CENTRE FOR ADVANCED SPATIAL ANALYSIS
Working Paper Series

Paper 18

DEFINING & DELINEATING THE CENTRAL AREAS OF TOWNS FOR STATISTICAL MONITORING USING CONTINUOUS SURFACE REPRESENTATIONS+

Mark Thurstain-Goodwin
*David Unwin

UCL

Centre for Advanced Spatial Analysis
University College London
1-19 Torrington Place
Gower Street
London WC1E 6BT

Tel: +44 (0) 171 391 1782
Fax: +44 (0) 171 813 2843
Email: casa@ucl.ac.uk
http://www.casa.ucl.ac.uk

http://www.casa.ucl.ac.uk/towncentres.pdf

Date: January 2000

ISSN: 1467-1298

ABSTRACT

The increasing availability of very high spatial resolution data using the unit post code as its geo-reference is making possible new kinds of urban analysis and modelling. However, at this resolution the granularity of the data used to represent urban functions makes it difficult to apply traditional analytical and modelling methods. An alternative suggested here is to use kernel density estimation to transform these data from point or area 'objects' into continuous surfaces of spatial densities. The use of this transformation is illustrated by a study in which we attempt to develop a robust, generally applicable methodology for identifying the central areas of UK towns for the purpose of statistical reporting and comparison. Continuous density transformations from unit post code data relating to a series of indicators of town centredness created using ArcView are normalised and then summed to give a composite ‘Index of Town Centredness’. Selection of key contours on these index surfaces enables town centres to be delineated. The work results from a study on behalf of DETR.

1. Introduction: the context

It is widely recognised that the growth of out-of-town retailing has had a deleterious impact on town centres in the UK. This was particularly prevalent during the 1980s and early 1990s when a relatively permissive planning system, combined with high levels of car ownership and low prices for land, prompted the major comparison retailers to follow the supermarkets and DIY outlets to out-of-town locations. Inevitably, those smaller retailers left behind in the town centres found it more difficult to compete, and, fearful of the wholesale decline of the town centres (as was witnessed in the United States), a number of policy instruments designed to mitigate against the worse impacts of the trend were introduced by central government.

The most important and successful of these was Planning Policy Guidance Note 6 (PPG 6), the aim of which was to preclude retail development in off-centre locations (Department of the Environment 1996). While this policy is acknowledged to have been successful, the lack of statistics on town centres meant that it was impossible to quantify its effects in any detail. Indeed, this whole era of out-of-town retail development occurred in an information vacuum - the last wholesale collection of retail statistics was the 1971 Census of Distribution. For such statistics to be useful it was found necessary to develop a consistent and rigorous approach to defining what is meant by the term 'town centre' that would also enable such centres to be delimited on the ground in some useful way for statistical aggregation of relevant data.

In 1996, the Department of Environment, Transport and the Regions (DETR) commissioned the Centre for Advanced Spatial Analysis (CASA) at University College London, and the Urban and Economic Development Group (URBED) to undertake a Feasibility Study to investigate the generation of consistent statistical areas of town centre activity. Using an off-the-shelf Geographic Information System (GIS), and utilising existing government datasets, it was demonstrated that it is possible to generate basis statistics such as employment, retail sales turnover and floorspace on a consistent basis for all town centres in the UK (DETR 1998).

The success of the Feasibility Study led the DETR to commission CASA and URBED to undertake a much wider Pilot Study covering the whole of London, as the first stage in a programme of national implementation. In this paper we first describe how the methodology developed in the Feasibility Study to facilitate the recognition and delineation of town centres using as a key component socio-economic data represented by continuous density surfaces was enhanced. In part 2, we outline some of the methodological problems involved in the idea of any 'town centre' and the data sources available to us. We also briefly report some unsatisfactory attempts to produce a methodology using relatively conventional taxonomic methods. The surface generation methodology adopted and presents results for the London region in Part 3. In Part 4 we conclude with some more general observations and speculations on the utility of these approaches.

2. Mapping town centres

But what exactly *is* a town centre? There are many viewpoints from which a definition of *town centre* could be approached. Although we might imagine that town centres could easily be represented as *objects* in a geographic information environment they are almost archetypal examples of geographic objects with indeterminate boundaries (Burrough et al. 1996). Among the issues that need to be addressed are:

- Although the centre is the part of the town where a multiplicity of different urban activities coalesce and coexist in a distinct place, there are many different types of centre. The key characteristics might, for example, include a retail core, an office centre, an area of high building density, or a concentration of visitor attractions. The concept can also be extended to include town centre residential areas, particularly in view of a trend towards people moving back into them.
- Despite the fact that we all recognise that town centres exist, and often realise when we are standing in one, it is unlikely that we would all agree on the criteria to be adopted in their recognition and mapping. The concept is essentially fuzzy.
- Because of this, the edge of the town centre is extremely difficult to define – it is indeterminate.

So, an initial difficulty relates to the nature of the objects themselves. How is it possible to define statistical aggregations for town centres –essentially a binary construct – if distinct boundaries cannot be defined? Even in town centres whose boundaries appear to be delineated by distinct geographical barriers (such as a ring road, or a river) the traditional statistical geography rarely bears much resemblance to the underlying urban geography. Figure 1a shows how the conventional enumeration district (ED) geography, used hitherto in urban analysis, could not be used to create aggregate statistics for Wolverhampton town centre (which is broadly defined by its ring road) since the three EDs within the town centre extend far beyond its boundary.

We can see, therefore, how any approach based on the classification of these small areas using the decennial census as a base seems doomed to fail. Land use and employment activities vary at scales well below this and until recently, no data have been available with sufficient spatial resolution to enable this issue to be tackled. In order to capture the heterogeneity of urban phenomena, we turned to the geography given by the postcode (Figure 1b).

FIGURE 1a: Wolverhampton Town Centre and ED geography

FIGURE 1a: Wolverhampton Town Centre and postcode geography

Postcodes are a generalisation of urban geography designed for the speedy delivery of the Royal Mail, and represent the spatial average of around 14-17 properties on a postman's delivery. The geography is *not* that of a clearly defined area on the ground and it is usual to represent them as point objects with varying spatial resolution (Raper et al. 1992). As an address based geography, the unit postcode (UPC) is well suited to the mapping of many different data sets and at the resolution required. Retailers and the insurance industry use UPC to tie their databases to maps (using attribute data generate from EPOS (Electronic Point of Sale) and customer surveys, for example). Companies such as the *Business Database* and *Experian* provide large databases to this end. Similarly, databases held by government agencies (such as the Valuation Office Agency (VOA), Office for National Statistics (ONS) and DETR are address-referenced but until this project, these have not been extensively mapped or analysed.

The broad information requirements are for data on town centre floor-space, employment and turnover of retail outlets for town and other shopping centres. The main sources for data geo-referenced at this level came from two government agencies – the VOA and the ONS. The VOA collects floorspace information on individual properties as part of the non-domestic rating valuation. The ONS holds employment and some turnover data on individual business on the Inter-Departmental Business Register. Collated from Inland Revenue sources, this dataset records the number of people employed, defined at 5 digit SIC (Standard Industrial Classification) and the total turnover (again by 5 digit SIC) for every company in the country. Both datasets are highly confidential and in order to help preserve confidentiality they have been aggregated to the UPC level.

Urban analysis is only rarely applied at this spatial scale using these sources of data. Of course, one the key problems is the accuracy of the data (both in terms of geo-referencing and attribution) although gross errors can easily be spotted when the data are contextualised by large-scale mapping in a GIS. This issue has started to be addressed in the research through the development of an on-line Data Verification Tool that enables local experts to evaluate and update the underlying raw data (CASA forthcoming). However, even if the data were 100% accurate, in the context of town centre definition, they represent a major challenge for analysis since data georeferenced at the unit postcode level have an intrinsic *granularity*.

Traditional, area-based classification techniques for example cluster analysis based on principal component scores will almost certainly fail when attempting to identify town centres using data at this scale. There are two related issues. First, for all variables examined the UPC data have extremely high variances, with a pre-dominance of zero values and hence extremely non-normal frequency distributions. This makes any classification that relies on the usual taxonomic indices of similarity (as in cluster analysis) very hard to apply. Second, it was also found that there is very little correlation amongst these variables which means that extraction of the principal components (PCA) gives very unsatisfactory results. An initial trial using employment and floor-space data for 901 UPC in the Andover area showed that, although the employment and the floor-space data correlate quite well, within either set the correlations are weak to non-existent and a PCA on the employment data produced no meaningful components. In fact, there were 10 components each of which 'explained' from 19.6 to 1.1 % of the variance. No single employment or floorspace variable appears to dominate or discriminate between the town centre and other parts of the urban area. Moreover, even within what can be regarded as a reasonable approximation to a town centre, a large proportion of the post codes do not have any employment of a type that might be held to be 'representative'. All that could be concluded was that town centre UPCs tend to have employment in the categories 'Civic and Public Administration', 'Commercial Office', 'Entertainment and Leisure', 'Hotel' and 'Retail' as well as higher than average total employment.

Finally, at this spatial scale, individual UPCs do not relate to well-defined spatial areas. All we have in readily accessible machine-readable format is a point reference, which we need to convert to define contiguous areas of the earth's surface. An option that was considered is to use a standard tessellation available in many GIS (Thiessen Polygons) based on these point references and then classify the available data using a contiguity constraint and the battery of available indicators. However, it is clear from the analysis presented above that this would prove technically very difficult and would not readily generalise to give a uniform methodology applicable to all towns.

3. Towards a surface based analysis

The availability of high spatial resolution socio-spatial data, such as the Martin/Bracken 200m population raster for GB (Bracken and Martin 1989) has led to a revival of interest in continuous representations expressed in terms of *potentials* and *spatial densities*. A number of recent papers have explored such transformations for visualisation (Langford and Unwin 1994, Dykes et al. 1998) and analysis (Donnay 1995, Turton 1999, Coombes and Raybould forthcoming). Although presented as a recent development these representations have much in common with the earlier tradition in geography referred to as *social physics* (Stewart 1947). Surface representations have also been used to explore other data sets, most notably to generate contour maps for property values (Knos 1968, Wyatt 1999).

The methodology we have adopted uses kernel density estimation (KDE) to create continuous surface representations of the four key factors that can be held to characterise town centres, namely: economy (which integrates the various town centre employment type ones might expect to find); property (the density of the buildings); diversity of use, and visitor attractions. The geographic variation of these factors is represented using a 20m grid of spatial densities. First, surfaces are generated for each of these four factors. These are then combined into a composite *Intensity of Town Centredness* (ITC) surface. Analysis of peaks on this surface provides a means of delimiting the geographical extent of town centres for which a suite of statistics can be generated. The surface generation process involves four distinct stages.

Step One: Defining the Characteristics

It is unlikely that if you where to ask a sample of ten people what was the key characteristic of a town centre you would find much agreement. Some might argue that retailing was the single most important function, while others would point out that it is the diversity of activities concentrated in a single place which define the town centre. During the Feasibility Study, which generated ITC surfaces for 12 different town centres around the country. Seven core characteristics were defined –

an Activities and Facilities component; a measure of the of the Diversity of Use; and Intensity of Use component (which described building densities); Pedestrian Gateways (which created a series of pedestrian catchment areas around pedestrian access points in the town centre); the relative lack of Residential Population; Visitor Attractions; and a composite measure of Commercial Office, Retail and Leisure Turnover.

During the London Pilot Study, this seven component model was re-appraised and trimmed to give the model shown in Figure 2. Surfaces were created for the various sub-components indicated at the top of the diagram, and these were then integrated into the four main components: Property, Economy, Diversity and Visitor Attractions. Note that three of the original components were dropped from configuration (CASA forthcoming).

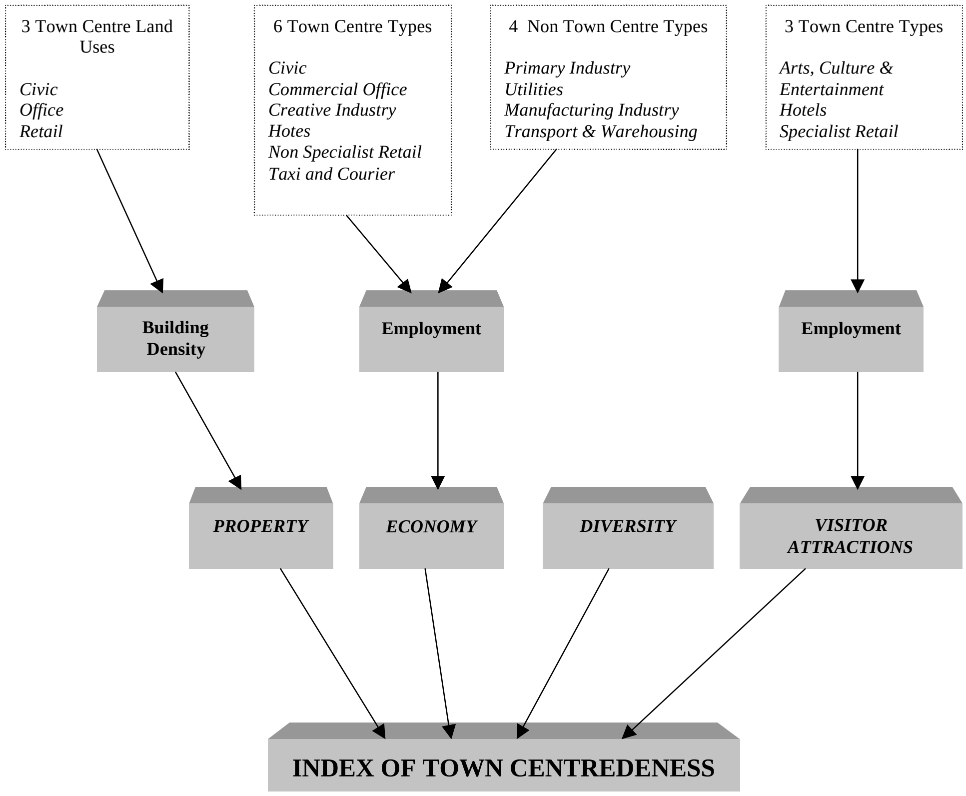

Figure 2: The characteristics of a town centre

Step Two: Two Dimensional KDE

Initially, each of the indicators shown as the top level of Figure 1 and geo-referenced by UPC, was mapped and transformed using a 300m bandwidth quadratic kernel density estimator (Silverman 1986, Gatrell et al. 1996). The method is known as kernel density estimation (KDE) because around each point at which the indicator is observed a circular area (the kernel) of defined bandwidth is created. This then has the value of the indicator at that point spread into it according to some appropriate function.

Summing all these values at all places, including those at which no incidences of the indicator variable were recorded, gives a surface of density estimates. Note that, in order for the method to be volume preserving, it is necessary that the spread function used integrates to unity, for example a normal distribution function. Almost all authors agree (Bailey and Gatrell 1995, Atkinson and Unwin 1998) that the choice of function is less critical than that of bandwidth, and it is an easy matter to show that the resulting surfaces become smoother as the bandwidth is increased. Experimentally, a series of tests using different bandwidths showed that a value of 300m is appropriate for this application, and that in this case the final results are relatively insensitive to most reasonable choices of bandwidth.

Each indicator is thus represented by a continuous spatial density surface (in numbers per square km, dimension L^{-2}). Note that, although we start with discontinuous point located data with a large proportion of zero entries, each surface must assign a local, non-zero density to every point in the plane. These individual surfaces can be visualised using conventional surface display techniques.

Step Three: Normalisation and Aggregation

The next step in the method took these individual density values and normalised them using a z-score normalisation. These normalised scores were then summed to provide the single index of town centredness (ITC). This operation is known in the geographical information science literature as map overlay and the method employed here is an *index overlay* (Bonham-Carter 1994). This can be formalised as the evaluation for every point on the plane of:

$$ITC = \Sigma_i a_i s_i$$

in which ITC is the index of town centredness, s_i is the transformed density score for a given indicator, and a_i is a weight for that indicator. The method assumes that $a_i = 1$ for all but a few indicators. A simple, easy to use interface was written in ArcView 3.1 that enables the weightings to be varied at will, including the masking out of indicators by assigning zero weights. We have experimented using this interface with the general result that the overall methodology seems very robust to changes in the inputs and the precise details of the procedure adopted. The ability to vary the weights is a useful feature, enabling the method to be tuned to reflect differing mixes of town centre activity.

Step 4: Results and Visualisations

Within the London region (defined as those areas lying within the M25) the seventeen base indicators covering over 0.5 million UPCs were subjected to KDE and index overlay in the manner outlined above. Figure 3 shows the resulting ITC surface visualised at this regional scale.

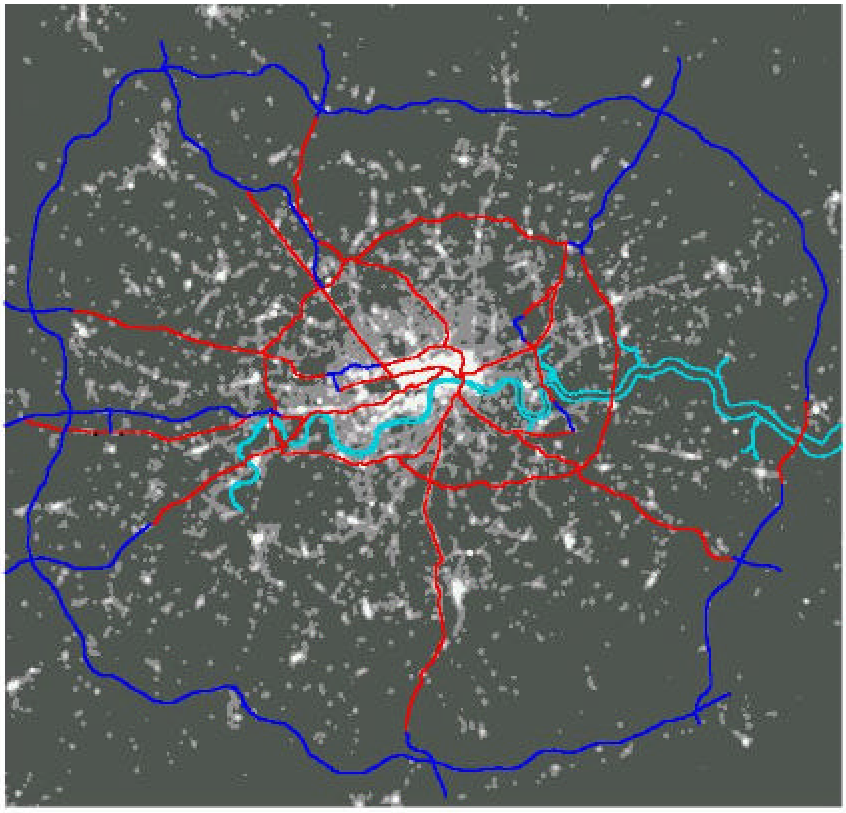

FIGURE 3: Intensity of Town Centredness within the M25

In this map, the lighter zones are those more likely to be areas of town centre activity. (recall that KDE is in fact a two-dimensional probability density indicator). Thus, the essential fuzziness of the town centre is embedded within the KDE process. The centre of London is clearly evident, but in addition the urban villages and the main strip centres along the major radial routes can be easily identified. For reasons that are not entirely clear to us, this visualisation is very similar to that produced by (Dykes et al. 1998) who used the local curvature of a population density surface for London. As far as we can see, at this scale the ITC surface identifies all of the major urban centres that make up the conurbation.

When we look at this surface at higher resolution then interesting detail becomes evident. Figure 4a shows the ITC surface for a portion of north London. At this scale of examination, the granularity of the urban landscape starts to become apparent, as do artefacts from the spatial distribution of UPCs. For example, the major parks and heath areas are, as expected, clearly visible. Interestingly however, University College London also shows no development or economic activity! This is, of course, because the College is represented by a single postcode which was not integrated in to the model.

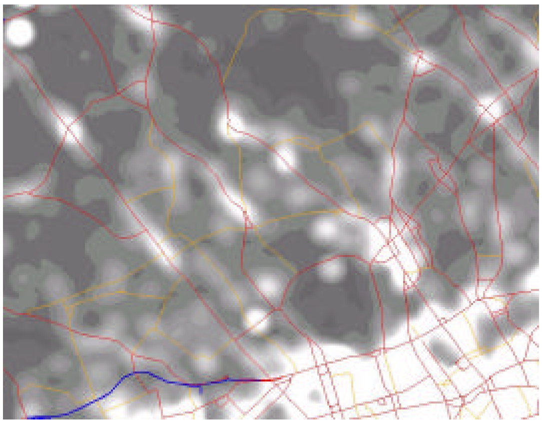

FIGURE 4A: ITC surface in North London

FIGURE 4B: Hampstead town centre

At an even smaller scale, we can begin to see the actual form of individual town centres. Figure 4b shows the ITC surface around Hampstead. At this spatial scale, the innate uncertainty of what we mean by the term *town centre* becomes evident.

4. Some Pitfalls and a Conclusion

The major technical objection to area based methods of urban analysis is essentially related to the uncertainty introduced by the well known, if often badly understood, modifiable areal unit problem (Openshaw 1987). From the viewpoint of a physical scientist, it can be argued that any arbitrary areal units, such as the ED of the census, must act as variable selective filters whose effects on how we 'sample' the underlying real geography is almost impossible to determine.

In developing these surface representations of the urban environment we are equally guilty of introducing uncertainty at almost every step in the procedure. Some of this uncertainty, associated with the selection of the indicator variables, their normalisation and weighting, is well known and relatively well understood. By choosing to model the geography using spatial density surface representations we are introducing further sources of uncertainty in respect of the geography of the UPC, and the technical details of the density estimator used. The issue that arises is thus whether or not these new types of uncertainty are any better or worse than that those of the more traditional area based taxonomies. It seems to us that surface representations of the type we have illustrated have two major advantages in urban analysis. First, they enable us to use very high spatial resolution data such as the UPC, or, for that matter, satellite derived land use (see Donnay 1995 for a similar argument) in which the granularity of the urban fabric becomes a major problem.

Second, they seem to allow the development and calibration of models of urban dynamics in which space is represented continuously and hence in a potentially richer way than the discrete zone-based method that are usually used. One of the key problems that beset the urban modeller in the 1960s and 1970s was that the models were simplistic - often constrained and limited by the available data and the technology. Furthermore, their representation of urban systems rarely corresponded to the wealth of implicit knowledge held by geographers, planners and other urban decision makers (Thurstain-Goodwin and Batty 1998). It was no surprise then that urban computer models were rejected as geographers in particular looked to new techniques to understand towns and cities. Perhaps the techniques we have outlined above can be used to start to map and to visualise aspects of the urban system hitherto unseen, in particular its innate fuzziness and granularity. And perhaps too these techniques will contribute in some small way to the rehabilitation of computer technology in contemporary human geography.

References

Atkinson, P., and Unwin, D. (1998). “Comparisons and problems in applying density estimation techniques to the distribution of Hepatitis A.” *Geographical Systems*, 3, 301-312.

Bailey, T. C., and Gatrell, A. C. (1995). *Interactive spatial data analysis*, Longman Scientific & Technical ;J. Wiley, Harlow Essex, England, New York, NY.

Bonham-Carter, G. (1994). *Geographic information systems for geoscientists : modelling with GIS*, Pergamon, Oxford ; New York.

Bracken, I., and Martin, D. (1989). “The generation of spatial population distributions from census centroid data.” *Environment and Planning A*, 21, 537-543.

Burrough, P. A., Frank, A. U., and European Science Foundation. (1996). *Geographic objects with indeterminate boundaries*, Taylor & Francis, London ; Bristol, PA.

CASA. (forthcoming). *Town Centres: Defining Boundaries for Statistical Monitoring - Pilot Study*, DETR, London.

Coombes, M. G., and Raybould, S. (forthcoming). “Public policy and population distribution: developing appropriate indicators of settlement patterns.” *Environment and Planning C : Government and Policy*.

Department of the Environment. (1996). *Revised PPG6: Town Centres and Retail Developments*, HMSO, London.

DETR. (1998). *Town Centres: Defining Boundaries for Statistical Monitoring: Feasibility Study*, TSO, London.

Donnay, J.-P. (1995). “Délimitation de l'hinterland des agglomérations urbaines au départ d'une image de télédétection.” *Revue belge de géographie*, 59, 325-331.

Dykes, J. A., Fisher, P. F., Stynes, K., Unwin, D., and Wood, J. (1998). “The use of the landscape metaphor in understanding population data.” *Environment and Planning B, Planning and Design*, 26, 281-295.

Gatrell, A., Bailey, T., Diggle, P., and Rowlingson, B. (1996). “Spatial Point Pattern Analysis and its Application in Geographical Epidemiology.” *Trans Inst Br Geogr*, 21, 256-274.

Knos, D. S. (1968). “The Distribution of Land Values in Topeka, Kansas.” Spatial analysis; a reader in statistical geography, B. J. L. Berry and D. F. Marble, eds., Prentice-Hall, Englewood Cliffs, N.J.,, 269-289.

Langford, M., and Unwin, D. (1994). “Generating and mapping population density surfaces within a geographical information system.” *Cartographic Journal*, 31, 21-26.

Openshaw, S. (1987). *The Modifiable Areal Unit Problem : Concepts and Techniques in Modern Geography 38*, Geo Books, Norwich.

Raper, J., Rhind, D., and Shepherd, J. (1992). *Postcodes : the new geography*, Longman Scientific and Technical :J. Wiley, Essex, England; New York, NY.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, Chapman and Hall, London ; New York.

Stewart, J. Q. (1947). “Empirical mathematical rules concerning the distribution and equilibrium of population.” *Geographical Review*, 37(461-485).

Thurstain-Goodwin, M., and Batty, S. (1998). “GIS and Town Centres: Exploratory Environments Involving Experts and Users.” *Built Environment*, 24(1), 43-56.

Turton, I. (1999). “Application of Pattern Recognition to Concept Discovery in Geography.” High Performance Computing, R. Allan, M. Guest, A. Simpson, D. Henty, and D. Nicole, eds., Kluwer Academic, London, 467-478.

Wyatt, P. (1999). “Geographical Analysis in Property Valuation.” , RICS, London.