Environment and Planning A 2000, volume 32, pages 1499 – 1518

DOI:10.1068/a29165

Defining locality boundaries with synthetic data

Mike Coombes
North East Regional Research Laboratory, Centre for Urban and Regional Development Studies, University of Newcastle, Claremont Bridge, Newcastle upon Tyne, NE1 7RU, England;
e-mail: mike.coombes@ncl.ac.uk
Received 14 November 1996; in revised form 2 May 2000

Abstract. Boundaries can be defined by applying a number of alternative techniques to various types of information; the choices made on these decisions should reflect the purpose for which the boundaries will be used. In this paper I report research to define a set of localities intended to provide an up-to-date and relevant definition of local communities across Britain for the very varied purposes of academic social scientists. First, the multidimensional nature of modern localities are outlined. Three established types of regionalisation procedure are then reviewed, leading to the identification of an appropriate method for analysing a suitable dataset. At the same time, it is concluded that a broadly based set of boundary definitions requires a more innovative approach in order to collate the many strands of evidence which are relevant to locality definitions. The response here is the development of *synthetic data* which codifies the critical information in any boundary set into a form which can then be combined with other, similar evidence. This leads to the first empirical challenge, which is to collect or create boundary sets which each provide a relevant strand of evidence for locality definitions. I apply the preferred regionalisation method to the synthetic data which has been created, and illustrate the localities which have been defined on this basis. I end by suggesting some other ways in which synthetic data might be analysed to provide insights into patterns of spatial association at the local scale.

Regionalisation is the branch of spatial analysis whose results take the form of boundaries. In most cases, the boundaries are defined so as to distinguish one cluster of interacting or related areas from the next by, for example, identifying local labour markets through the analysis of commuting data. Methodological development has so far been limited by the need to choose the single 'least worst' run of a single method on a single dataset. The new spatial data handling opportunities provided by geographic information systems (GIS)—together with hugely increased computing power—should now open up other opportunities, including the possibility of adopting a more 'fuzzy' approach in which the evidence need not be limited to that of a single analysis on a single dataset. In this paper I report the development of a new method which was stimulated by the need[(1)] in Britain for locality boundaries for social science researchers with a diversity of interests. The definition of locality boundaries on the basis of the evidence from *any* single dataset alone cannot consistently provide an adequate set of boundaries which will be relevant to a wide range of researchers.

I begin by highlighting the challenge which defining localities poses, not so much because of the contested nature of the locality concept itself, but because the concept is multidimensional and so not well suited to being represented by analysing a dataset related to a single topic. In the second section of the paper I then outline the main types of regionalisation method, reviewing their potential value for the definition of multi-issue locality boundaries. Next, I summarise a particular regionalisation algorithm

[(1)] The set of locality definitions produced by the research reported in this paper are available free to academic researchers via the Mimas national data service at Manchester University or by e-mail from the author. They are also available on a CD supplied with *The Census Data System* (Rees et al, 2000).

which can be customised for application to different datasets, and I set out the key idea of synthetic data, which allows the information present in many different datasets to be brought together. I go on to detail the initial application of this form of data integration to define locality boundaries in Britain, first identifying the relevant information which could be collated in the form of synthetic data, and then reporting the results of the analysis using the regionalisation algorithm described earlier. I conclude by previewing some further developments in boundary definition which are opened up by the innovation of synthetic data.

Localities: problems of definition

The term 'locality' is used here to denote an area—in contemporary Britain, in the context of this paper—within which particular socioeconomic processes have created a distinct set of circumstances and pattern of interaction which, at that time, constitute a separably identifiable part of the country (Massey, 1991; 1995). Localities were traditionally identified by reference to static features such as topography, which can create separateness through barriers such as hills or rivers, but recently they have been seen as clusterings of local flows and interactions, an approach which is at least partly rooted in the models of time geography (Pred, 1984; Thrift, 1983). In other words, defining localities calls for boundaries which indicate where one localised complex of flows and characteristics shades into the fairly distinct clusters in adjacent areas. This is clearly not to say that such a boundary demarcates an area within which there is no variability, just as it is not to claim that two localities which are separated by such a boundary have no links or characteristics in common. The idea is similar to that which underlies boundaries drawn in many other contexts: for example, Upton and Widdowson (1996) say that the boundaries of their dialect maps should be read as showing "areas shading into one another, rather than being sharply demarcated" (page xviii). In this paper, I seek to devise a method to identify areas which are consistently distinctive enough to be considered as localities in their own right; the aim in so doing is to provide a satisfactory set of locality definitions in Britain at the end of the 20th century.

A basic assumption here is that locality boundary definitions should be consistently defined, so as to avoid the problems which arise from 'comparing apples with pears' in subsequent social science research for which the definitions are being made available. Comparing areas *without* the confidence that they have been consistently defined creates problems for both in-depth local studies—where the definition of the areas must intimately shape the findings—and also in comparative spatial analyses, for which the importance of study-area boundaries in shaping results has been termed the modifiable areal unit problem (Openshaw and Taylor, 1981).

As in most countries, local authority areas are the primary candidates in Britain as a set of predefined locality boundaries. These areas have recently been revised but the new set of boundaries have, if anything, been defined in a way which makes them even less relevant for researchers than were the old ones. The problem is the lack of consistency in the approach to local authority boundary definition—an inconsistency notably heightened in the recent revisions which not only proceeded on entirely different bases in Scotland and Wales, but also excluded Greater London and the other metropolitan counties from the English review, where a consistent approach even to the rest of that country was rejected explicitly.

The crucial issue in defining British local government boundaries—as for much of modern social and economic geography and policy (Greenhalgh et al, 1997)—is the extent to which the urban areas remain detached from the surrounding, more prosperous rural areas with which, in reality, they are intimately linked. The local authority district boundaries of some towns and cities (such as Leeds and Carlisle) embrace quite

Figure 1. Local authority district boundaries to the north of the London region.

large tracts of rural land, while other similar cities (for example, Leicester and Norwich) do not even include the whole of their own built-up areas. Figure 1 shows the district boundaries in the northern half of the London region, readily illustrating their inconsistent relationship to built-up areas (shaded in the map). For example, the northern part of the map shows Bedford and Cambridge which are two similarly sized county towns: Bedford urban area forms just part of a district whose boundaries are extensive enough to include all the rural areas nearer to it than to any other towns, whereas the Cambridge urban area makes up a district of its own and thus excludes all the neighbouring suburban and rural areas. Inspection of other parts of the map throws up similarly inconsistent treatment of nearby towns and cities of similar size: for example, Aylesbury and Luton, or Stevenage and Chelmsford (all in the central part of figure 1). The first objective for this analysis is thus to provide a consistently defined set of boundaries, but this objective is to be combined with the aim of creating a *multipurpose* set of locality boundaries of potential value for a wide range of different social and economic studies.

It is this second objective which militates against the reductionism of previous approaches (a reductionism which, in fact, was as much a result of the technical limitations faced by earlier regionalisations as it was of narrow objectives). For example, many sets of boundaries used in locality studies have been based exclusively on labour-market patterns. Of course, for some specific purposes it is most appropriate for the boundary definitions to be based on the analysis of the single dataset which is peculiarly relevant to that purpose: for example, Maclennan and Bannister (1995) call for more analyses of migration patterns as a basis for identifying local housing-market areas. In contrast, the objective here is to define areas suitable for a *broad* range of social science purposes. A brief indication of the concerns which such locality definitions might need to represent can be gleaned from the thirty-one proposed locality studies collected in Cooke (1986). Over half of the studies did take local labour-market patterns as the key defining feature of the locality they proposed to study. Most of the others adopted a local authority district boundary—although few rationalised this choice to any extent. The other feature which attracted attention was local social and/or economic history, often focusing on a particular local industry which set an area apart from its neighbours through the development of a distinctive social structure. These three social science emphases—on areas' population structures and histories, their current administrative boundaries, and their labour-market patterns—can be combined with more physical aspects of localities, namely the landscape-related features (for example, the identity of a readily defined valley) and the consideration of access and facilities (for example, the hinterland for the services of a town or city). These five different facets to the notion of locality cannot be reduced to one single dominant aspect, according to which localities' boundaries can then be adequately defined. As a result, the second objective for any analysis which aims to produce genuinely multi-issue definitions of localities is that it must reflect *most* of the facets of locality touched on here, that is to say, local institutions, demography, economy, facilities, and landscape.

In summary, the concept of locality is multifaceted and, as a result, it is scarcely to be expected that appropriate locality definitions can be found 'off-the-shelf' among the available sets of boundaries in Britain (such as those of local authorities). Hence, the objective here is to apply a method of boundary definitions which can provide consistently a multifaceted interpretation of the locality concept. In the next section of this paper, I review regionalisation methods with the express aim of identifying one or more with the potential to create consistent multifaceted definitions.

Regionalisation methods: an overview

Regionalisation was seen as a crucial challenge by pioneering quantitative geographers. For example, Hess and Weaver (1965) demonstrated the potential of new methods to meet the demand for objective boundary definitions which arose with the 'redistricting' task of creating electoral areas which consistently meet such statistical criteria as having roughly similar population sizes. In response, the International Geographical Union convened a Commission on Methods of Economic Regionalisation (see, for example, Hamilton, 1969). The emphasis of the Commission's work was on the range of specific approaches to boundary definition which were then in use or under development; not surprisingly, much of their detailed discussion is now of limited interest when compared with more recent methodological developments. However, the key issues covered by the Commission were included in a broader review of methods for grouping areas by Spence and Taylor (1970), and their review established some important principles which remain relevant. In particular, regionalisation was placed in the context of other taxonomic techniques, wherein it was defined as the form of area classification within which each class is normally a single group of contiguous areas. Thus, methods of *functional* regionalisation—that is, those methods which are specifically designed to create boundaries from the analysis of datasets on commuting, migration flows, or other forms of *interaction* between areas—can be set alongside different forms of analysis which may also be used for regionalisation if they can be constrained so that they group only contiguous areas.

In practice, there are many methods of numerical taxonomy (Sokal and Sneath, 1963) which could be combined with a contiguity constraint to produce a regionalisation from an affinity matrix (for example, the output from a factor analysis of areas' attributes). One important potential advantage of this approach is that it would allow regionalisation to draw on information other than interaction datasets, which are the customary prerequisites of functional regionalisation. This is likely to be essential if the analysis is to come anywhere close to providing a multifaceted definition of localities because interaction datasets are scarce and usually restricted to just a few aspects of local geography, such as commuting patterns. However, there is one crucial advantage for regionalisation of analysing interaction datasets: the 'friction of distance', which tends to restrict people's patterns of movement, although varying between different groups of the population (Simpson, 1992) in general causes the strongest interactions to be between nearby areas. As a result, contiguous groupings are inherently more likely to be produced from the analysis of interaction data than they are from the analysis of other types of data. The fact that a regionalisation based on an affinity matrix requires a contiguity constraint to be imposed on the analysis, explicitly shows that the required contiguous groupings must be suboptimal with respect to the criteria for selecting the strongest link; otherwise they would have emerged without the constraint. For example, in a dataset of neighbourhoods, the city centre of Manchester is likely to have greatest affinity with the centres of other cities. However, regionalisation will require that each city centre is grouped with other parts of its own city. Therefore, a contiguity constraint is required in order to prevent the analysis from grouping all the city centres together (without their suburbs), because this is likely to be the optimal grouping in terms of their affinity alone.

The disadvantages of contiguity constraints include the influence on the results of "irregular base areas" (Spence and Taylor, 1970, page 30). There are numerous bizarre boundaries among areas such as parishes or wards which are the common 'building-block' areas for which the data needed for regionalisation analyses can be obtained, and these boundaries will restrict the options as to which areas can be grouped at every step of an analysis which is explicitly contiguity constrained. At the time that they were writing, Spence and Taylor (1970) needed to recognise the pragmatic counter-argument that including a contiguity constraint vastly speeds up the analysis, for the very reason that it restricts to a handful the number of permutations considered when assessing which is the most appropriate grouping. This argument now has little strength because the huge increase in computer speed since 1970 allows vast numbers of permutations to be evaluated even within a very large analysis. As a result, methods which require an explicit contiguity constraint at every stage of the analysis will inevitably tend to fall well short of providing an optimal form of boundary definition.

A typology of methods

As part of their review, Spence and Taylor (1970) indirectly drew attention to the fact that regionalisation research has been characterised by the diversity of methods used, and it is still true that there has yet to be any convergence on a single widely recognised 'best practice' approach. Even so, there remain some fundamentally distinct options which have been discussed in the literature and yet have rarely, if ever, been implemented. For example, almost all methods have been developed to meet the commonplace practical requirement for nonoverlapping regions; very different methods might well be required to produce well-defined overlapping regions. It is also notable that almost all regionalisation methods are agglomerative, rather than starting with the country as a whole and then seeking the most appropriate series of partitions. Setting aside these alternatives, regionalisation methods can be seen to fall into the three types of agglomerative procedure which are now briefly outlined.

The first category of procedure is typified by cluster analysis and customarily proceeds in a single step from individual building-block areas to the final set of regions. Instead of a readily recognisable key rule or measurement—such as the size of the commuting flows between two areas—the grouping of *clustering* methods usually relies on abstract measurements of the relative similarity in the overall statistical properties of the areas, as measured in an affinity matrix. Spence and Taylor (1970) cite the analysis of Hagood (1943) as the first quantitative regionalisation which they place within the clustering type of method. They also include within this type the graph theoretic method of Nystuen and Dacey (1961), and most redistricting techniques (such as Horn, 1995) can be added too. Other methods which are forms of clustering, as defined here, are the location–allocation algorithm of Lolonis and Armstrong (1993) and the automatic zoning program of Openshaw and Rao (1995).

Spence and Taylor (1970) termed the second type of procedure *hierarchical* because the criterion for grouping areas is gradually lowered so that, step by step, the process builds upon earlier groupings until all the areas satisfy the criterion which determines when the procedure ends (for example, when the regions are all of a certain size, or when there are a certain number of them). Hierarchical methods developed rapidly in the 1970s as computerised interaction data such as commuting flow matrices began to be more frequently available. Smart (1974) led the way with an algorithm which, remarkably enough, was in fact applied manually to a matrix of flows between nearly 2000 base areas. Masser and Brown (1975) and Slater (1976) were among those who developed computerised hierarchical algorithms, presenting the results as dendrograms in which the choice of the number of regions defined is made clear. Masser and Scheurwater (1980) confirmed that these methods all accept groupings which were to a greater or lesser extent suboptimal in order to ensure the contiguity of the regions defined (and also, at that time, as a way of reducing the computing task by limiting the number of possible permutations). Computational capacity need no longer constrain these methods, although the French hierarchical algorithm finds it necessary to retain a contiguity constraint in order to obtain results which satisfy users' understanding of their local geography (Documentation Francaise, 1996).

A third kind of procedure applies a fixed rule or rules which identifies those areas which are sufficiently linked to be combined in the grouping process. The number of regions produced by *rules-based* methods is thus not known at the outset and, in some methods, there is the option of the analysis being completed with some areas remaining unallocated to any group. This can occur because rules-based methods often start by defining an initial set of core areas from which to build up the final regions; the noncore areas are thus residual and only became part of a region by being linked, directly or indirectly, to one of these core areas. The two aforementioned approaches tend to start with each building-block area as, in effect, a potential region in its own right, so that if any of them is not linked to others during those types of analyses then they are deemed to be single area regions (rather than being unallocated areas, as are the equivalent unlinked areas with rules-based methods).

Another distinction between the rules-based approach and the other two types is that the former is likely to have devised its criteria for grouping areas from a geographical model, in contrast to the more general principles underlying a method such as graph theory or factor analysis. For example, the rules-based method for defining metropolitan statistical areas in the USA continues to be based on a model of metropolitan areas which assumes that each one is centred on at least one city, and these cities in turn provide the focus for the patterns of commuting flows (Spotila, 1999). By contrast, hierarchical and clustering methods are based on more generalised principles (such as minimising the level of interaction, or of similarity across group boundaries) and these principles rarely reflect a specifically geographical model of urban or regional systems. The conceptual basis of rules-based methods is not a sufficiently strong advantage by itself to choose the rules-based approach for this research, because it is unlikely that a single geographical model of localities provides the ideal basis for the definition of boundaries which are to be used for many different social science purposes.

Identifying appropriate methods for defining localities

All three approaches include methods which allow statistical or other objectives to be set for the analysis. For example, the regions might be required to be as compact in shape as possible (Martin, 1997), and perhaps to be subject to a number of other constraints such as maximum and/or minimum population sizes. The task here will be to set relevant objectives which the locality definitions need to meet, and against which the definitions should be seen to be the 'optimal' solution. Another objective relates to the number of regions produced: in some regionalisations an absolute number of regions is required (for example, 'define four parliamentary constituencies within the county'), but here there is the broader task of identifying as many localities as possible which all meet the specified objectives.

The hierarchical method of grouping areas step-by-step also progressively raises the statistical characteristics of the areas (for example, the population size of the smallest remaining region). As a result, these methods allow the objectives for the analysis to be set in terms of either the required number of regions or the statistical properties which the defined regions must possess. It is the latter option which is relevant here, not least because there is no prior knowledge of how many localities Britain should be divided into. The rules-based approaches also provide an appropriate option in this respect, because for them the number of regions is always an unknown at the start of the analysis. By way of contrast, clustering methods usually specify the required number of regions at the outset, which makes them less appropriate for this research than the rules-based and hierarchical approaches. There *is* the option of producing several clustering analyses—each with a different number of regions—then selecting the option which performs best against the statistical objectives. Even so, the most familiar clustering methods cannot ensure that every region produced will meet the minimum statistical objectives set at the outset. Adding this constraint, along with the explicit contiguity constraint which clustering methods require, reduces the options open to the analyses to such an extent that the results are likely to be far from optimal.

Hierarchical procedures are inherently prone to providing suboptimal results at larger scales, for the simple reason that the area groupings made at the early stages of the analysis severely restrict the options available at the later stages. For example, a grouping of European countries might be expected to link Luxembourg with Belgium and then with the Netherlands in its early stages; in a hierarchical procedure, these early groupings would then prevent the later stages creating what may be the optimal broader groupings in which, for example, the Netherlands could be linked with other northern countries which speak Germanic languages, whilst Belgium and Luxembourg were grouped with France and other more southern countries. In general terms, the problem here is that the most appropriate set of results at any one scale of resolution often needs to be able to prioritise one subset of the data used for the analysis, when the most appropriate set of results at a different scale needs to focus on a different subset of the data. Thus, hierarchical methods will be suboptimal because their groupings of areas at earlier steps of the procedure remain preserved throughout the rest of the procedure—and thus shape all of the subsequent groupings—when they are unlikely to be very relevant at this broader scale of analysis.

Most of the early rules-based methods were implicitly also hierarchical, because they proceeded sequentially and so their early groupings were preserved in the later results. More recently, some rules-based methods have been developed to 'escape' from being hierarchical through procedures which ensure that the initial groupings need not be preserved within the later groupings' boundaries. For example, Coombes et al (1982) developed a multistep rules-based method which featured this form of procedure, and Coombes et al (1986) made this form of self-optimisation the central component of the method to define travel-to-work areas (TTWAs). In addition, the TTWA method routinely produced contiguous regions without using an explicit contiguity constraint, relying upon the self-optimisation feature to maximise the inherent tendency of interaction data strongly to link nearby areas. Eurostat (1992) recommended that the

TTWA method—or the generalisation of it known as the European Regionalisation Algorithm (ERA)—should be seen as the standard for defining local labour-market areas in European countries. Variants of ERA have indeed been successfully applied in Italy (Sforzi et al, 1997) and parts of Spain (Casado-Diáz, 1996). Henry et al (1995) provided a different rules-based approach, which also features an optimisation procedure, but this uses a contiguity constraint and so must be prone to producing less optimal results. In a recent contribution to the national debate in the USA on how further to develop their rules-based definition of (non)metropolitan areas, Frey and Speare (1995) cited as 'state of the art' the methodological developments underpinning the ERA method in particular.

This review has concluded that methods can only consistently approach an optimal solution if they dispense with any explicit contiguity constraints—which disallows clustering methods—and can avoid the 'chaining' problems which are characteristic of hierarchical methods. Turning then to the rules-based methods, ERA provides a plausible candidate as the 'best practice' regionalisation procedure available.

The basic methodology

Defining boundaries with a procedure such as ERA prompts the analyst to specify the criteria which the defined areas have to meet. In general, it is assumed that the objective is to maximise the number of separable regions capable of meeting these criteria (*not* to maximise a predefined number of regions' values on these criteria). Statistical criteria are specified for ERA—as both 'target' and absolute minimum values—on each of two (or more) key criteria (for example, population size and local separateness). If there are two criteria, then regions must meet *both* the minimum requirements *and* at least one of the higher target values (or be close enough to both target values to pass a test for the 'trading off' of the value on one criterion against that on the other). The standard ERA criteria are: size of population, and self-containment (in other words, a measure of separateness shown by the relative lack of linkage across the proposed locality's outer boundary).

The basic features of the ERA method derive from the TTWA algorithm (Coombes et al, 1986). Although remaining a multistep method, the ERA procedure has been simplified to remove numerous steps which require the setting of applications-specific parameters (for example, 'select areas which score higher than x on parameter y'). Reducing the procedure to its basic components makes it much more transferable, in that it can be used with different datasets and for different purposes without long periods of experimentation to determine the optimal settings on numerous parameters. There are up to five steps in the ERA procedure:

(1) identify foci around which the regions are to be built;

(2) group any foci which are closely interconnected (for example, areas identified as foci which were really separate only as a result of the vagaries of the building-block areas' boundaries);

(3) progressively associate all nonfoci areas to one or other region;

(4) disassemble regions which fail to meet the statistical criteria (starting with the region furthest from meeting the criteria) and associate their constituent areas individually with one of the other regions;

(5) consider further 'optimising' these draft definitions, perhaps by inputting them to a procedure which attempts to improve the regions' overall score on the statistical criteria by shifting the region boundaries by one building-block area at a time and/or by carrying out a consultation process (but only accepting well-argued changes which do not lead to any region failing the statistical criteria).

If the application is one for which nodal regions are not particularly appropriate, then every building-block area can be deemed to be a potential region in its own right: the ERA analysis then starts, and may indeed finish, with step (4) in these cases. It is step (4) which gives ERA its self-optimisation characteristic, and which allows the groupings to escape from being hierarchical while also tending to produce contiguous regions without using an explicit contiguity constraint.

Even if it is accepted that the ERA software is the 'best practice' method available, its use cannot sidestep the inherent limitations of creating locality boundary definitions which are based on a single dataset. In the first place, of course, the evidence on locality boundaries which is provided by any single dataset will be inadequate when multifaceted definitions are required by researchers for many different purposes. Commuting flows are probably the most widely used data source in regionalisation, but commuting patterns alone can be at best no more than a very partial proxy for the more rounded approach to boundary definitions called for by, for example, Claval (1987). In short, many other facets of localities need to be reflected in their definitions.

A second fundamental problem is that any regionalisation analysis is only one of many possible analyses of the dataset concerned. No matter how good the regionalisation algorithm, the very best it can aim for is optimality in terms of a specific selection of criteria. These criteria will include a number of basic parameters—such as a minimum or maximum size—and the results of the analysis will be sensitive to the values on these parameters. With most algorithms there are also technical parameters which provide the mechanism by which the expertise and experience of the analyst can contribute to producing more optimal results. Thus, although a single run of a regionalisation procedure is strictly deterministic, there is a wider form of flexibility to the analysis because the results produced would change if key decisions made by the analyst were altered.

One aspect of this sensitivity is the fact that any single analysis of a large number of areas customarily produces suboptimal (or even positively paradoxical) results for at least some areas. Yet a set of only slightly differing analyses may well, between them, include an appropriate boundary for every part of the country, because the unsatisfactory boundaries produced by each of the individual analyses tend to be in different areas. The traditional response to this sensitivity has been to select the single *least worst* set of results, then manually adjust the minority of unsatisfactory boundaries. A more appropriate response now is to see all these slightly different sets of results as parts of a 'fuzzy' picture, underlying which is the recurring pattern that the analysis aims to reveal. To reveal this underlying picture, the methodological innovation here hinges on splitting the whole regionalisation procedure into two phases:

phase 1—compile numerous analyses from numerous datasets; and then

phase 2—collate the results from these analyses within a single synthesing analysis.

It is asserted here that ERA provides an appropriate algorithm for much of the first phase of this strategy, so it is in the second phase where the key innovation is required. The solution centres on creating *synthetic data* which provide the basis for phase 2 of the method, using as input the range of evidence provided by the phase 1 analyses (together with any other relevant and collatable information).

Synthetic data

The essential basis for the synthetic data is the understanding that a set of boundaries is a classification in which each building-block area is allocated to one and only one region. Thus, each analysis undertaken in phase 1 of this approach can produce a classification of all the 10 529 areas from which the locality definitions here are to be composed. For each area, the initial information in each classification is the region number to which that area is assigned. The crucial step here comes from realising that this information can be transformed into binary data in a matrix by taking each pair of areas and identifying whether they are (1) or are not (0) classified into the same region. In this way, each classification becomes expressed as a binary matrix of $10\,529 \times 10\,529$ cells (although the matrix is in fact symmetrical, so only half of it is needed). For example, if area B was in the same region as area C but in a different region to area D then the cell BC would take the value 1 while cell BD would be 0 (and cell CD would also take the value 0).

The crucial benefit from reexpressing each separate classification in this binary form is that these matrices can then be cumulated to produce the synthetic data needed. For example, if the results from three phase 1 analyses were collated in this way, the value in each cell of the synthetic data matrix would vary from 0 (for any pair of areas which were not in the same region according to any of those analyses) up to 3 (for any pair of areas which all three analyses had put into the same region). In GIS terms, it is analogous to layering the sets of boundaries on top of each other and counting the number of layers in which there is no boundary between each pair of areas. It can be seen that this approach provides an assessment of the 'strength of evidence' that two areas should be grouped together. The regionalisation can then be seen to be searching for area groupings which recur in most of the input classifications; this operationalises a principle foreshadowed by Cole (1921) who saw that different sets of regions often have much in common in terms of their grouping of areas, even though their boundaries appear to be very different. The final synthetic dataset then provides the ideal basis for the second phase of the definitional procedure. This dataset is analogous to an interaction matrix because it represents the level of connectedness of pairs of areas and, like a commuting or migration matrix, it can be analysed with a variant of ERA which has been optimised for this purpose.

The *methodological* innovation of creating synthetic data removes the technical limitations which arise from relying upon a single analysis of a single dataset. In particular, something of a 'fuzzy' approach becomes possible because, in phase 1, more than one form of analysis can be applied to the same dataset, thereby accepting that each of these may provide a different but equally valid insight into the patterns lying within that dataset. Each of these analyses can then provide a separate input classification to the synthetic dataset, which will then assess which findings these analyses have in common and those on which they differ.

The *substantive* benefit of the synthetic data method is the ability to draw upon analyses of different datasets. The synthetic data require inputs in the form of boundaries: this allows the final definitions to draw on evidence not only from prior analyses of flow datasets but also from relevant nonflow information which can be expressed as boundaries. In short, creating the synthetic data within the two phase regionalisation method removes the key limitation in any single analysis of a single dataset (that is to say that only one dataset can be analysed, so multifaceted definitions could not be created). I now turn to an application of the synthetic data approach to the challenge of creating a consistent set of multifaceted locality definitions in Britain in the 1990s.

Localities in Britain

The first practical task is to complete phase one of the analysis, and thereby to collect together information on many different facets of localities. In practice, it was possible to bring a substantial range of relevant data together, although it is certainly not possible to claim that this information constitutes a fully comprehensive mapping of all aspects of British localities. Any such data compilation must remain constrained by pervasive limits on available information, and especially the scarcity of interaction data. There is also an inevitable degree of contentiousness in the choices made: for example, the absence here of information representing factors such as landscape characteristics is certainly open to criticism. Even so, some advances in terms of building a multifaceted set of locality definition are worthy of note.

(1) The scarce flow datasets have been disaggregated into distinct groupings so that, for example, the pattern of migration by older people can be considered separately (and as a result it is neither 'dominating' nor 'drowned out' by the rather different patterns of other age groupings).

(2) A confidential dataset on the 'journey-to-bank' patterns of account holders with a major high street bank provided access to some rarely available information on flows within the service sector (the raw data could not be made available for the analysis, but a prior analysis had produced a set of bank branch hinterland boundaries and these provided a valuable input to the synthetic data).

(3) The synthetic dataset has also been enriched by taking as a further form of input a range of existing sets of boundaries which are arguably relevant to this analysis of which areas are better kept together and which kept separate.

Table 1 summarises the information collated by phase 1 of the research. As suggested previously, numerous boundary sets can be defined from recent data by carrying out runs of the ERA software (for example, the nine sets of migration areas are each derived from the same type of ERA run, but applied to different subsets of 1990–91 Census migration data). Other boundary sets were not produced by analyses but are distinct 'geographies' which, like local authorities (figure 1), are included because they are aspects of modern life at the locality scale.

Table 1. Boundary sets for the locality definitions.

Facets of localities	1986–1995	pre-1986
Institutions	Parliamentary constituencies	Earlier local authority areas (2 sets)
	Local authority areas (3 sets)	
Demography	Migration areas (9 sets)	Functional regionalisation (4 sets)
Economy	Enterprise Council areas	Job Centre catchment areas
	Commuting areas (14 sets)	1981-based travel-to-work areas
Facilities	Bank branch catchment areas	
	Postcode districts	Postcode areas
Landscape		

Table 1 introduces a time dimension into the discussion here, showing that some earlier geographical patterns are also drawn upon as the inputs to the synthetic data. The argument for their inclusion here follows the emphasis in Johnston (1991) on the key role in local culture of the social and economic processes in prior periods. At a more mundane level, table 1 makes some distinctions which are more indicative than definitive in practice. For example, postcode *areas* (such as the large area around Chelmsford where addresses have postcodes beginning with CM) are put in the pre-1986 column because there has been stability in their boundaries; by contrast, postcode *districts* (such as the small area within which addresses have CM2 postcodes) are placed in the column for more up-to-date boundary sets, because there are annual changes in postcodes at this level of detail. Figure 2 (over) shows the postcode area boundaries in the north of the London region. This same part of the country is used for all of the maps in this paper so that the different boundary sets can be directly compared, and named places can be located using figure 1.

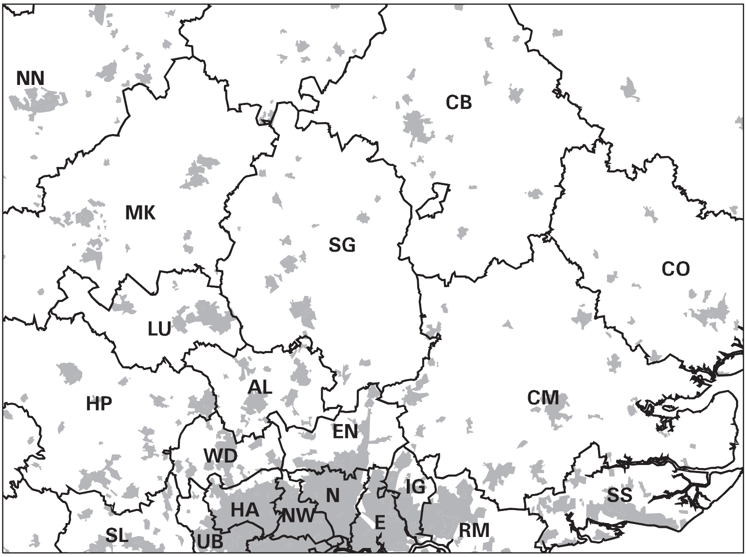


Figure 2. Postcode area boundaries.

Table 1 also sketches an allocation of the thirty-nine boundary sets between the five broad facets of localities which were identified in the first section of this paper. Once again, there are some debatable allocations of boundary sets between the different facets: for example, the functional region definitions (Coombes et al, 1982) drew upon both demographic and economic factors, so these areas could equally well have been allocated to the row below that in which they appear. One of the major problems which was encountered was the scarcity of boundary sets on the landscape-related aspect of localities; fortunately the case can be made that this is the least relevant aspect of localities to these boundary definitions' main purpose as units for social and economic studies.

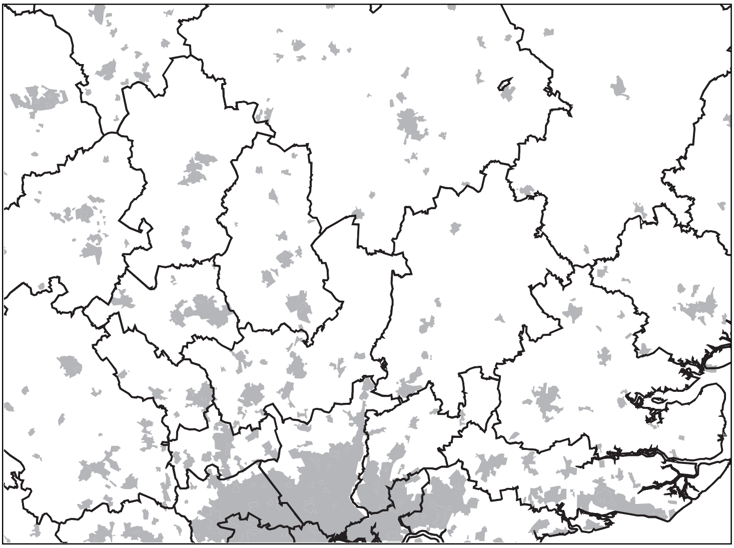


Figure 3. Migration areas.

Figure 3 shows the boundaries which result from analysing all migrants together so as to define, in effect, a set of housing-market areas. It is notable that in London (where higher proportions of people live in local authority housing) many of these boundaries exactly follow district boundaries (figure 1) because of the relative difficulty of transferring tenancies from one district to another. Even so, the boundaries around all the towns and cities in more rural areas (such as Cambridge and Luton) see the urban areas grouped with their rural hinterlands; all the towns and cities which have tightly bounded districts (figure 1) are shown here (figure 3) to be fully integrated with their surrounding countryside in terms of migration patterns.

Both the census migration and commuting datasets have been disaggregated so that, for example, migration analysis by age group can reflect the important contrasts in migration behaviour by life-stage (Flowerdew and Boyle, 1992). A key reason for these disaggregated analyses was to prevent the eventual locality definitions being dominated by the behaviour of a single predominant group (for example, middle-aged middle-class men). Figures 4 and 5 (see over) illustrate the variation which this analysis strategy has brought to the locality definitions: they show the labour-market areas of part-time workers and car-using commuters, respectively. These analyses are the result of applying the same variant of ERA with which TTWAs are defined (Coombes et al, 1986); the results differ dramatically because of these different groups' contrasting commuting trip lengths which, in turn, produce very differently sized local labour-market boundaries. Again, taking the Bedford and Cambridge areas as examples, the part-time workers' labour markets (figure 4) can be seen to be quite tightly defined around each town, with two distinct part-time workers' labour markets existing in the area of small towns lying between the two county towns. Figure 5 reveals the very strongly contrasting pattern for workers who commute by car, with all of these small towns, in this case, part of a very large labour-market area centred on Cambridge and its rapidly growing environs. It is notable that, as with these two maps, analysing smaller groups' flow data can lead to some fragmentation of the boundaries because of the small numbers involved in a matrix of flows in excess of 10 000 areas. The response here has been to include those boundary sets which group the vast majority of areas into intuitively plausible boundaries; the small minority of anomalous results—which are

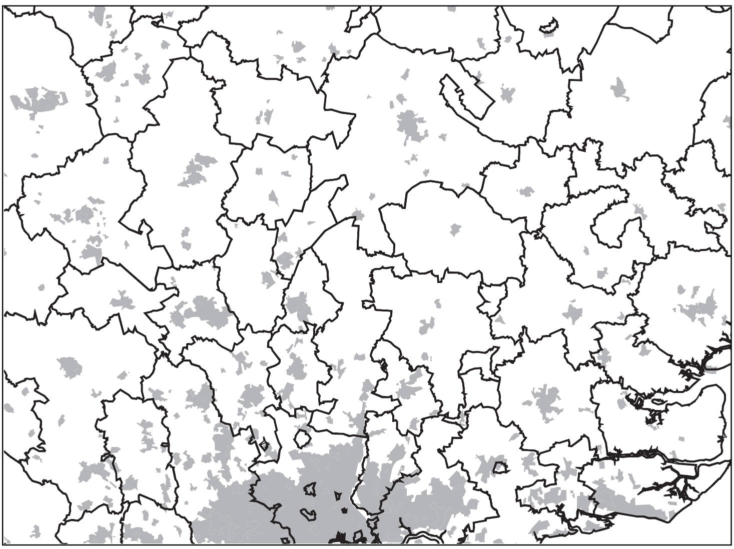


Figure 4. Labour-market areas of part-time workers.

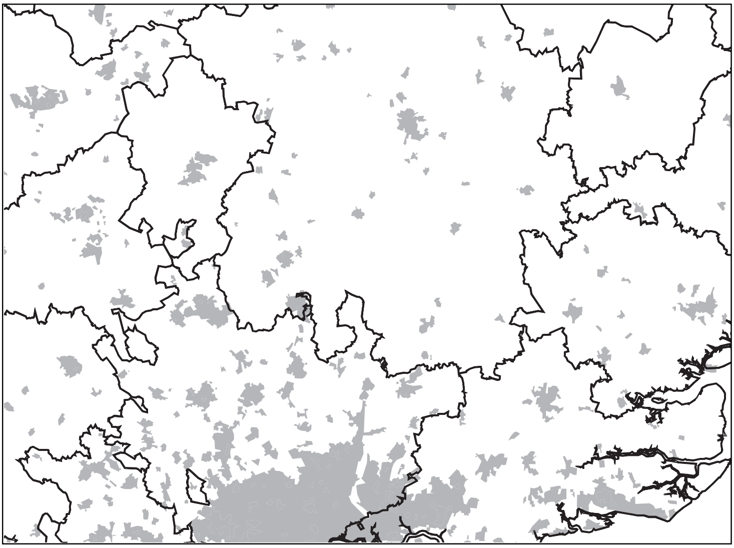


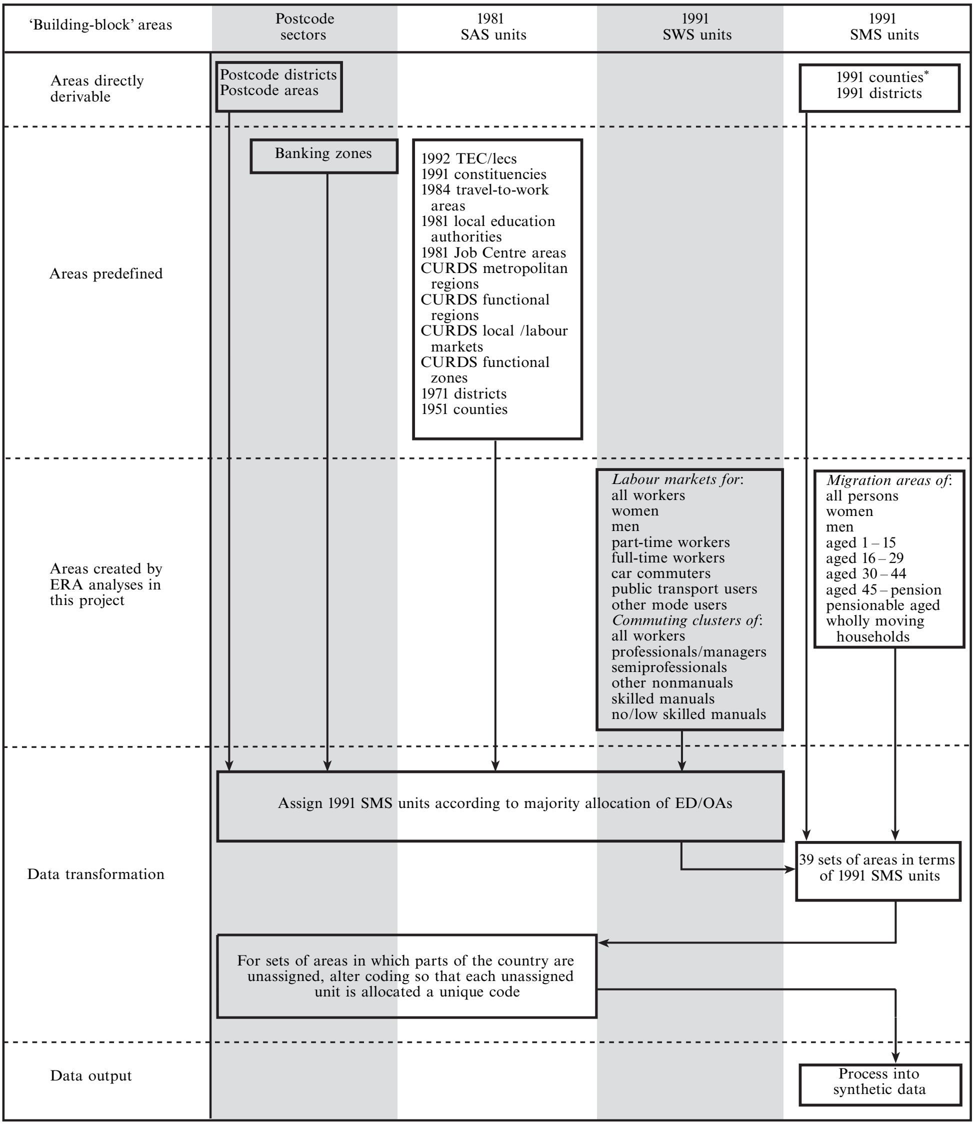
Figure 5. Labour-market areas of car-using commuters.

to be expected from any single analysis, as noted earlier—provide part of the complexity which the synthetic dataset is seeking to handle.

A final note on the phase 1 analyses also relates to the commuting analyses. Figure 6 provides some further information on the boundary sets selected as inputs to the synthetic data, and shows that some are termed *commuting clusters* (Coombes et al, 1996). In these ERA analyses, the statistical criteria which each area must satisfy were lowered from the TTWA values (the basis of the local labour-markets cited in figure 6) so as to identify potentially neighbourhood-level commuting patterns. This variation of the ERA methodology serves to indicate both the valuable flexibility of the ERA software and also a way in which the synthetic data can incorporate something of a 'fuzzy' procedure, through carrying out different analyses on the same data so as to reduce the risk of idiosyncratic results due to the purely technical effects of the method used.

Figure 6 not only lists the thirty-nine sets of boundaries which have been compiled within this phase of the research, it also shows that the inputs were obtained in terms of four principal sets of building-block areas (as a result of compiling boundaries from different sources and eras). Bringing this disparate information together onto the common building blocks of 1991 Census areas would have posed a greater problem were it not for earlier work by Owen et al (1986) to compile the majority of the boundary sets which were accessed in terms of 1981 wards. The contribution of Atkins et al (1993) was also essential here, because this provided the link between 1981 and 1991 Census areas. Figure 6 outlines the procedure through which all the boundary sets become expressed as a 'best fit' allocation[(2)] of 1991 wards, allowing each then to contribute to the compilation of synthetic data.

[(2)] The procedure used to achieve this involved an initial point-in-polygon allocation to a finer set of areas [that is to say, 1991 enumeration districts (EDs) in England and Wales, and 1991 output areas (OAs) in Scotland] so as to maximise the precision of the results. The procedure can be illustrated with the 1981-based TTWAs: (1) obtain the areas' definitions in terms of a look-up file (for the TTWAs this is in terms of 1981 wards); (2) use a preexisting file to allocate 1991 EDs or OAs to the areas [in this case, the Atkins et al (1993) file links 1991 EDs or OAs to 1981 wards]; (3) use the nesting of 1991 EDs or OAs into 1991 wards to allocate the latter to the areas (where a ward's EDs or OAs are allocated to differing TTWAs, refer to the EDs' or OAs' population and then allocate the ward to whichever TTWA embraced most of the ward's population on this basis).



Note: *Unlike in the Census data, Greater London is treated as a single unit (as in each metropolitan county and Scottish region). Key: SAS = small area statistics, SMS = special migration statistics, SWS = special workplace statistics, ED/OAs = enumeration districts (in England and Wales)/output areas (in Scotland).

Figure 6. Compiling synthetic data from thirty-nine sets of boundaries.

The banking zones were 'weighted' by being entered into the eventual synthetic dataset four times. This step ensured that this rare evidence on British people's journey-to-services could have a substantial influence on the results when (because the data could not be reanalysed to draw out different strands of information on service-centre hinterlands) it might otherwise have been overwhelmed by the much more extensive evidence on commuting and migration patterns. This device of multiple inclusion of a single input also demonstrates another methodological option within the synthetic data strategy.

Following this process of data collation, the synthetic dataset was then generated by applying the principles outlined earlier in this paper. Given the heterogeneity of the inputs from phase 1 of the analysis, there was every possibility for the values in the

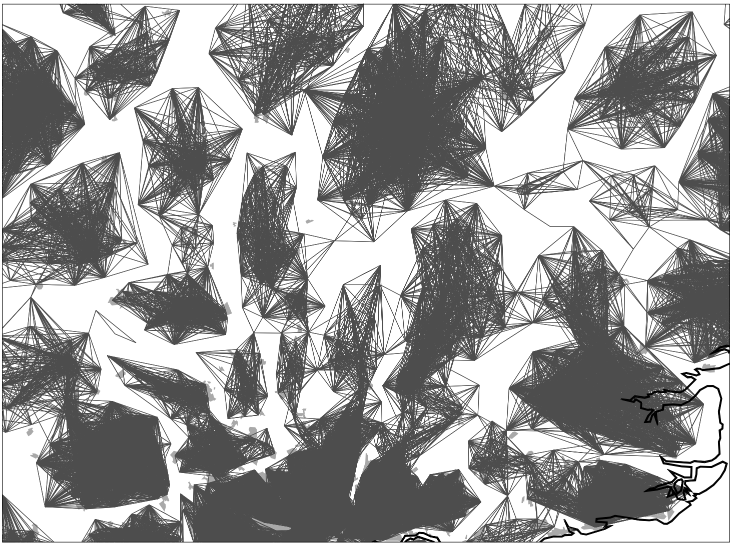


Figure 7. High synthetic data values.

synthetic data matrix to be very diffused—that is, for there to be evidence of at least some 'linkage' between any one ward and many other wards. The outcome is indeed a matrix which includes substantially fewer zeros (pairs of wards with no link at all) than does any of the census commuting or migration datasets. The next question, then, is whether the values are semirandomly scattered across the matrix or whether they constitute evidence of recurring linkage patterns across the varied inputs. Figure 7 reveals that the latter is very much the case: here, a line is drawn between pairs of adjacent wards whose value in the synthetic data is at least 60% of the maximum possible. Figures 1–5 showed that the input data had included some classifications typified by very large-scale groupings, some classifications in which many neighbouring areas were kept separate, and also some which included fragmented boundaries. Yet, the synthetic data succeed in finding recurring patterns of linkage at the very scale which perhaps the term 'locality' suggests in the modern British context. Bedford and Cambridge emerge unsurprisingly as separate clusters, with the former having significant links with an area to the south which is also linked to the substantial Luton-centred cluster. It is particularly interesting to see that some gaps are visible within London's massive cluster of linkages. Perhaps unsurprisingly there is little linkage across the River Thames in the east of the capital, but it was certainly less predictable that the small River Lea would also cause such a gulf to emerge (figure 7 shows this as a gap, 'pointing north' towards Cambridge, in London's massed linkages at the centre of the southern edge of the map). Even more remarkable is the gap which is visible in London's western suburbs; here a corridor of radial roads, railways, and a canal (to the south of Wembley) combines with other land uses substantially to inhibit interaction between the north-western and western suburbs, as can be seen from the fact that only in one of the previous maps (figure 5) did the boundaries group both sets of suburbs together.

As intimated earlier, the synthetic data matrix can yield a regionalisation by applying the ERA to it, after setting the statistical criteria to appropriate levels. In this analysis, the TTWA population minimum of 7000 was adopted whilst the target value was set at 50 000 (Coombes et al, 1982). The other key measure is self-containment, which assesses how far any boundary is grouping together areas which tend to be linked to each other, rather than to others lying across that boundary. This measure is

Figure 8. Localities.

entirely appropriate for an analysis of the synthetic data, and it ensures that areas which become grouped into the same locality must have been likely to have been grouped together by most of the input sets of areas. The set of localities defined in this way are illustrated in figure 8 and do appear to conform to 'common sense' expectations, such as that they should generally include a main town—perhaps together with some nearby satellite towns—along with its more rural hinterland. For example, the boundaries around Bedford and Cambridge (figure 8) can be clearly seen to reflect the principal patterns in the synthetic data (figure 7), with the small intervening towns grouped into other localities in the way which the synthetic data had foreshadowed. London also provides notable examples of patterns in the synthetic data being reproduced in the locality boundaries. The capital is partially subdivided, with the widely known East End emerging as a separate locality (whose boundary includes the Lea valley, noted earlier as a 'gap' in the synthetic data) and another boundary following the previously mentioned 'gap' to the south of Wembley as it groups the capital's western suburbs into a locality which also includes the Heathrow area.

It is not coincidental that the number of localities defined here (307) is quite similar to the 281 local labour-market areas of Coombes et al (1982), the 322 TTWAs which were defined with the 1981 Census data, the 297 TTWAs defined using 1991 Census data (ONS and Coombes, 1998), and indeed to the 408 local authorities resulting from Britain's administrative boundary revisions during the 1990s. The urban system of Britain clearly leads to a degree of commonality in sets of boundaries. By returning to the example of Bedford and Cambridge once again, these two towns are repeatedly recognised as individual areas by all the sets of boundaries discussed above, but the partition of the area lying between them varies markedly from one set of boundaries to the next. From a more geographical viewpoint, further breakdown of these areas would tend to produce less comparable areas because inevitably some of the smaller areas will more often be exclusively made up of suburbs, rural areas, or some other type of neighbourhood which arguably is little more than a component part of the wider urban-centred localities. It is these whole localities which can be most meaningfully compared with each other.

Of course, there is a colossal number of different ways in which 10 529 building-block areas can be combined into approximately 300 groupings. Thus, the numerical similarity between the localities and other sets of areas is potentially misleading: a comparison of figure 8 with figure 1 demonstrates that the locality boundaries are distinct from those of local authorities, even though the latter had formed part of the input to the synthetic data. It is not appropriate here to dwell on a description of the localities' boundaries, so only a few key points will be made by way of illustration.

(a) Unlike functional regions and TTWAs, the localities provide a breakdown of Greater London (and one which, unlike the district boundaries, is consistent with the treatment of the other conurbations).

(b) Unlike districts and TTWAs, the localities all include at least one identifiable urban centre.

(c) Unlike all other sets of boundaries, the localities are defined by reference to the latest census evidence of both commuting and migration patterns, and *also* to a range of other evidence including the unique journey-to-service information of the banking zones.

Future opportunities

At this stage, it seems clear that some new lines of research have been opened up by this work. The more obvious focus is on the new locality definitions themselves; not only their potential value for a diversity of social science enquiries but also their characteristics as geographical objects of study in their own right. Examples of research questions raised by these definitions—and their basis in the innovation of synthetic data—include the following lines of enquiry, among many possible future developments.

Which of the input sets of boundaries was most influential in shaping the locality boundaries which emerged here?

How would the synthetic data methodology cope with adding in noncontiguous classifications such as geodemographic profiling systems?

How sensitive are the results obtained from the synthetic data to this particular mix of input regionalisations (for example, what would be the effect of including some very much larger groupings such as standard regions)?

In what other ways might the synthetic dataset be analysed, given its distinctive features such as the fact that it is a symmetrical matrix?

Further research along these lines can indicate how far the innovation of synthetic data itself has wider potential as a form of spatial analysis for, in particular, visualising spatial association. At the very least, this research can be seen to have demonstrated one way in which regionalisation can break out of the rigid confines of relying upon a single analysis, and the reductionism of using a single information source as a proxy for all the other aspects of localities. The recent resurgent social science interest in 'the local' centres on the intimate linkage between economic, social, and cultural processes (Amin and Graham, 1997) so it is all the more important for regionalisation simultaneously to draw on information about many facets of local geography. It is hoped that the development of synthetic data has, in this way, provided locality boundaries which approach the objective outlined by Claval (1987) as a set of boundaries which were "significant in relation to a whole series of criteria at the same time" (page 167).

Acknowledgements. This paper draws upon the Localities and City Regions research project sponsored by ESRC (project H507255129). The results, which are reported here to illustrate the argument of the paper, could not have been created without the inputs of the other members of the Localities and City Regions research team: Colin Wymer (School of Architecture, Planning and Landscape) and David Atkins (formerly of the Geography Department) in Newcastle University, and Stan Openshaw (Leeds University's Centre for Computational Geography). CURDS colleagues contributing were Simon Raybould who produced the maps, Ann Rooke who produced figure 6, and Amanda Stonehouse who prepared the text and table. Thanks are also due to the journal's editorial team for supporting resubmission of the paper after a protracted delay.

References

Amin A, Graham S, 1997, "The ordinary city" *Transactions of the Institute of British Geographers, New Series* **22** 411 – 429
Atkins D J, Charlton M E, Dorling D F L, Wymer C, 1993, "Connecting the 1981 and 1991 Censuses", RR 93/9, North East Regional Research Centre, Centre for Urban and Regional Development Studies, University of Newcastle upon Tyne, Newcastle upon Tyne
Casado-Diáz J M, 1996, "Mercados laborales locales: análisis preliminar del caso valenciano" [Local labour markets: preliminary analysis of the Valencian case] *Revista de Estudios Regionales* **45** 129 – 155
Claval P, 1987, "The region as a geographical, economic and cultural concept" *International Social Science Journal* **39** 159 – 172
Cole G D H, 1921 *The Future of Local Government* (Cassell, London)
Cooke P H (Ed.), 1986 *Global Restructuring, Local Response* (ESRC, London)
Coombes M G, Dixon J S, Goddard J B, Openshaw S, Taylor P J, 1982, "Functional regions for the population census of Britain", in *Geography and the Urban Environment (Volume 5)* Eds D T Herbert, R J Johnston (John Wiley, Chichester, Sussex) pp 63 – 112
Coombes M G, Green A E, Openshaw S, 1986, "An efficient algorithm to generate official statistical reporting areas: the case of the 1984 travel-to-work areas revision in Britain" *Journal of the Operational Research Society* **37** 943 – 953
Coombes M G, Atkins D, Wymer C, 1996, "Workplace and travel to work patterns", in *Urban Trends in England: Latest Evaluation from the 1991 Census* Eds D Atkins, T Champion, M G Coombes, D Dorling, R Woodward (The Stationery Office, London) pp 123 – 144
Documentation Francaise, 1996 *Pays, Nouveau Territoire du Developpement Local* (DATAR, Paris)
Eurostat, 1992, "Study on employment zones", report E/LOC/20 (Eurostat, Luxembourg)
Flowerdew R, Boyle P, 1992, "Migration trends for the West Midlands: suburbanisation, counterurbanisation or rural depopulation?", in *Migration Processes and Patterns (Volume 2): Population Redistribution in the United Kingdom* Eds J Stillwell, P Rees, P Boden (Belhaven, London) pp 144 – 161
Frey W, Speare A, 1995, "Metropolitan areas as functional communities", in *Metropolitan and Nonmetropolitan Areas: New Approaches to Geographical Definition* Eds D C Dahmann, J D Fitzsimmons, WP-12, Population Division, US Bureau of the Census, Washington, DC, pp 139 – 190
Greenhalgh L, Landry C, Solesbury W, Warpole K, 1997 *The Richness of Cities: Urban Policy in a New Landscape* (Comedia, Stroud, Glos)
Hagood M J, 1943, "Statistical methods for the delineation of regions applied to data on agriculture and population" *Social Forces* **21** 288 – 297
Hamilton F E I, 1969 *Regional Economic Analysis in Britain and the Commonwealth* (Weidenfeld and Nicolson, London)
Henry M, Barkley D, Brooks K, 1995, "RAS: a regional analysis system integrated with Arc/Info" *Computers, Environment and Urban Systems* **18** 37 – 56
Hess S W, Weaver J B, 1965, "Nonpartisan political redistricting by computer" *Operations Research* **13** 998 – 1006
Horn M E T, 1995, "Solution techniques for large regional partitioning problems" *Geographical Analysis* **27** 230 – 248
Johnston R J, 1991 *A Question of Place: Exploring the Practice of Human Geography* (Blackwell, Oxford)
Lolonis P, Armstrong M P, 1993, "Location-allocation models as decision aids in delineating administrative regions" *Computers, Environment and Urban Systems* **17** 153 – 174
Maclennan D, Bannister J, 1995, "Housing research: making the connections" *Urban Studies* **32** 1581 – 1585
Martin D, 1997, "From enumeration districts to output areas: experiments in the automated creation of a census output geography" *Population Trends* **88** 36 – 42
Masser I, Brown P J B, 1975, "Hierarchical aggregation procedures for interaction data" *Environment and Planning A* **7** 509 – 523
Masser I, Scheurwater J, 1980, "Functional regionalisation of spatial interaction data: an evaluation of some suggested strategies" *Environment and Planning A* **12** 1357 – 1382
Massey D, 1991, "The political place of locality studies" *Environment and Planning A* **23** 267 – 281
Massey D, 1995, "The conceptualisation of place", in *A Place in the World? Places, Cultures and Globalisation* Eds D Massey, P Jess (Oxford University Press, Oxford) pp 45 – 85

Nystuen J D, Dacey M F, 1961, "A graph theory interpretation of nodal regions" *Regional Science Association Papers and Proceedings* **7** 29 – 42
ONS, Coombes M G, 1998, "1991-based travel-to-work areas", Office for National Statistics, London
Openshaw S, Rao L, 1995, "Algorithms for reengineering 1991 Census geography" *Environment and Planning A* **27** 425 – 446
Openshaw S, Taylor P J, 1981, "The modifiable areal unit problem", in *Quantitative Geography* Eds N Wrigley, R J Bennett (Routledge, London) pp 60 – 70
Owen D W, Green A E, Coombes M G, 1986, "Using the social and economic data on the BBC Domesday interactive videodisc" *Transactions of the Institute of British Geographers, New Series* **11** 305 – 314
Pred A, 1984, "Place as historically contingent process: structuration and the time – geography of becoming places" *Annals of the Association of American Geographers* **74** 279 – 297
Rees P, Martin D, Williamson P (Eds), 2000 *The Census Data System* (The Stationery Office, London)
Sforzi F, Openshaw S, Wymer C, 1997, "Le procedura di identificazione dei sistemi locali del lavoro" [The procedure to identify local labour market areas], in *I Sistemi Locali del Lavoro 1991* Ed. F Sforzi (ISTAT, Rome) pp 235 – 242
Simpson W, 1992 *Urban Structure and the Labour Market* (Clarendon Press, Oxford)
Slater P B, 1976, "A hierarchical regionalisation of Japanese prefectures using 1972 interprefectural migration flows" *Regional Studies* **10** 123 – 132
Smart M, 1974, "Labour market areas: uses and definitions" *Progress in Planning* **2** 239 – 353
Sokal R R, Sneath P H A, 1963 *Principles of Numerical Taxonomy* (Freeman, San Francisco, CA)
Spence N A, Taylor P J, 1970, "Quantitative methods in regional taxonomy" *Progress in Geography* **2** 1 – 64
Spotila J T, 1999, "Recommendations from the Metropolitan Area Standards Review Committee to the Office of Management and Budget concerning changes to the standards for defining metropolitan areas" *Federal Register* **64** 56628 – 56644, 20 October
Thrift N J, 1983, "On the determination of social action in space and time" *Environment and Planning D: Society and Space* **1** 23 – 57
Upton C, Widdowson J D A, 1996 *An Atlas of English Dialects* (Oxford University Press, Oxford)