*Environment and Planning A* 1998, volume 30, pages 1835–1838

# Guest editorial

## Geocomputation: solving geographical problems with new computing power

Geocomputation can be defined as the process of applying computing technology to geographical problems. It could be argued that geocomputation has been around for almost as long as computers have been in general scientific use, and certainly since the 1960s when standard mainframe computers could be used. However, for a long while the size and complexity of the computations that could be handled were rather limited, both by central processor memory and by central processor speed. Most applications involved using statistical or mathematical methods that existed prior to computers with data sets or problem sizes that could not feasibly be handled manually.

The collection of papers published in this issue and continued in the next shows how things have moved on since the 1960s as a result of the ever-increasing speed of both scalar computers and because of the availability of parallel computer architectures. The papers are focused on methods, but with illustrations of their geographic application.

The methods covered include:
neural networks (Fischer, 1998; Mitchell et al, 1998; Openshaw, 1998),
genetic programming (Diplock, 1998; Openshaw, 1998; Turton and Openshaw, 1998),
simulation using cellular automata (Batty, 1998),
simulation using random sampling of model parameters (Turton and Openshaw, 1998),
fuzzy logic modelling (Openshaw, 1998),
geographically weighted regression (Fotheringham et al, 1998).

The geographic problems to which these methods are applied range widely:
(a) the prediction of spatial interaction between origins and destinations—of telephone traffic between exchanges in Austria (Fischer, 1998), or journey-to-work flows in Durham (Diplock, 1998);
(b) the projection of European Union regional populations (Turton and Openshaw, 1998);
(c) the factors explaining the spatial distribution of limiting long-term illness (see Fotheringham et al, 1998);
(d) the reconstruction of household types for small areas within cities (Mitchell et al, 1998); and
(e) the evolution of urban form over time (Batty, 1998).

### The potential for geocomputation

Turton and Openshaw (1998) review the potential that high-performance computing (HPC) releases for geographical analysis, arguing that there are enormous opportunities waiting to be tapped. They provide illustrations of that potential from their own work on porting existing model code from serial computers to the T3D parallel computer at the University of Edinburgh. Examples include the ability to calibrate spatial interaction models with very large numbers of origins and destinations, and the capacity to search among a very large number of possible solutions for optimal locations for 60 stores across 822 sites. Another application involves the computation of confidence limits around population forecasts by sampling distributions of the driving parameters a very large number of times. Finally, the design of zones for analysis purposes is a combinatorial optimisation problem discussed by the authors which can be exhaustively examined by using high-performance computing.

**New methods for modelling spatial interaction**
Openshaw (1998) separates the problem of predicting the flows between origins and destinations into accounting and interaction tasks. The first task is to ensure that predictions match known information on external predictions of total flows leaving origins or arriving at destinations. The second task focuses on the prediction of the interaction of origin with destination. New technologies can then be used, he argues, to search for the best functional form that converts available measures of impedance between origin and destination into interaction predictions. He describes a genetic model breeder in which the pieces of genetic code that mutate or cross-over between generations are mathematical operators or functions rather than genes or binary digits. This technology is described in detail by Diplock (1998). A second technique applied by Openshaw to spatial interaction models is that of artificial neural networks which are trained to match input variables to knowledge in a training data set and the resulting weights are used in prediction. This technology is explained at greater length by Fischer (1998). A final set of methods involves fuzzy logic: sets of simple rules that specify, for example, that you get big flows with small impedances and small flows with big impedances.

**Computational neural networks**
Fischer (1998) provides a systematic exposition of computational neural networks. Readers will find this a very useful introduction to a complex field. The properties of a neural network—its processing elements, its topology, and its learning method—are discussed. A fourfold taxonomy of computational neural networks is proposed which will help users in choosing the particular technique to use.

**Genetic programming**
Diplock (1998) provides a detailed account of this new method which applies the idea of a genetic algorithm not to parts of objects (for example, genes) but to equation components, building these components up into new model functions. The idea is to generate models of spatial interaction which would not have been 'thought up' by practitioners on the basis of current theory. Considerable improvements in goodness of fit can be effected by using this method for capturing diversity, but there is still the challenge of interpreting 'the Frankenstein monsters' the genetic programming method throws up.

**Geographical variation in model fits**
We routinely fit explanatory linear models to geographic phenomena, choosing some universe of interest within which to work. Fotheringham et al (1998) show how simplistic such a standard approach can be in an application which shows that the factors explaining the level of age-specific limiting long-term illness (LLTI) vary dramatically in their predictive power from area to area within a geographical region.

To calibrate a regression model for an area requires selection of a set of neighbouring areas, weighted appropriately by a distance-decay function of influence. The authors propose and implement a method for finding a suitable distance-decay parameter. The results of their analysis of ward level LLTI in northeastern England show wide variation in the influence of unemployment, social class, and population density. They identify the special influence of former employment in coal mining as a missing variable in the analysis. It is always pleasing when a hypothesis tested in a much cruder way (Rees, 1995; Senior, 1998) continues to be robust in more sophisticated analysis. The method deserves wider use in geographic analysis of social phenomena.

**Recovering social composition of small area populations**

A very large amount of information on households and individuals is gathered in the decennial census and released as a set of aggregate tables for small areas. The techniques used to describe the sociogeographic mosaics that are our cities, ranging from social area analysis through factorial ecology to geodemographic classification, fail to capture the essential mixed composition of any area in terms of the types of households that live there. Mitchell et al (1998) show how this failure can be put right through an application of neural network training to the 1% (Household) Samples of Anonymised Records from the UK 1991 Census and a set of synthetic ED (enumeration district) tables constructed from the SAR. When the real ED tables are fed into the classifier an estimate of the composition of the ED population in terms of household types is obtained. Instead of having to designate small areas just as 'struggling', they can be characterised as composed of 90% 'struggling' households together with 10% 'aspiring' households, for example. Perhaps in the next census in 2001 such computational ingenuity can be applied to the classification of the country's 24 or 25 million households and these compositional tables can be published directly.

**Understanding urban dynamics**

Geographers have paid most attention to cross-sectional phenomena, neglecting the study of dynamics. Batty's work has always been an exception and in his paper he shows how contemporary cellular modelling and visualisation tools can be used to implement a simple model of urban development. He demonstrates how a wide variety of complex evolutions can occur as a result of modest shifts in key model parameters. The cellular automata method was developed in the 1950s by the Swedish geographer Hägerstrand and his army of labouring doctoral students. In the late 1990s the computing technology described by Batty does their work in seconds. So far such tools have been used to gain an understanding of the critical events and structural parameters that lead to different evolutionary paths. Applied to realistic urban systems, Batty's model could become a laboratory for developer – planner – resident debate about the degree to which household projections for areas with greenfield and brownfield sites become self-fulfilling prophecies.

This set of papers represents the cutting edge of the application of computational methods to geographic problems. You may not agree with all of the claims made by the authors or understand all the methods described. However, as a set they will, I hope, stimulate and provoke, and perhaps inspire use of one or other methods in your own research.

P Rees, I Turton

**References**


Batty M, 1998, "Urban evolution on the desktop: simulation using extended cellular automata" *Environment and Planning A* **30**(11) forthcoming

Diplock G, 1998, "Building new spatial interaction models using genetic programming and a supercomputer" *Environment and Planning A* **30** 1893 – 1904

Fischer M, 1998, "Computational neural networks: a new paradigm" *Environment and Planning A* **30** 1873 – 1891

Fotheringham S, Brunsdon C, Charlton M, 1998, "Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis" *Environment and Planning A* **30**(11) forthcoming

Mitchell R, Martin D, Foody G, 1998, "Unmixing aggregate data: estimating the social composition of enumeration districts" *Environment and Planning A* **30**(11) forthcoming

Openshaw S, 1998, "Neural network, genetic and fuzzy logic models of spatial interaction" *Environment and Planning A* **30** 1857 – 1872

Rees P, 1995, "Putting the Census on the researcher's desk", in *The Census Users' Handbook* Ed. S Openshaw (GeoInformation International, Cambridge) pp 27 – 81

Senior M, 1998, "Area variations in self-perceived limiting long term illness in Britain, 1991: is the Welsh experience exceptional?" *Regional Studies* **32** 265 – 280
Turton I, Openshaw S, 1998, "High performance computing and geography: developments, issues and case studies" *Environment and Planning A* **30** 1839 – 1856