The spatially discontinuous choropleth map is a poor representation of the underlying continuous distribution of population density. A possible alternative is to derive dasymetric maps at a fine spatial resolution by making use of satellite imagery in a geographical information system. However, there are cartographic problems when these maps are displayed and further processing is needed in order to obtain approximations to a continuous density surface. Isarithmic maps of these density surfaces retain a high degree of spatial accuracy while providing pleasing and highly adaptable presentations.

The methods used to generate dasymetric and isarithmic maps are readily implemented in most raster based geographical information systems. For example, the classification of remotely sensed imagery, the subsequent processing and integration of data, and most of the cartographic display, were all undertaken in this work using the low cost IDRISI GIS that operates on standard IBM PC compatible hardware..

Generating and mapping population density surfaces within a geographical information system

M. Langford[1] *and D. J. Unwin*[2]

[1] Midlands Regional Research Laboratory, Department of Geography, University of Leicester, Leicester LE1 7RH
[2] South East Regional Research Laboratory, Department of Geography, Birkbeck College London, London W1P 1PA

INTRODUCTION

"The Universe – some information to help you live in it

⋮

4. Population: None
It is known that there are [sic] an infinite number of worlds, simply because there is an infinite amount of space for them to be in. However, not every one of them is inhabited. Therefore, there must be a finite number of inhabited worlds. Any finite number divided by infinity is as near to nothing as makes no odds, so the average population of all the planets in the Universe can be said to be zero. From this it follows that the population of the whole Universe is also zero, and that any people you meet from time to time are merely products of a deranged imagination." (Douglas Adams, 1980, *The Restaurant at the End of the Universe*, London, Pan Books, pp. 113–114.)

The logic in this quotation from comic science fiction is appealing and highlights a major source of difficulty in mapping the density of population. This is simply the fact that any measurement of population density is inherently linked to the area over which it is calculated. In our experience most presentations of population density data seem to pay scant attention to the effects of these areal units and yet, depending on the boundaries of space chosen, it is possible to obtain virtually any density measure. Densities recorded from a massively large area, such as in Adam's Universe, will tend towards zero. At the other end of the scale individual people each occupy a small area of the earth's surface, and generate very high population densities if recorded at such local levels.

Useful measures of population density are obtained from areal units of less extreme scale but the basic problem still remains. For example, Dixon (1972) cites evidence that when all non-residential land uses are included in the area measure (for example, parks, roads, and gardens) residential densities in parts of UK cities range from 1,000 km^{-2} to 6,000 km^{-2}. This compares to more realistic densities (where these land covers are excluded from the calculation) which in parts of Glasgow are suspected to reach 50,000 km^{-2}, and in the world's most congested cities may rise to in excess of 100,000 km^{-2}.

Cartographic techniques for mapping volumetric data that have been collected over areas fall traditionally into three approaches: the choropleth and dasymetric map, both of which utilise area symbolism, and the isarithm map based on line symbolism. Conventionally, population density is represented using the choropleth map, in which the densities are calculated from a series of arbitrary areas (normally administrative or enumeration zones) and displayed using various shading schemes. The dominance of choropleth representation has undoubtedly been strengthened by the widespread adoption of computer-assisted cartography, since this is the only technique currently implemented in many mapping packages.

In this paper we suggest an alternative to the conventional choropleth representation. First, we explain how a choropleth map can grossly distort and mask the underlying reality of the population density distribution. Second, using satellite remotely sensed imagery to identify the settled area, we produce a dasymetric representation. Although this provides a better representation of the real distribution, it is not a totally adequate solution. Finally, we present an alternative representation (and some derivatives) generated within a geographical information system (GIS), that more closely models the population density surface.

The various maps are produced for a study area comprised of 49 census wards that make up the three local government administrative districts of Oadby and Wigston, Leicester, and Charnwood, all in northern Leicestershire, UK. As shown in the sketch map of *Figure 1*, this region has two major cities, Leicester in the south and the smaller one of Loughborough in the north-west. In addition, there are a number of villages most of which are located along the valley and beside the railway line connecting the two cities.

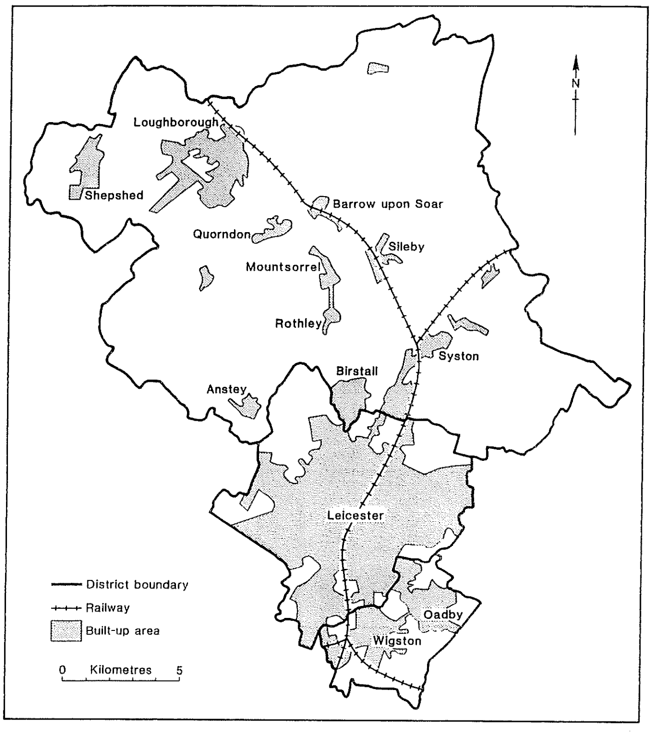

Figure 1. The study area of northern Leicestershire.

THE CONVENTIONAL CHOROPLETH MAP

Figure 2 shows a simple five-class choropleth map of the density of population over the wards of the study area as inferred from the 1981 Census of Population. This was calculated in the conventional way outlined in, for example, Unwin (1981). The population of each ward was divided by its area and then, for the purposes of generalisation

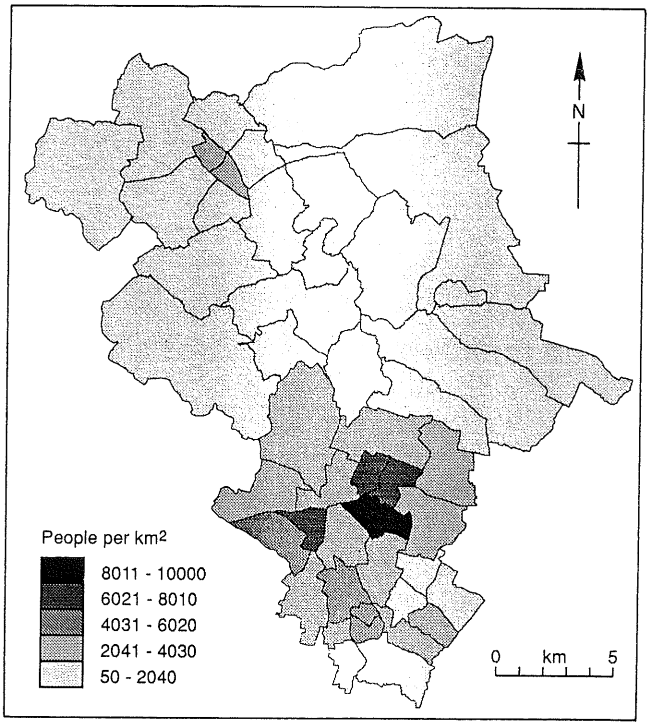

Figure 2. Choropleth population density map (1981 Census).

(Dobson, 1973), these data were allocated to a specific class using a quantile scheme (Evans, 1977), and finally each ward area was shaded appropriately. The population densities calculated in this manner suggest a range of 50 km^{-2} to 10,000 km^{-2}.

Unfortunately, there are at least three very significant problems with this form of representation. Although Baxter (1976) has summarised these as being related to 'computer drawn maps' they are in fact common to any map using areal densities. First, there is a general tendency for larger areal units to provide lower measures of population density (Massey and Stephan, 1977), and conversely, as the base area is reduced, both the peak values and the variance of population density estimates tend to increase. Second, they act as low-pass spatial filters that suppress any high frequency fluctuations occurring within them leading to a generalisation effect. As areas become larger the level of generalisation increases and, since the areal units are normally irregularly shaped, the effect is directionally variable and very difficult to quantify. Finally, the actual location of the boundaries is arbitrary, and in most cases will bear no logical relationship to the location of genuine discontinuities in the property being mapped. If the location of the boundaries were to be changed it is highly likely that a very different looking map would be produced. This is the well-known modifiable areal unit problem that pervades many aspects of the statistical analysis of zone data (see Openshaw, 1984).

Bearing these problems in mind the impression of population density distribution generated by *Figure 2* must be treated very cautiously. For example, although population density appears to increase according to expectation as we progress towards the centre of Leicester, this might be as much due to the fact that the administrative wards get smaller as it is to underlying reality. Note that densities close to the centre of Loughborough appear to be lower than those found in Leicester, whereas in reality they are likely to be very similar since there are no fundamental differences in the housing characteristics of these two areas. Furthermore, the suggestion that population density changes instantaneously at the ward boundaries whilst remaining uniform within the zone is clearly erroneous. Overall, it is very difficult to assess the extent to which the perceived distribution is dependent upon the shape, size and location of the areal units.

A further problem with the conventional choropleth representation is the dependence of its appearance on the procedure used to classify the data (see Evans, 1977). In particular both the method of classification and the number of classes that are used can substantially alter its final appearance. When density data are allocated to a set of classes some level of 'error' is inevitably introduced, although several schemes have been devised to minimise this (Jenks, 1977; Unwin, 1981; Fisher, 1990). In respect to the number of classes that are employed, a general rule is that seven should be regarded as the upper limit (see example, Gilmartin and Shelton, 1989).

The so-called 'continuous choropleth' maps proposed by Tobler (1973), where the spacing of lines or dots is proportional to the values being mapped, is a possible solution to the class allocation problem. The initial attempts to implement the idea were cartographically rather crude and it has been suggested that these maps fail to generalise the patterns sufficiently for the map user to distinguish them (Dobson, 1973; Muller, 1979). However, the use of finely graduated grey scales in facsimile equipment (Muller and Honsaker, 1978) or modern graphics terminals, together with improved software, have ensured that continuous choropleth representations are becoming more widely used.

Although the effects of modifiable boundary location and class interval selection have traditionally been considered the most serious in choropleth mapping, we believe that simple variability in the size and shape of areal units may be of equal importance. In making comparisons of densities from one areal unit to another map users are not comparing like with like, since the measures have been recorded at different levels of generalisation. Unless the density is very uniform throughout the study area any comparison of results must be considered unreliable. A partial solution to this problem is to use a standardised areal unit such as the grid squares suggested by Browne and Millington (1983), but unfortunately data are only seldom available in this form. The 1971 UK Census of Population was available as aggregations at the kilometre square level, a fact which made the automated mapping of the entire country relatively straightforward (CRU *et al.*, 1980).

THE DASYMETRIC ALTERNATIVE

An alternative representation which has been available for many years (Wright, 1936), although not widely used, is the dasymetric map. This approach makes use of additional geographical information about the distribution of population, for example, the known distribution of housing. Each density measure is then calculated by assuming that the areal unit's population is concentrated in the areas identified from this additional knowledge, and the result is mapped by shading only these 'populated' areas. In effect, the representation of spatial discontinuities in the data is divorced from the boundaries of the collection units. Provided that the additional information is of a suitable nature, the discontinuities shown on the dasymetric map are more likely to be realistic guides to the genuine underlying distribution of the property that is being mapped. At first sight this is an attractive option since, with the rapid increase in the use of GIS for storing, manipulating and displaying spatially referenced data sets, it seems almost inevitable that appropriate sources of additional information will become available.

In this study use is made of modern remotely sensed data that are readily integrated into a GIS based on the raster data structure, and this enables the populated areas (as inferred from residential housing) of each census ward to be mapped with considerable precision. *Figure 3* shows the areal distribution of housing within the study area, obtained from a 1984 Landsat Thematic Mapper multispectral image classified in the manner described by Langford *et al.* (1991). For the present study, the various land covers recognised in the original classification have been collapsed within the GIS into a simple binary division into 'occupied' and 'unoccupied' classes. Occupied pixels were those classified as some form of residential housing, while the unoccupied pixels equate to all other land uses.

Each pixel of the image represents a 30 m × 30 m area on the ground. This resolution is probably sufficient to define the populated areas in at least as much detail as could be obtained from a 1:50,000 digital topographic database. Selecting populated areas by this method may also be less arbitrary than if they had been derived from traditional map data since, although there are undoubtedly errors in the classification, the selection has been undertaken using consistent criteria throughout the region.

A global classification accuracy of between 85% and 95% is typical for imagery of this type, although it should also be recognised that there may be considerable variation in the spatial distribution of the error. Another potential problem with this data source that is that whilst it may be possible to identify 'housing' pixels, we cannot be sure that these pixels are truly occupied. Furthermore, high-rise and

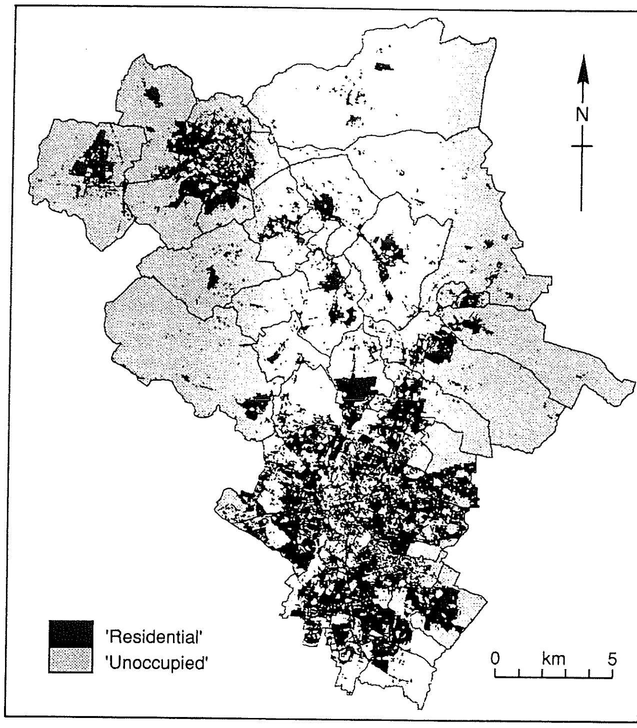

Figure 3. Distribution of residential housing (from classified Landsat TM imagery).

low-rise buildings are not easily differentiated. However, providing that the housing type is relatively uniform within the study region and that unoccupancy remains a low proportion of the total, these should not present any major problems.

In creating the dasymetric representation, it is assumed that the entire population for each census ward is concentrated within the area designated as occupied on the classified image. This results in estimated population densities with a range of 2,820 km^{-2} to 14,900 km^{-2}, clearly wider than that obtained by the earlier choropleth mapping.

A dasymetric map based upon the distribution shown in *Figure 3* can be generated by shading the occupied pixels in accordance with their estimated population density values. This is a feasible proposition when displaying the data on a video graphics terminal where grey levels can be finely controlled, but it becomes less so for paper based output where the halftone screens necessary to simulate grey levels often interfere with the complex pixel patterns. However, this situation is clearly dependant on current technology and, as improvements are made to electrostatic plotters and other high resolution devices, it may prove to be ephemeral.

Even when displayed on a graphics terminal the map is a poor cartographic product. This is principally because the reader can see individual pixels and too much fine spatial detail. Furthermore, there are likely to be additional problems caused by simultaneous contrast when pixels of one shade are placed next to others with very different shades. The problem of excessive detail is always a danger in dasymetric mapping since there is no obvious scale at which it should be applied. In this sense every dasymetric representation is arbitrary. The solution to many of these difficulties is to generalise the data for cartographic display, but an alternative is to reconsider population density in terms of a continuous surface.

MODELLING POPULATION DENSITY AS A SURFACE

A sensible alternative to the dasymetric map lies in the recognition that population density is a continuous function even though population itself is discrete. Whilst Dixon (1972) argues that population density 'cannot be observed at a point', this is only partially correct. It is true that, without referring to an area over which it is evaluated, population density at a point is incapable of definition. However, at any point in the area of interest it is possible to count the number of people falling within a given search radius. Division of this count by the area searched will yield a population density for that point, and such estimates can be obtained for any point. It is therefore possible to construct a continuous surface of population density (see Unwin, 1981; Martin, 1989) and it is suggested that this is best represented cartographically in a way which recognises its inherent continuity.

Obviously, data on the distribution of individuals are not available using conventional census data. Although population counts are released for enumeration districts, a finer spatial zonation than wards, the only spatial reference readily available for use in a GIS is the estimated zone centroid. However, using the satellite remotely sensed imagery, it is possible to infer the distribution of the population at a finer spatial resolution still. Whilst it has been shown that the dasymetric data, generated in the fashion outlined above, have little value in direct cartographic display, this does not exclude the fact that they have significant analytical value. For instance, these data have a direct interpretation as population density estimates over 30 m × 30 m parcels of land, and this provides us with a means of modelling population density as a continuous surface.

As a first step consider the problem of the irregular shaped and variably sized areal units. Clearly, the population can be re-mapped onto a regular lattice, say 1 km^2 grid squares, simply by aggregating the dasymetric populations over the desired units on a pixel-by-pixel basis. Thus, should a new areal unit span across a ward boundary, the estimated population will be a summation that takes into account both the population densities within each contributing ward, and the proportions of occupied and unoccupied land cover within the new areal unit. *Figure 4* is a pseudo-3-D representation of data that have been calculated in this manner to produce an estimate of the 1981 population as recorded on a regular 1 km^2 grid lattice. It is worth noting that this method of cross-area estimation has the desirable pycnophylactic property (Tobler, 1979) which ensures that total volume, in this case population, is maintained; people are not destroyed or manufactured during the re-distribution process.

A close approximation to the continuous surface of population density is achieved simply by using the well known 'floating grid' or 'floating window' procedure. This technique has long been recognised as an attractive method for estimating areal density distributions (Schmid and MacCannell, 1955; Porter, 1957). The same calculations used to generate *Figure 4* are performed, but the aggregate population is divided by the area of the kernel to derive a new density measure and this is assigned to the central pixel. The process is then repeated, with the location of the kernel shifted by a single pixel each time, until the entire area has been covered. This results in a discrete approximation to a continuous population density surface. Each pixel on this surface holds the magnitude of the population density at that point, calculated using a surrounding areal unit of consistent size and shape. It is a simple matter to alter both the shape of the search unit (circular is most appropriate although square is computationally efficient), its size (thereby adjusting the level of generalisation of the estimated surface) and the weightings used in the density estimation.

Although fairly novel cartographically, this form of density estimation is often used in statistics to estimate underlying probability (rather than population) density

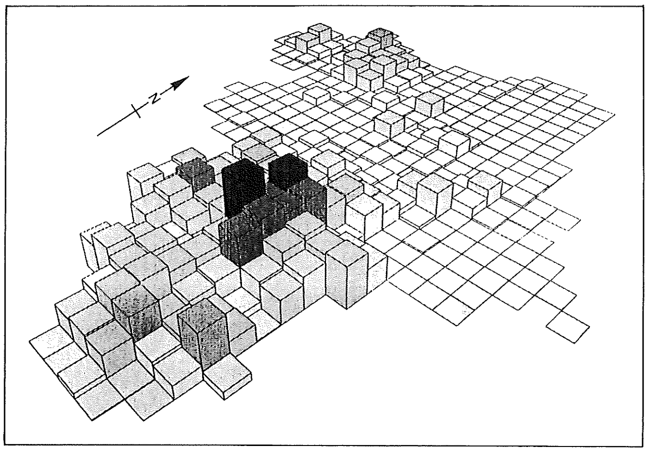

Figure 4. Population density on a 1 km grid.

functions (Silverman, 1986). Two dimensional density estimation of this type has been used to map the distributions of point located data such as victims of specified diseases such as leukaemia (Bithell, 1990), and the incidence of crime (Brunsdon, 1991). An accessible introduction with examples related to population distribution is provided by Gatrell (1994).

This technique is simple to implement using the neighbourhood or spatial filtering functions that are found in most raster based GIS. The results are illustrated in *Figures 5a* and *5b*, which show the population density surface recorded over a 0.5 km search radius. In *Figure 5a* the data are represented through the use of colour shaded contours or isopleths (reduced to grey tones for publication), while the same data are shown in *Figure 5b* as a perspective plot which helps to reinforce the concept of a density surface. *Figures 5c* and *5d* are the equivalent plots where the density data have been evaluated over a 1.0 km search unit. The effect of search radius (or areal unit size) on the level of generalisation is very clear, with *Figures 5c* and *5d* being much 'smoother' surfaces than those seen in *Figures 5a* and *5b*.

A large attraction of modelling population density in this fashion is that the opportunities for displaying the data are wide and varied. A standard isopleth map may be produced using any of the computer contouring routines that are widely available, or alternatively a simple continuous grey tone image might be used. Many of these representations are considerably enhanced by the use of colour when it is available. There are in the literature numerous examples where 'surface shadowing' effects generated by an imaginary light source are used to visualise surface models. It is also possible to compute the vector properties of gradient magnitude and direction from the surface and display these in some suitable manner (for example, Lavin and Cerveny, 1987).

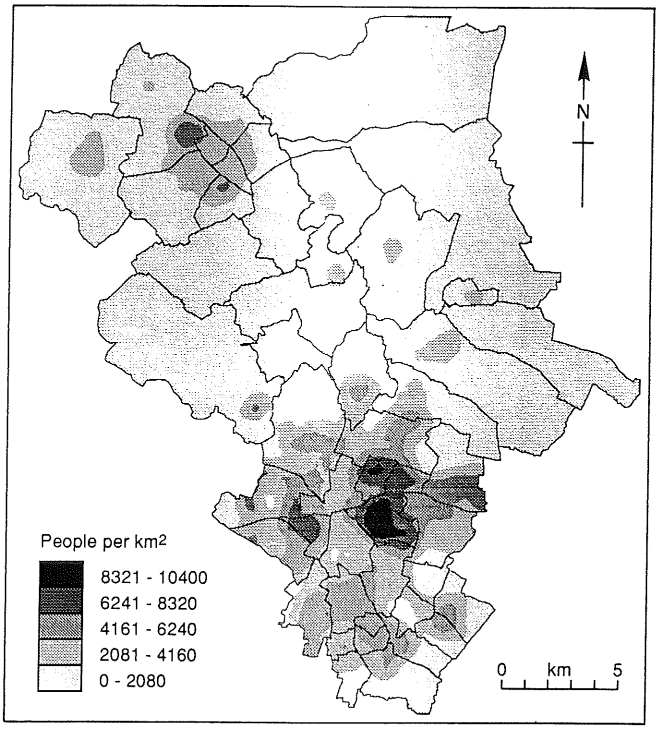

Figure 5a. Population density surface (0.5 km search radius).

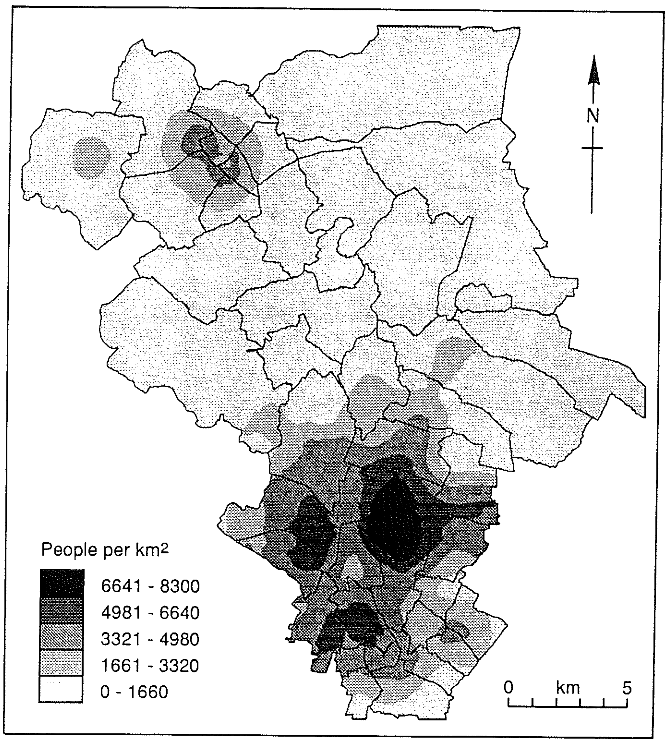

Figure 5c. Population density surface (1 km search radius).

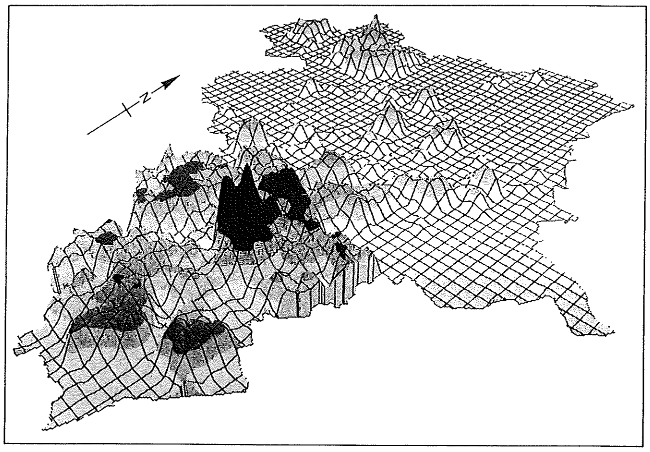

Figure 5b. Pseudo-3-D population density surface (0.5 km search radius).

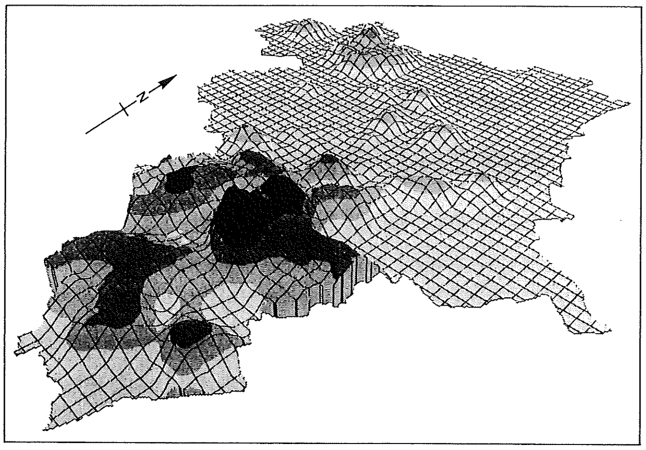

Figure 5d. Pseudo-3-D population density surface (1 km search radius).

CONCLUSIONS

Where the purpose of a population density map is to convey an accurate impression of density distribution the conventional choropleth map representation is a poor choice. It encompasses many problems including data reported at variable levels of generalisation, arbitrary zonation and the false spatial continuities/discontinuities that this generates, and difficulties in the selection and representation of classes.

A dasymetric map allows some of these difficulties to be overcome. In particular, the spatial discontinuities in the mapped data are largely divorced from the boundaries of the collection zones and thus it should be a better indication of the true geography. However, dasymetric mapping is arbitrary in the sense that there is no obvious scale at which it should be applied. It has been shown that if the spatial resolution is too fine the cartographic product becomes unsatisfactory. This is not an indication that the derived data are themselves a poor product, merely that the method of cartographic presentation is at fault.

Population density is a continuous function and in order to present an effective and accurate impression of its distribution we need a scheme that recognises this continuity. The fine spatial resolution of dasymetric data derived within a geographical information system that combines classified remotely sensed imagery with census ward statistics, can be exploited to provide a close approximation of the continuous population density surface. This is achieved by simply summing the dasymetric population distribution within a chosen search radius on a pixel-by-pixel basis, and allowing this window to 'float' across the study area, thus repeating the estimate at fine spatial intervals.

This approach has many attractive features. First, the density measures are recorded at a consistent level of generalisation by an areal unit which is both well defined and easily modifiable. Second, the product is a close approximation to a continuous density surface and provides the reader with an accurate impression of population distribution. Third, there are numerous options for displaying these data, such as isopleths, continuous grey tones, colour 'density slices', and pseudo-3-D visualisations among a wealth of others. Finally, the method is easily implemented in an 'integrated' raster based GIS, and the results lend themselves well to modern analytical operations and raster display technology.

REFERENCES

Baxter, R. S., 1976. Some methodological issues in computer drawn maps, *The Cartographic Journal*, 13: 145–155.

Bithell, J. F., 1990. An application of density estimation to geographical epidemiology, *Statistics in Medicine*, 9: 691–701.

Browne, J. J. and Millington, A. C., 1983. An evaluation of the use of grid squares in computerised choropleth maps, *The Cartographic Journal*, 20: 71–75.

Brunsdon, C., 1991. Estimating probability surfaces in GIS: an adaptive technique. In: Herts, J., Ottens, H. F. L., and Scholten, H. J. (eds.), *Proceedings Second European Conference on Geographical Information Systems*, EGIS Foundation, Utrecht, Netherlands, 155–164.

CRU/OPCS/GRO(S), 1980. *People in Britain – A Census Atlas*.

Dixon, O. M., 1972. Method and progress in choropleth mapping of population density, *The Cartographic Journal*, 9: 19–29.

Dobson, M. W., 1973. Choropleth maps without class intervals: A comment, *Geographical Analysis*, 5: 358–360.

Evans, I. S., 1977. The selection of class intervals, *Transactions, Institute of British Geographers*, (2): 98–124.

Fisher, P., 1990. A Turbo Pascal unit for optimal partitions, *Computers and Geosciences*, 16 (5): 717–732.

Gatrell, A., 1994. Density estimation and the visualisation of point patterns. In: Hearnshaw, H. M. and Unwin, D. J. (eds.), *Visualization in Geographical Information Systems*, John Wiley and Sons, Chichester, 65–75.

Gilmartin, P., 1988. The design of choropleth shading for maps on 2- and 4-bit color graphics monitors, *Cartographica*, 25 (4): 1–10.

Gilmartin, P. and Shelton, E., 1989. Choropleth maps on high resolution CRTs: The effects of number of classes and hue on communication, *Cartographica*, 26 (2): 40–52.

Jenks, G. F., 1977. *Optimal data classification for choropleth maps*, Occasional Paper No. 2, Department of Geography, University of Kansas. 25pp.

Langford, M., Maguire, D. J. and Unwin, D. J., 1991. The areal interpolation problem: estimating population using remote sensing in a GIS framework, In: Masser, I. and Blakemore, M. (eds.), *Handling Geographical Information: Methodology and Potential Applications*, Longman, London, 55–77.

Lavin, S. J. and Cerveny, R. S., 1987. Unit-vector density mapping, *The Cartographic Journal*, 24: 131–141.

Massey, D. S. and Stephan, G. E., 1977. The size-density hypothesis in Great Britain: analysis of a deviant case; *Demography*, 14 (3): 351–361.

Martin, D., 1989. Mapping population data from zone centroid locations, *Transactions, Institute of British Geographers*, 14: 90–97.

Muller, J. C. and Honsaker, J. L., 1978. Choropleth map production by facsimile, *The Cartographic Journal*, 15: 14–19.

Muller, J. C., 1979. Perceptions of continuously shaded maps, *Annals of the Association of American Geographers*, 69: 240–249.

Openshaw, S., 1984. The modifiable area unit problem. Norwich: Geo Abstracts, CATMOG 38.

Porter, P. W., 1957. Putting the isopleth in its place, *The Minnesota Academy of Science Proceedings*, XXV–XXVI, 372–379.

Schmid, C. F. and MacCannell, E. H., 1955. Basic problems, techniques and theory of isopleth mapping, *Journal of the American Statistical Association*, 50: 220–239.

Silverman, B. W., 1986. *Density estimation for statistics and data analysis*, Chapman and Hall, London.

Tobler, W. R., 1973. Choropleth maps without class intervals, *Geographic Analysis*, 5, 262–265.

Tobler, W. R., 1979. Smooth pycnophylactic interpolation for geographic regions, *Journal of the American Statistical Association*, 74: 519–530.

Unwin, D. J., 1981. *Introductory Spatial Analysis*, London: Methuen.

Wright, J. K., 1936. A method of mapping densities of population with Cape Cod as an example, *Geographic Review*, 26: 103–110.