (1995), Holt *et al.* (1996) and Steel and Holt (1996), although they have not found general acceptance and still rely on inferences to empirical regularities. An alternative computationally intensive ‘solution’ to this problem is that demonstrated by Fotheringham and Wong (1991) who provide analytical results not just for one set of zones but for a variety of zoning systems. By comparing the results across a large variety of different zoning systems, the stability or instability of a particular result can be assessed visually or statistically. Results (for example, parameter estimates from a regression model) that are relatively stable to variations in reporting units are more reliable, *ceteris paribus*, than those which are relatively unstable. With data drawn from census units in the Buffalo Metropolitan Area, the relationship between mean family income and a series of independent variables is examined within both a linear and nonlinear modelling framework. The relationship with the percentage of elderly in a spatial unit is most interesting: by varying the spatial scale at which the data are collected, the parameter estimate for the elderly variable is consistently insignificant when the data are obtained from systems of 800 zones but is consistently significant and negative when the data are aggregated to 200 or fewer zones. Thus, two very different interpretations can be drawn from the same underlying data, which is clearly worrying!

A similar inconsistency is found even when scale (that is, the number of zones) remains constant but the arrangement of these zones is allowed to vary. An examination of the parameter estimate for the elderly variable estimated with 150 different spatial arrangements of the data at the same spatial scale produced the following results: the majority of the zoning systems yield the conclusion that there is no significant relationship between mean family income and the proportion of the elderly; a substantial number of zoning systems yield the conclusion that there is a significant *negative* relationship between the two variables; and two zoning systems yield the conclusion that there is a significant *positive* relationship between income and the elderly! It should be noted that the results reported by Fotheringham and Wong (1991), as well as being computationally intensive, rely heavily on the ability to combine large numbers of zones into realistic aggregates large numbers of times based on their topological relationships and that this is greatly facilitated by having the data stored in a GIS.

A similar, computationally intensive sensitivity analysis is described by Fotheringham *et al.* (1995) for a set of techniques for locational analysis known as location–allocation modelling. These techniques provide information not only on the optimal locations for a set of facilities but also on the demand that is served by each facility. Common to almost all applications of location–allocation modelling is that demand is computed for a set of aggregate zones. An issue analogous to that investigated by Fotheringham and Wong (1991) is to what extent the outcome of a location–allocation procedure is affected by the particular way in which demand is aggregated. This is a particularly relevant question in location-allocation modelling because the outputs from such a procedure are taken to be the *optimal locations* for a set of facilities and the *optimal allocation* of demand to those facilities and consequently often carry a great deal of weight in decision-making. However, if the results can be varied substantially simply by varying the scale of the analysis or by altering the arrangement of the reporting units for which the data are collected, as Fotheringham *et al.* (1995) demonstrate, the word 'optimal' in this situation would need to be used with a great deal of caution.

Given that policy decisions are often guided by the analysis of spatial data in aggregate zones and that the results of such analysis appear to be dependent on the nature of the zoning system used, there is an increasing need for computationally intensive techniques such as those described above. In order to provide a convincing set of results for any spatial analysis using aggregated data, it is necessary to demonstrate that the results are likely to hold regardless of the type of zoning system used. If consistency cannot be demonstrated then the results may be mere artifacts of the particular zoning system used instead of reflecting any underlying process.

V Example 3: computational issues in geographically weighted regression

Discussion of the background to geographically weighted regression (GWR) can be found in last year's report on quantitative methods (Fotheringham, 1997) and so is dealt with only extremely briefly and is confined to the computational aspects of the technique. GWR is a relatively simple, although computationally complex, procedure that extends the traditional global regression framework by allowing local rather than global parameters to be estimated. The model has the general form:

$$y_i = a_{io} + \Sigma_k a_{ik} x_{ik} + \varepsilon_i$$

where y represents the dependent variable, x_k represents the kth independent variable, ε represents an error term and a_{ik} is the value of the kth parameter at location i. The parameters of the model represent the nature of the relationship between y and each x around each point i. In the calibration of this model it is therefore assumed that observed data near to point i have more influence in the estimation of the a_{ik}s than do data located further from point i. Hence, the calibration procedure is more complex than with ordinary regression: although a weighted least squares approach is used, the data are weighted according to their location with respect to point i and therefore the weights vary with each point i rather than remaining fixed. The estimator for the parameters in GWR is:

$$a_i = (x^t w_i x)^{-1} x^t w_i y$$

where w_i is an n by n matrix whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of observed data for point i.

Operationally and computationally, the GWR framework incorporates some interesting issues connected with spatial processes. The first is the definition and calibration of the spatial weighting function. A weighting function has to be defined which weights data in close proximity to point i more than data which are further away. One possibility is to set the weights equal to one within a prespecified distance of point i and zero beyond this distance. This is relatively easy to compute but as it assumes a discrete spatial process, it is perhaps not very realistic for most processes. An example of this type of weighting function is given in Fotheringham *et al.* (1996) and in Charlton *et al.* (1997). A more realistic, but more computationally intensive, weighting function is a continuous decreasing function of distance in which a distance-decay parameter is calibrated for the function (Brunsdon *et al.*, 1996; Fotheringham *et al*., 1997a; 1997b).

Given that a weighting function and, where necessary, a calibration procedure have been selected, a further element of realism and computational complexity can be added by allowing the weighting function to vary spatially. That is, in what is described above, a global weighting function is calibrated but it is possible to allow the function to vary across space, presumably with the kernel becoming larger in sparser areas and smaller in more populated areas. There are several ways in which a spatially adaptive kernel can be computed. One is to allow a point-specific distance-decay parameter to be calibrated. A second is to base the weighting function on the x nearest neighbours of point i and so a continuous distance-based measure is replaced with an ordinal topological measure. A third is to set a constraint so that every point has the same sum of weighted data. This creates a weighting function with the added attraction of having a constant number of 'virtual' degrees of freedom. This constant could be given exogenously or, with yet another computational complexity, it could be calibrated within the GWR routine. One further computational complexity is to allow the weighting functions in GWR to vary across the parameters and even to allow 'mixed' models in which some of the parameters are fixed over space while the others are allowed to vary spatially.

Clearly, this is an exciting frontier of geocomputational analysis because it allows us to input the types of spatial processes we think operate in a given situation. It also provides us with a great deal of additional *spatial* output in the forms of maps of parameter estimates, goodness-of-fit statistics and other regression diagnostics. Issues of space are brought to the fore in the previously aspatial, yet very powerful, statistical technique of regression.

VI Example 4: artificial intelligence

Artificial intelligence (AI) is concerned with the computer replication of characteristics, such as learning and reasoning, normally associated with human intelligence. There are many components to AI but three of the main ones are heuristic searches, neurocomputing and evolutionary computing (Diplock, 1996; Openshaw and Openshaw, 1997).

Examples of heuristic searches in geography include the newer forms of point pattern analysis as described by Openshaw *et al.* (1987) and Fotheringham and Zhan (1996), and which were discussed in Fotheringham (1997). The essence of these techniques is that local, rather than global point pattern analysis is undertaken to identify locally interesting clusters of points.

Neurocomputing, often equated with neural networks, is concerned with simulating processes operating in the human brain. Neural nets consist of sets of nodes between which weighted connections are established (Zahedi, 1991). The networks are 'trained' or calibrated with data on inputs and outputs. Supervised networks can be trained to represent connections between inputs and outputs with a priori knowledge embedded in them; unsupervised nets contain no such a priori knowledge. Examples of the use of neural nets in geography for model calibration can be found in Fischer (1994) and Fischer and Gopal (1994).

Evolutionary computing techniques involve replicating aspects of evolution, and so are used to represent the growth of an object. The work of Batty *et al*. (1989) and Fotheringham *et al*. (1989) on simulating urban growth through diffusion-limited aggregation and that of Wong and Fotheringham (1990) on the development of urban systems provides early examples of evolutionary computing techniques applied to geographical problems. More recently, Diplock (1996) and Diplock and Openshaw (1996) have applied genetic algorithms to the task of building spatial interaction models using strings of characters to represent solutions to the problem as an analogy to the operation of chromosomes. Currently, the cutting edge of research in this area is on genetic programming pioneered by Koza (1992), which attempts to provide computer programs flexible enough to be applied to a multitude of problems. The ultimate aim is to produce computer programs which can solve problems for which they have not been programmed (Diplock, 1996; Openshaw and Openshaw, 1997).

VII Summary

There are at least two constraints on undertaking quantitative empirical research within geography. One is our ability to think about how spatial processes operate and to produce insights which lead to improved forms of spatial models. The other is the set of tools we have to test and refine these models. These tools might be used for data collection (e.g., GPS receivers, weather stations, stream gauges, etc.) or for data analysis (computers). In the early stages of computer use, it was relatively easy to derive models which could not be implemented because of the lack of computer power. This was an era when the second constraint was more binding than the first: the level of technology lagged behind our ability to think spatially. We are now no longer in this era. We are now in a situation where the critical constraint is more likely to be our ability to derive new ways of modelling spatial processes and analysing spatial data. The increases in computer power within the last 20 years have been so enormous that the technological constraint is much less binding than it once was. The challenge now is to make full use of the technology to improve our understanding of spatial processes. It is hoped the examples given above suggest ways in which this is being done. However, we are just at the beginning of this new era of virtually unconstrained computing power and there is still much to be done in revamping our analytical methods to take advantage of the situation. In many instances the change is so profound that it can alter our whole way of thinking about issues: the development of experimental significance testing procedures and the subsequent decline in the reliance on theoretical distributions is a case in point. The movement from global modelling to local modelling is another. Who knows what changes the next decade will bring? This is an exciting time to be involved in spatial analysis.

References

Anselin, L. 1990: *SpaceStat: a program for the statistical analysis of spatial data*. Santa Barbara, CA: NCGIA Publications, Department of Geography, University of California at Santa Barbara.

Anselin, L., Dodson, R.F. and **Hudak, S.** 1993: Linking GIS and spatial data analysis in practice. *Geographical Systems* 1, 3–23.

Arbia, G. 1989: *Spatial data configuration in statistical analysis of regional systems*. Dordrecht: Kluwer.

Batty, M., Longley, P.A. and **Fotheringham, A.S.** 1989: Urban growth and form:scaling, fractal geometry and diffusion-limited aggregation. *Environment and Planning A* 21, 1447–72.

Brunsdon, C., Fotheringham, A.S. and **Charlton, M.E.** 1996: Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, 28, 281–98.

Burrough, P.A. 1990: Methods of spatial analysis in GIS. *International Journal of Geographical Information Systems* 4, 221–23.

Charlton, M.E., Fotheringham, A.S. and **Brunsdon, C.** 1997: The geography of relationships: an investigation of spatial non-stationarity. In Bocquet-Appel, J.-P., Courgeau, D. and Pumain, D., editors, *Spatial analysis of biodemographic data*, Montrouge: John Libbey Eurotext, 23–47.

Diaconis, P. and **Efron, B.** 1983: Computer intensive methods in statistics. *Scientific American* 248, 5116–30.

Ding, Y. and **Fotheringham, A.S.** 1992: The integration of spatial analysis and GIS. *Computers, Environment and Urban Systems* 16, 3–19.

Diplock, G. 1996: The application of evolutionary computing techniques to spatial interaction modelling. PhD thesis, School of Geography, University of Leeds.

Diplock, G. and **Openshaw, S.** 1996: Using simple genetic algorithms to calibrate spatial interaction models. *Geographical Analysis* 28, 262–79.

Efron, B. and **Gong, G.** 1983: A leisurely look at the bootstrap, the jacknife and cross-validation. *American Statistician* 37, 36–48.

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, 221–35.

Fischer, M.M. and **Gopal, S.** 1994: Artificial neural networks: a new approach to modeling interregional telecommunication flows. *Journal of Regional Science* 34, 503–27.

Fischer M., Scholten, H.J. and **Unwin, D.**, editors, 1996: *Spatial analytical perspectives on GIS*. London: Taylor & Francis.

Fotheringham, A. S. 1997: Trends in quantitative methods. I. Stressing the local. *Progress in Human Geography* 21, 88–96.

Fotheringham, A.S., Batty, M. and **Longley, P.A.** 1989: Diffusion-limited aggregation and the fractal nature of urban growth. *Papers of the Regional Science Association* 67, 55–69.

Fotheringham, A.S. and **Charlton, M.E.** 1994: GIS and exploratory spatial data analysis: an overview of some research issues. *Geographical Systems* 1, 315–27.

Fotheringham, A.S., Charlton, M.E. and **Brunsdon, C.** 1996: The geography of parameter space: an investigation into spatial non-stationarity. *International Journal of Geographic Information Systems* 10, 605–27.

—— 1997a: Measuring spatial variations in relationships with geographically weighted regression. In Fischer, M.M. and Getis, A., editors, *Recent developments in spatial analysis, spatial statistics, behavioural modeling and neurocomputing*, London: Springer-Verlag, in press.

—— 1997b: Two techniques for exploring non-stationarity in geographical data. *Geographical Systems* 4, 59–82.

—— 1995: The zone definition problem and location-allocation modeling. *Geographical Analysis* 27, 60–77.

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

—— , editors, 1994: *Spatial analysis and GIS*. London: Taylor & Francis.

Fotheringham, A.S. and **Wong, D.** 1991: The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A* 23, 1025–44.

Fotheringham, A.S. and **Zhan, F.** 1996: A comparison of three exploratory methods for cluster detection in spatial point patterns. *Geographical Analysis* 28, 200–18.

Goodchild, M.F. 1987: A spatial analytical perspective on geographic information systems. *International Journal of Geographical Information Systems* 1, 327–34.

Graham, E. 1997: Philosophies underlying human geography research. In Flowerdew, R. and Martin, D., editors, *Methods in human geography: a guide for students doing a research project*, Harlow: Addison Wesley Longman, 6–30.

Green, M. and **Flowerdew, R.** 1996: New evidence on the modifiable areal unit problem. In Longley, P. and Batty, M., editors, *Spatial analysis: modelling in a GIS environment*, Cambridge: GeoInformation International, 41–54.

Harper, E.A. and **Manheim, M.L.** 1990: Geographic information systems in transportation planning: a case for a geographic-based information services strategy. *Regional Development Dialogue* 11, 188–212.

Holt, D., Steel, D.G. and **Tranmer, M.** 1996: Area homogeneity and the modifiable areal unit problem. *Geographical Systems* 3, 181–200.

Koza, J.R. 1992: *Genetic programming: on the programming of computers by means of natural selection*, Cambridge, MA: MIT Press.

Mooney, C.Z. and **Duval, R.D.** 1993: *Bootstrapping: a nonparametric approach to statistical inference*. *Sage Series in Quantitative Applications in the Social Sciences*. Newbury Park, CA: Sage.

Openshaw, S. 1983: From data crunching to model crunching: the dawn of a new era. *Environment and Planning A* 15, 1011–13.

—— 1984: *The modifiable areal unit problem. Concepts and Techniques in Modern Geography* 38. Norwich: Geo Books.

—— 1988: Building and automated modelling system to explore a universe of spatial interaction models. *Geographical Analysis* 20, 31–46.

—— 1994: Computational human geography: towards a research agenda. *Environment and Planning A* 26, 499–505.

—— 1995: Human systems modelling as a new grand challenge area in science. *Environment and Planning A* 27, 159–64.

Openshaw, S. and **Abrahart, R.J.** 1996: Geocomputation. Abstracted in *Proceedings GeoComputation '96, 1st international conference on geocomputation*, University of Leeds, 17–19 September, 665–66.

Openshaw, S., Abrahart, R.J. and **Harris, T.E.**, editors, 1998: *Geocomputation*. Reading: Gordon & Breach.

Openshaw, S., Charlton, M.E., Wymer, C. and **Craft, A.W.** 1987: A mark I geographical analysis machine for the automated analysis of point data sets. *International Journal of Geographical Information Systems* 1, 359–77.

Openshaw, S. and **Openshaw, C.** 1997: *Artificial intelligence in geography*. Chichester: Wiley.

Rhind, D. 1988: A GIS research agenda. *International Journal of Geographical Information Systems* 2, 23–28.

Steel, D.G. and **Holt, D.** 1996: Rules for random aggregation. *Environment and Planning A* 28, 957–78.

Veneris, Y. 1984: Computerised dreams cannot make positivism creative! *Environment and Planning A* 16, 269–72.

Waller, L.A. and **Turnbull, B.W.** 1993: The effects of scale on tests for disease clustering. *Statistics in Medicine* 12, 1869–83.

Wong, D. and **Fotheringham, A.S.** 1990: Urban systems as examples of bounded chaos: exploring the relationship between fractal dimension, rank-size, and rural-to-urban migration. *Geografiska Annaler* 72B, 89–99.

Wrigley, N. 1995: Revisiting the modifiable areal unit problem and the ecological fallacy. In Cliff, A.D., Gould, P.R., Hoare, A.G. and Thrift, N.J., editors, *Diffusing geography* London: Blackwell, 49–71.

Xia, F.F. and **Fotheringham, A.S.** 1993: Exploratory spatial data analysis with GIS: the development of the ESDA module under Arc/Info. *GIS/LIS '93 Proceedings* 2, 801–10.

Zahedi, F. 1991: An introduction to neural networks and a comparison with artificial intelligence and expert systems. *Interfaces* 21, 25–38.