**Algorithms for reengineering 1991 Census geography**

S Openshaw, L Rao1f

School of Geography, University of Leeds, Leeds LS2 9JT, England Received 22 April 1994; in revised form 6 October 1994

**Abstract.** The availability of GIS technology and digital boundaries of census output areas now makes it possible for users to design their own census geography. Three algorithms are described that can be used for this purpose. An Arc/Info implementation is briefly outlined and case studies presented to demonstrate some of the results of explicitly designing zoning systems for use with 1991 Census data.

1. **Background**

In the United Kingdom, spatial census data from the 1991 Census are reported for a standard set of geographical areas (enumeration districts-EDs, wards, districts, counties, regions, countries) that nest into each other (see Denham, 1993). However, for the first time a complete set of digital boundary data is available for the smallest geographic units (census EDs and output areas in Scotland) for which census statistics are reported. Digital boundaries suitable for use with census data were previously only available for much larger geographical units; for example, districts in the 1970s; then wards and postcode sectors in the late 1980s. It is now easy through geographic information systems (GISs) to take zonal digital boundaries and reaggregate them to whatever higher level of standard census geography is considered relevant; for example, from EDs to wards to districts to counties (see Charlton et al, 1995). So the 1991 small area boundary data files present the first real opportunity that census users have had to build their own census geographies from the smallest census building blocks and thus escape from the tyranny of an arbitrary imposed and fixed set of census geographies. They can now design or engineer whatever spatial zoning systems are considered most suitable for a given purpose, using spatial building blocks which are probably small enough to make the exercise worthwhile. The principal difficulty now is the absence of zone-design tools in GIS and the need to have an explicit rationale for the zone-design process considered best for any specific census application.

The justification·for analysing census variables for fixed sets of ward or district units was historical and based mainly on convenience and feasibility rather than because wards or districts were considered to be the best or most meaningful geographical objects to study. Quite often they were the only objects that could be analysed. For some purposes, these are sensible geographical objects; for example, districts are obviously useful for studying processes that reflect the functioning of local authorities. However, the use of districts or wards to study geographical patterns in many census variables, such as unemployment or car ownership, is much more prob­ lematical. In the former, unemployment reflects the spatial organisation of the local economy which may be independent of local government boundaries. In the latter case, the underlying processes probably operate on a much more local scale, perhaps at a neighbourhood level, and may reflect the operation of patterns and processes for

11 Present address: GIS/Trans Ltd, 675 Massachusetts Avenue, Cambridge, MA02140, USA

which districts and wards need not be sensible geographies. Likewise few would claim that attempts to create EDs that roughly equalise census-enumerator workloads is a relevant zone-design criterion for the spatial analysis of census data. The resultant census geography is arbitrary, highly nonrandom, and provides a variable level of representational accuracy for the households and people it contains. There is a risk of obtaining biased or misleading results when data for possibly inappropriate areal units are studied. Once there was little or no choice; now there is.

This uncertainty as to which zones to use leads to the well-known modifiable areal unit problem (MAUP) (Kendall and Yule, 1950; Openshaw, 1977a; 1977b; 1978a; 1978b; 1984; Openshaw and Taylor, 1979; 1981). Indeed, the MAUP becomes of even greater practical significance once the user is freed from having to use fixed, usually 'official', sets of areas for geographic analysis (Openshaw, 1989). The problem now is that of how best to exploit the new-found freedom of user-controlled zone design and flexible higher scale geographies by developing sensible and appropriate zoning systems for particular purposes. It is noted that the recent rediscovery that different zoning system representations of the same data may or may not provide differ­ ent results and, that this or that statistic may be observed to vary for this or that arbitrary aggregation, is no longer surprising although still useful as a means of improving end-user awareness (Amrhein, 1992; Amrhein and Flowerdew, 1992; Dudley, 1991; Fotheringham and Wong, 1991). The principal value of such studies is to provide independent confirmation of the reality of the problem but without offering much prospect of a solution, with the limited exception of Arbia (1989).

Another closely associated geographical issue is the spatial representation problem. There is an implicit assumption in mapping census data that the spatial variations in colour across the map are real and not largely an artefact of the mapping and data­ aggregation process. Monmonier (1991) points out how to adjust or fiddle the message being communicated by the map through careful selection of choropleth class intervals. Although interesting, this still ignores aspects of the wider spatial representation issue. The map is reporting in a cartographic form values for areal objects that need not be the same everywhere on the map and, indeed, seldom are. Census EDs, wards, and district entities vary in physical size, and they also vary in terms of the numbers of people they contain. This is well known. More important, however, is the observation that they also vary in the extent to which they represent one specific geographic object (a piece of space or zone) that is supposedly compa­ rable with other objects of the same logical type. The problem in geographical analysis is that data are reported for objects that are not basically the same; they are not strictly speaking comparable entities; and they have often no natural signifi­ cance or innate meaningfulness in a geographical sense. There are also problems of how best to represent geographically discontinuous phenomena by zonal geography. Select any two census EDs, have a close look at them and in particular at the areas and communities (or parts of community) they overlie and presumably try to represent. The key question is whether or not the ED boundaries make any sense in relation to the microspatial patterns and processes relevant to those small pieces of the earth's surface they represent. The problem is that they seldom do. As Morphet (1993) argues "the only significant ed boundaries are those to be found between differ­ ent but homogenous eds" (page 1274) but then observed that "out of many thousand ed boundaries in the 1981 census map of Newcastle, only two, amounting to less than 200 m of boundary length, are significantly located at the boundary of areas of different social composition" (page 1276). Openshaw and Gillard (1978) in an even earlier study of boundary location in census classification noted that in their study

region the only stable physical boundaries were located along railway lines.

From a geographical perspective most census areas are not natural nor comparable areal entities. A ward or ED located in the middle of Birmingham can be a very different geographic entity from a ward or ED located in a small market town or in a rural area. These zonal geographies are not everywhere sensitive to the under­ lying sociospatial patterns they represent. An ED in a large town might be entirely composed of council housing, whereas in a small town the council estates are smaller and the equivalent sized ED now has a mix of different housing types. So, although census EDs do have some broad similarities: for instance, they have the same areal label (that is, census EDs) and they have been created by the same census geography planning process; there are also many differences. They vary in physical size; they vary in population size; they vary in terms of their internal socioeconomic, demographic, and morphological heterogeneity, because they vary in terms of their ability to represent the underlying microdata. They are as like as chalk and cheese when viewed from a geographical perspective. They present one small area-aggregated view of census data that varies enormously from one ED to another in terms of the amount of scale and aggregational noise, bias, and distortion they have added to the data. Of course all data-aggregation procedures cause information loss and generalisation, but the problem with census data is that the nature and extent of the loss of information are not consistent from one ED to the next, and they are also uncontrolled.

It is now much too late for the census agencies to redefine the 1991 Census EDs in order to provide a more consistent spatial representation of the census micro­ data; although this could be done in 2001. The only remedial action open to the census user is to hope that by reaggregating some or all of the ED data some of the worst problems can be avoided. The 1991 Census presents the first real chance users have had to reengineer their census geographies to suit their needs better rather than have to rely on a small number of standard but arbitrary census geographies. The challenge is how best. to do this efficiently and optimally. However, it is also important to be constructive. The new opportunity provided by the availability of 1991 Census ED boundaries is not to demonstrate the universality of MAUP effects or to manipulate results by gerrymandering the spatial aggregation used but it is to design new zoning systems that may help users recover from the MAUP. It is with this objective in mind that in section 2 we describe three algorithms designed to allow the census user to reaggregate the 1991 small area census data into large areal entities that are less arbitrary in terms of a specific purpose or application. In section 3 we briefly outline the coupling of these methods to the Arc/Info GIS (ESRI Inc., Redlands, CA). In section 4 we briefly present some examples of use and in section 5 discuss possible future developments.

1. **Unconstrained zone-design methods**
   1. **Designing zoning systems**

The task here is to start with data at one scale and then reaggregate them to create a new set of regions designed to be more suitable for a specific purpose than the original building blocks. In a census context, it is to be hoped that the loss of information through the additional spatial aggregation can be offset by developing zonations that, although cruder than the original building blocks, are in some specific ways judged to be more useful. Clearly, this will not always be appropriate and given the size of the ED building blocks from the 1991 Census it restricts the zonal customisation to mesoscale geographical units. Ideally, the technology needs to be applied to much smaller spatial units than census EDs.

The zone-design process involves the aggregation of spatial data for *N* zones into *M* regions, where *M* is less than *N* and usually a small fraction of *N* in size. The *M* output regions should be formed of internally connected, contiguous, zones. Provided *N* is not too small, and *M* is not large in relation to *N;* there are many, many, billions of different ways of aggregating *N* areas into *M* regions. The asso­ ciated scale problem is much less binding and involves the selection of a suitable value for *M.* There are a maximum of *N* - 2 choices and it is clear that the degree of scale-induced variability will in general be considerably less than that caused by aggregational variability. Tobler and Moellering (1972) suggest that higher levels of generalisation (that is, small values of M} might provide more explanation than larger values about the variance of a spatial variable. However, in general it would seem that scale is something that only needs to be broadly right, whereas the aggregation needs to be optimised. These problems can be avoided only if accept­ able frame-independent spatial analysis and modelling procedures can be developed, see Tobler (1989}. The problem is that for census analysis frame-independent methods might not make much geographical sense and do not yet exist.

The earliest zone-design procedures might be considered to be the regionalisa­ tion algorithms that were concerned with either the multivariate classification of data subject to a contiguity restriction or the functional regionalisation of flow data. Electoral redistricting might be considered another variant of the problem as indeed is location-allocation partitioning (see Goodchild, 1979}. Openshaw (1978b; 1984) argued that these are all special cases of a more general zone design problem that was the geographer's equivalent to a form of nonlinear optimisation problem in which the parameter to be optimised was the classification of *N* areas into *M* regions.

This can be stated as the following combinatorial optimisation problem

optimise F(Z), (1)

subject to constraints on *Z* to ensure the *M* regions are internally connected, F(Z} can be any function defined on data for the *M* regions in *Z,* and *Z* is the allocation of each of *N* zones to one of *M* regions such that each zone is assigned to only one region.

The original automatic zoning procedure (AZP} of Openshaw (1977a; 1978a; 1978b} used a heuristic procedure that evolved by trial and error from an iterative relocation method commonly used in numerical taxonomy. Subsequent work extended the AZP to allow additional constraints to be added; for instance, on the configura­ tion of the *M* regions and/or on the nature of the aggregated data being generated by the zoning system. In principle, once you can solve the unconstrained zone-design problem the imposition of various other side conditions in the form of equality and inequality constraints can be handled through penalty function methods. Openshaw (1978a; 1978b} gives examples of the fit of a linear regression being optimised subject to constraints that the residuals should have zero spatial autocorrelation and be homoscedastic. The greatest practical difficulty is the choice of objective func­ tion, and that of deciding what constraints can be sensibly imposed on the data. The zoning system acts as a pattern detector and the nature of the patterns it detects depends on the function being optimised, and how this interacts with the data being represented by the set of *M* regions, as well as any additional constraints imposed on the zoning system being estimated. In geographical analysis, zone design, result, and pattern all interact, and there is a risk of causing considerable damage to unknown spatial patterns by accident. Virtually all aspects of these interactions are at present only dimly understood.

* 1. **AZP, a mildly steepest descent algy,ri -** . . . . .· ..

The original Openshaw (1977at'f97ifa(1978b}'A.z:P 'algorithm is a fairly simple local boundary optimiser. It involves the following:

*Step 1* Start by generating a random zoning system of *N* small zones into *M*

regions, *M* < *N.*

*Step 2* Make a list of the *M* regions.

*Step 3* Select and remove any region *K* at random from this list.

*Step 4* Identify a set of zones bordering on members of region *K* that could be moved into region *K* without destroying the internal contiguity of the donor region(s).

*Step 5* Randomly select zones from this list until either there is a local improvement in the current value of the objective function *or* a move that is equivalendy as good as the current best. Then make the move, update the list of candidate zones, and return to step 4 or else repeat step *5* until the list is exhausted.

*Step 6* When the list for region *K* is exhausted return to step 3, select another region, and repeat steps 4- 6.

*Step* 7 Repeat steps 2 - 6 until no further improving moves are made.

This AZP algorithm can work with any type of objective function that is sensi­ tive to the aggregation of data for *N* zones into *M* regions that is being evaluated in step 5. Examples include functions computed directly from the data (for instance, the sum of squared deviations from average zone size), and also functions which represent the goodness of fit of a model applied to the data (for example, the fit of a linear regression model or the performance of a spatial interaction model). A number of studies have used this AZP algorithm seemingly quite successfully (for example, Openshaw, 1977a; 1977b; 1978a; Openshaw and Taylor, 1979). However, this work was restricted to fairly small data sets. Openshaw (1977a) used 1219 100- metre grid squares, but this might be regarded as an 'easy' set of building blocks to rezone because of their Queen's case contiguity regularity. Most of the other uses of the algorithm were on the grid-square-like 99 Iowa counties, aggregating to various numbers of regions down to 6 regions; or 73-zone spatial interaction journey-to­ work data for Durham County (aggregating to 22 regions); and some unpublished political redistricting involving small numbers of zones and regions in Scotland.

An ability to be able to optimise any function is important because the under­ lying purpose in designing the AZP was threefold: first, to identify the approximate limits of aggregation effects on any statistic, or model, by the simple expedient of first maximising and then minimising its value by systematically manipulating the zoning system; second, to demonstrate that the MAUP is endemic and affects all the results of studying spatially aggregated data and that, despite claims to the contrary, there are probably no geographically meaningful exceptions in the form of zoning-system-invariant spatial models. Third, and of greater longer term signifi­ cance, it provides a basis for developing a new approach to the aggregation problem based on two observations. First, the estimation of the zoning system is the geographical equivalent to the estimation of parameters in a model. Second it is apparent that the zoning system that best fits a particular statistic or model provides a map-based visualisation of the interaction between data, model or statistic, and aggregation. It was conjectured that this might provide useful insights about spatial pattern and process, although until GIS existed it was virtually impossible to test on realistic data.

* 1. **Problems with the AZP algorithm**

The AZP algorithm is not in the steepest descent direction because the search is local to a selected region. This is deliberate. A steepest descent version would scan

the global list of all possible single moves that could be made for all *M* regions and then select the best. Surprisingly, this is much less effective than the basic **AZP** algorithm, and it is also considerably more expensive because of the much greater number of different zoning systems that would have to be evaluated before a move could be made. More serious is the observation that the steepest descent algorithm often gets stuck, whereas the local optimiser follows a much more gradual route towards a local optimum and this seems to be much better suited to zone-design problems. However, the AZP method can become stuck, although the usual solution of rerunning the algorithm from a number of different randomly generated starting zonings and then selecting the best often works reasonably well.

A major problem in the pre-GIS 1970s, was the need to create the contiguity lists by hand and this tended to restrict the size of the data set that could be experimented with. The grid-square data were an exception because of the ease by which the contiguities could be generated automatically. Another difficulty con­ cerned mapping the *M* regions that were produced at a time when the cartography was manual. GIS has removed both problems. It is observed, however, that the performance of the simple AZP algorithm seems to diminish as the N-into-M aggregation problem becomes harder both because of increasing *N* and also because of irregularities in zone size and shape; for example, compare census ED topology with that of districts. The availability of larger data sets causes other difficulties. The number of alternative ways of aggregating *N* zones into *M* regions explodes combinatorially as *N* increases. This greater aggregational freedom helps when *N* is not too big but then as *N* continues to increase it starts to reduce the efficiency of the algorithm as the number of local suboptima rapidly increases. This combina­ torial complexity and hardness of the optimisation task appear to interact with the nature of the function being optimised.

* 1. **A simulated annealing variant of AZP**

The principal problem with AZP is that of being trapped by a local suboptimum which is not very good and which is dependent on the starting zoning system. There are ways of reducing this problem; for example, by extending the search process to consider two, three, or more moves at a time in the search for a local improvement. This greatly increases compute times, although it has been used successfully to overcome the limitation of the basic AZP's single-move heuristic in redistricting applications. The real question now with much faster computer hard­ ware is whether or not a different heuristic might provide much better results.

A very widely used method in combinatorial optimisation is that of simulated annealing (see Aarts and Korst, 1989). Simulated annealing sometimes called the Metropolis technique, was originally developed to solve a hard optimisation prob­ lems in physics (Metropolis et al, 1953). It has subsequently developed into a global optimisation method particularly suitable for problems with multiple suboptima (Kirkpatrick et al, 1983). There have so far been only a few geographical applications of it (for example, see Openshaw, 1988). It is essentially Monte Carlo optimisation procedure that can escape from local suboptima. It provides a robust optimisation for problems which are otherwise hard to solve. It is therefore inherently attractive as a potential AZP solution method.

The basic simulated annealing AZP method (AZP-SA) is simple to apply. The local search step 5 in AZP algorithm only need be modified as follows:

*Step 5* Randomly sample this list until there is a local improvement in the objective function or an equivalently good move. Then make the move. Otherwise make the

move with a probability given Boltzmann's equation,

*R(0,* 1) < exp ( ; )), (2)

where

*Vf* is the change in objective function caused by the move,

*T(k)* is the 'temperature' being applied at annealing time step *k,*

*R(0,* 1) is a uniformly distributed random number in the range 0.0 to 1.0.

The hardest part is determining a suitable slow-cooling schedule for the annealing process with the temperature being lowered in a sufficiently large number of steps to ensure that a very good result has had sufficient time to emerge. There are some proofs that establish the global optimality of the simulated annealing method but these require a very gradual annealing process such that

*T(k)* = ). (3)

This is much slower than most algorithms which typically use an exponentially decreasing- annealing schedule, namely

*T(k)* = *fT(k-1),* (4)

where /is typically between 0.8 to 0.95 (for example, see Press et al, 1986).

The power of the simulated annealing algorithm is that it permits moves which result in a worse value of the objective function but with a probability that dimin­ ishes gradually, through iteration time. This means that the basic AZP algorithm now has to be embedded in an iteration loop that slowly reduces *T* from some arbitrarily high initial value. The initial value of *T(0)* would be set such that about half of all moves would be accepted. It is important to ensure that the zoning system is properly 'melted' before applying an annealing and cooling schedule to it. *Step a* Set *T(0), k* = 0.

*Step b* Apply AZP with the modified step 5 until either MAXIT (a user-defined maximum number of) iterations or convergence or at least a minimum of *Q* simulated annealing moves have been made.

*Step* c Update *T* and *k*

*T(k)* = *0.85T(k-1)*

*k* = *k+l.*

*Step d* Repeat steps b and c until no further moves occur over at least three different *k* values.

The local serial-move heuristic used in AZP is still useful with waves of change rippling through the zoning system. Experience demonstrates that this method produces what look like extremely good results. As computer hardware becomes faster so the about two to four orders of magnitude increase in compute times over the basic AZP method will become insignificant.

* 1. **A tabu search heuristic**

Another relatively new heuristic that can be applied to zone-design problems is tabu search (see Glover, 1977; 1986.). This started out as an integer programming method but is now regarded as being suitable for a wide class of combinatorial optimisation problems which are NP hard (Bland and Dawson, 1991; Glover and Llaguna, 1992; Hertz and de Werra, 1987). The AZP is certainly of this form and any heuristic that is capable of generating near-optimal solutions to hard combina­ torial problems which may have many different local optima is of potential interest here. Simulated annealing certainly meets these needs, but it is computationally

very expensive. The hope is expressed that tabu search might be quicker and equivalently good.

A basic tabu version of the AZP is specified as follows

*Step 1* Find the global best move that is not prohibited or tabu.

*Step 2* Make this move if it is an improvement or equivalent in value, else:

*Step 3* If no improving move can be made, then see if a tabu move can be made which improves on the current local best (termed an aspiration move), else:

*Step 4* If there is no improving move and no aspirational move, then make the best move even if it is nonimproving (that is, results in a worse value of the objective function).

*Step 5* Tabu the reverse move for *R* iterations.

*Step 6* Return to step 1.

The tabu algorithm is a powerful optimisation tool that allows the search process to escape from local optima whilst avoiding cyclical behaviour. The only real requirement is that the list of possible moves is even in its coverage of the solution space, which it clearly is in the zone-design context. The key algorithmic parameter is the length of tabu period, *R.* If a zone *I* is moved from region *K* to region *L,* then the reverse move is prohibited for *R* subsequent iterations. Clearly the value of *R* should not equal or exceed the number of possible moves. The purpose of tabu-ing reverse moves is to force the algorithm to look elsewhere for new or better results, even if it means making nonimproving moves. This results in a very aggres­ sive search process that is constantly trying to discover better results without repeating itself too much.

The principal practical problem with AZP-tabu search is defining a suitable

tabu length for *R.* The results can critically depend on specifying a good value for *R* but seemingly there is no easy or good way to do this. It also seems to be problem specific. One solution that works well with zone-design problems is simply to rerun the method repeatedly starting with a small *R* value and then steadily increasing it. Each tabu suboptimisation problem is then run to completion. This works very well especially when the best result is selected from a small number of different random starts.

A more pleasing solution to specifying a good *R* value is to adapt the reactive tabu search of Battiti and Tecciolli (1993; 1994) to AZP problems. This algorithm provides a more efficient exploration that becomes even more powerful the harder the problem becomes. It involves the following procedure.

*Step 1* Start with a random zoning system.

*Step 2* Set the length of prohibition (R = 1) and the average number of iterations between repetitions (r•v = 1.0).

*Step 3* Define the list of all possible moves that are not tabu and retain regional con­ nectivity.

*Step 4* Find the best nontabu move.

*Step 5* Make the move. Update the tabu status.

*Step 6* Look up the current zoning system in a list of all zoning systems visited so far during the search. If not found then go to step 10.

*Step* 7 If it is found and it has been visited more than K1 times already and this cyclical behaviour has been found on at least *K2* other occasions (involving other zones), then go to step 11.

*Step 8* Update a moving average of the repetition interval ,•v, and increase the prohibition period *R* to *l.lR.*

*Step 9* If the number of iterations since *R* was last changed exceeds *r"V,* then decrease *R* to max(0.9R, 1).

*Step 10* Save the zoning system and go to step 12.

*Step 11* Delete all stored zoning systems and make *P* random moves, *P* = l +½ *r•v,*

and update tabu to preclude a return to the previous state.

*Step 12* Repeat steps 3 - 11 until either no further improvements are made or a maximum number of iterations are exceeded.

The reactive tabu dynamically adjusts the prohibition period *R* as the search proceeds. There is a memory mechanism that is used to prevent repeat results. If the same zoning system is repeated more than a small number of time (K1 = 3) by more than a small number of different regionalisations (K2 = 3} then a series of random moves are made and *R* is increased. The random moves are designed to escape from the neighbourhood of the current solution and the tabu is used to pro­ hibit an early return there. The avoidance of cycles is important in order to ensure that the available time is spent in an efficient exploration of the search space. The principal problem now is the potentially large number of iterations that may be needed.

* 1. **Parallel algorithms**

The prospect of teraflop computing speeds on terabyte memory hardware during the late 1990s, makes it extremely worthwhile to consider developing parallel zone­ design algorithms to allow larger numbers of zones to be handled or more complex zonal optimisation problems. Clearly the basic AZP algorithm as described in section 2.2 is not suitable for parallel implementation, although, depending on the objective function being used, there may well be some local vectorisation opportu­ nities. The AZP is not a parallel algorithm because of its local sequential search for improvements in the objective function. It could be transformed into one only by performing the local search in parallel on different parts of the map which are not contiguous, provided the objective function was decomposable so that the contribu­ tion of any of the *M* regions to it could be examined and changed independently of all others. Surprisingly, this is often the case and the benefits of this form of parallel boundary optimisation increases with the size of *N.* Developing parallel versions of these algorithms is a matter of some priority as compute speeds (not data) are becoming the principal barrier to the design of census zones.

1. **Developing a zone design system (ZDES)**
   1. **Benefits of developing an Arc/Info interface**

The algorithms described in section 2 are not standalone software. The AZP methods require contiguity information as inputs and then some means of mapping the resulting optimal zonations. GIS provides an ideal environment within which to perform these functions and is a natural way of disseminating zone design tools in a form that GIS literate users can appreciate. In this section we describe the develop­ ment of a system called ZDES (zone design system}.

The choice of GIS often seems to present major philosophical problems. Here Arc/Info is used because it is a de facto worldwide standard, it operates on many different computer platforms, it has a large number of research users in Britain and many other countries, and there is nothing to preclude the development of inter­ faces for other GISs if needed later. The AZP software is written in FORTRAN 77 and is GIS independent; only the interface is Arc/Info specific. The objective is to develop a loosely coupled zone-design system for Arc/Info within which the link­ ages between the GIS and AZP functions are completely transparent to the user and which would also permit the AZP optimisation to take place on a different machine from that with the GIS on it. What follows assumes Arc/Info version 6.1.1 (or higher) operated on UNIX workstations. Arc/Info provides a set of macro

programming facilities (AML) for linking functions within and outside Arc/Info and also allows the users to build the menu-driven applications. Under OpenLook or Motif, sophisticated graphic user interfaces can be easily created by using Arc/Info's THREAD - MENU - POPUP tools.

* 1. **Generating contiguity tables**

AZP requires a topologically 'cleaned' data structure from which logically consistent interzone contiguities can be generated. In the Arc/Info data model, the rela­ tionships between spatial objects should be identifiable in a coverage. This provides an easy means of generating the contiguity table from any input zoning system or polygon coverage. In fact there are two Arc/Info methods for obtaining the con­ tiguity matrix. The first is by reading the AAT. In a cleaned polygon coverage, every arc is shared by two polygons, so in its AAT there is a left and a right polygon. From this a contiguity table can be produced. The second method is by spatial query. In ARCPLOT, the RESELECT command performs a spatial query for identifying the ADJACENT polygons of any selected polygon. The query could also have a buffer distance. By changing this distance, the spatial contiguity of zones no longer has to be restricted to strictly physically map-based contiguities. A much more flexible definition of contiguity can be used, one that is able to bridge holes in the physical map topology caused by barriers; for example, a river.

* 1. **Building a ZDES in Arc/Info**

The AZP algorithms described in section 2 are general purpose in that they can approximately optimise any function with a value that depends on spatial data. A number of generally applicable functions have been developed: equal value zoning, correlation, distance function, autocorrelation, space partitioning, similarity.

The most useful, census-relevant, function is probably the equality zoning; in which regions are devised that are of near equal value in terms of a selected variable; for example, population size or numbers of economically active people. This has some similarities with a cartogram except that here the aggregation process reduces the number of original zones, whereas the cartogram is purely a carto­ graphic device. Equal value zoning is useful because it controls for spatial size variation and there are a number of applications for which this is helpful. Cole (1993) notes "The 10 per cent SAS have only been released at ED/OA level to provide users with a primary building block which can be used as a basis for flexible area aggregation" (page 230). Clearly aggregating EDs to produce zones of an equal and sufficient size to yield reliable data would make the 10% census data more reliable.

Optimising a zoning system with respect to a correlation target is of interest mainly to demonstrate aggregation or modifiable areal unit effects. However, it may also have some value as the basis for spatial analysis via mapping zoning systems that match a particular correlation target. This is discussed further in the next section. The distance function option is a further development in this direction in that the AZP procedures attempt to fit a distance decay function to the data with respect to a user-selected central point. Visualising the zoning systems that match such a model may also have some intrinsic geographical value. The spatial auto­ correlation function is of interest because it represents a measure of global map pattern, so zonings that for a specific variable match particular autocorrelation targets may sometimes be of interest. The space-partitioning function is equivalent to a form of location - allocation model. The regions are designed to maximise internal weighted accessibility. The result is a highly efficient partitioning of space.

The similarity-based zoning provides an AZP version of a contiguity-constrained classification or regionalisation of the data in which regions are developed that have a similar level of internal heterogeneity with respect to the variables being used.

Other zone-design aspects are also relevant. In particular, it is not unusual to impose barriers upon the aggregation process; for example, to preserve key linear geographic features or to retain higher level administrative boundaries in the reengineered zoning system. This can be easily introduced via deletion of contiguity links. The hard part is identifying and then deleting the relevant links; and clearly a GIS can help here; for example, buffering and polygon-overlay functions can be used to add barriers to the contiguity information. In interactive zone design, editing the contiguities is an important part of the process.

Finally, once an approximately optimal zoning system is defined a new coverage

needs to be generated that can be used subsequently in other spatial data-handling operations. It is clear then that to be a viable zone-design system the procedures need to be embedded within a GIS as the various separate functions need to be integrated. Table 1 provides a summary of the major steps in this process. If zone­ design methods are to be packaged for more general use, then this complexity needs to be hidden behind suitable user-friendly interfaces.

**Table 1.** Principal steps in zone design.

*Step 1: select zone design options and targets*

Set output zoning number Set zoning function

Set optimisation method Set zoning constraints

Set process parameters (CPU time and iteration number)

*Step 2: create input data* Read initial coverage Generate contiguity matrix

Reselect arc segments (AAT records} Alter contiguity matrix for any barriers Select data to be used in zone design

*Step 3: run automatic zoning procedure to search for best achievable result*

Initial a random starting zoning Optimise

Check results Check performance

Write out the best result

*Step 4: generate new results coverage*

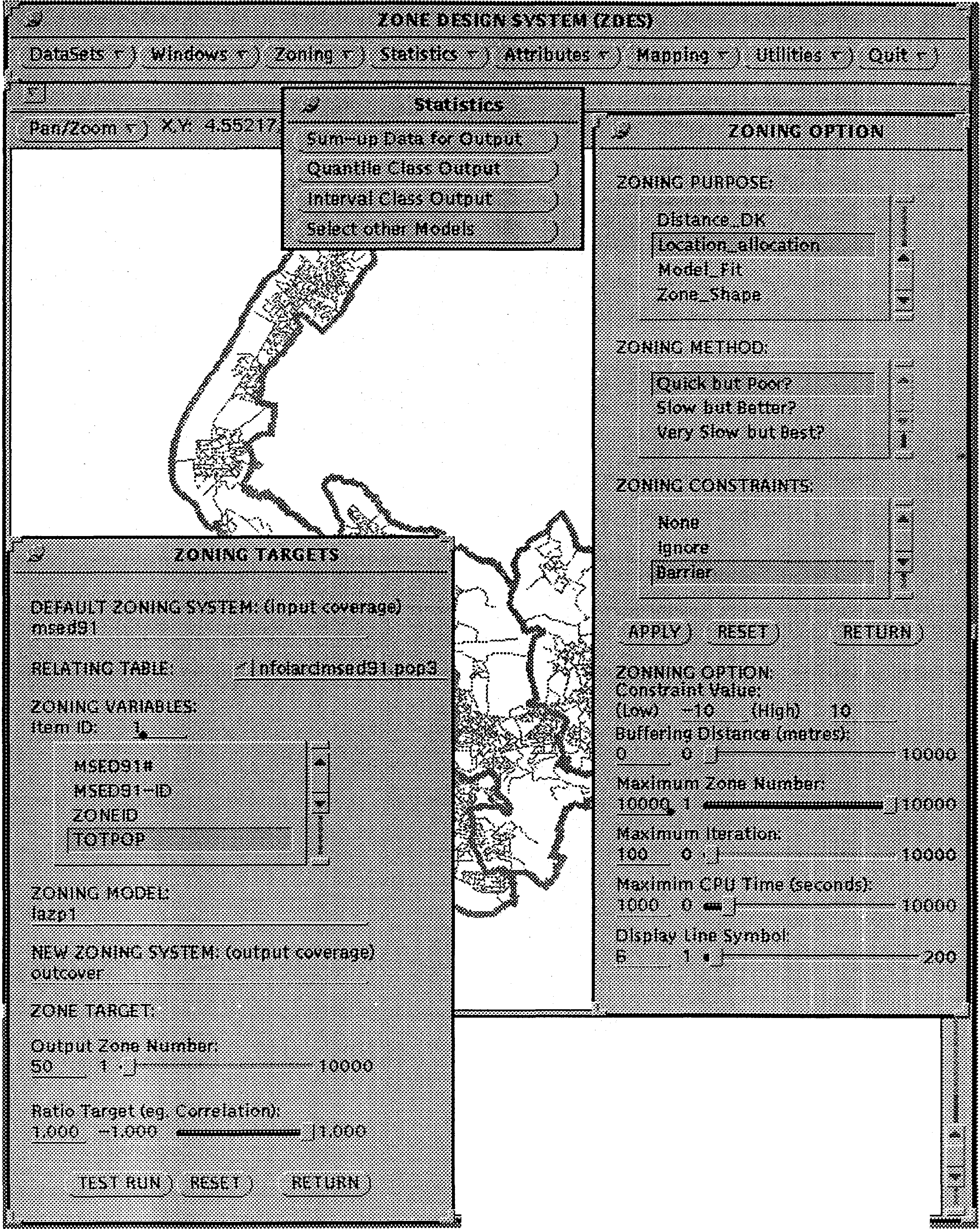
Assign the regionalisation to the input coverage Dissolve old to produce new coverage

Transfer data to new coverage

* 1. **Developing an easy-to-use ZDES command**

Two sets of AML programs have been written in order to make AZP an easy-to-use application in Arc/Info. One is called ZDTools which allows the user to run AZP from a command line; the other is called ZDMenu which provides the user with a menu-driven system that will help to explore the various functions relating to zone design. Figure 1 (see over), provides a screendump of this menu-driven interface to ZDES, outlining some of the choices. The results described in section 4 are produced by this system.

Dissolving msed91 by class to create outcover Creating outcover.PAT format...



Creating dissolve table... Dissol vi ng...

Number of Polygons (Input.Output) - 2926

51

Number of Arcs (Input.Output) 27671 106

Creating outcover.PAT...

**Figure 1.** A screen from the zone-design system.

# Census-data case studies

* 1. **Reengineering census EDs for Merseyside**

Census digital ED boundaries and census data from the 1991 Census for the 2926 EDs in Merseyside are used to illustrate the ZDES system. Initially attention is focused on the distribution of old aged persons (males aged 65 and over and females aged 60 and over) on the grounds that these data are interesting in a substantive sense and they are also a reasonably 'hard' data set to rezone. The hardness results from the sparsity of the data; not all EDs have old people in them

(indeed 113 had none) and thiswill provide a good test of the various zone-design methods. Additionally, the sepkhHion of the Wirral from the rest of Merseyside by the river Mersey splits the 2629 EDs into two superregions or contiguity islands. This is interesting not least because this might be expected to be a common feature with digital census boundaries for small areas; for example, because physical features can cause logical contiguity breaks in census ED geography. There is no easy solution to the problem other than to redefine what is meant by contiguity and then to add 'artificial' linkages to the data to reflect this view.

One immediate consequence of the two contiguity islands is that the initial random zoning-system generator based on Openshaw (1977c) will no longer ensure that each subregion has an appropriate share of the initial *M* regions. The algo­ rithm was modified first to scan the contiguity information to identify the number of internally connected islands within it; second to identify each 'island's' allocation of the *M* regions according to its share of a relevant variable (that is, population); and third, to generate random zoning systems within each island. This might be regarded as spatially stratified random zoning. Another solution to the same problem is to use seed points as 'cores' for the random zoning system. ZDES offers all three. In some ways this problem might also be regarded as a basic spatial representational issue because it concerns how many regions a given island or subregion has to represent it. It reemphasises the risks of applying global spatial analysis and spatial models to census data without stopping to think about its structure; for example, if you were fitting a spatial regression model would you have one or two sets of parameters?

As a test of the relative efficiency and performance of the three AZP algorithms, interest here focused on the rezoning of the census EDs to produce regions that contained about the same numbers of old people.

The objective function is

t

F(Z) = abs( *o;;P;* - ) , (5)

where

***f>;;***

{ 1, if zone *i* is in region *j,*

, 0, otherwise,

*P;* is the number of old people in ED zone *i,*

*Tj* is the target size for region *j.*

Note that considerable compute time can be saved by updating rather than recom­ puting equation (5) every time any zone moves. This also makes the compute time needed to calculate the objective function invariant of *N* and *M;* a highly useful property whenever it can be done; although the total number of function evalua­ tions depend on *N* and *M.*

Table 2 (see over) shows that generally the performance of the basic algorithm is good for small numbers of regions but starts to deteriorate once more than 100 regions are involved. The basic tabu version is better but becomes less reliable with over 150 regions. The reactive tabu would be even better but is unaffordable even with these data because of very long compute times. The simulated annealing results are best throughout. However, the compute times in table 3 (see over), provide a compelling reason why in a workstation environment the poorer results of the basic algorithm might be acceptable in many practical applications. It also reemphasises the need to develop parallel algorithms for zone design.

**Table 2.** Results for equal zone size [equation (5)) zonations.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number | Algorithm• | |  | |
| of regions  required | AZP | AZP-TABU | | **AZP-SA** |
| 10 | 15.6 | 12.1 | 15.2 | |
| 20 | 153.8 | 238.3 | 64.0 | |
| 30 | 232.7 | 128.2 | 158.3 | |
| 40 | 538.8 | 253.2 | 287.9 | |
| 50 | 815.7 | 569.2 | 367.8 | |
| 75 | 1061.4 | 986.1 | 585.4 | |
| 100 | 2239.0 | 1491.8 | 922.6 | |
| 150 | 3800.6 | 4839.8 | 156.7 | |
| 200 | 6860.2 | 10693.9 | 1765.9 | |

* The AZP (automatic zoning procedure) result is the best from starting with 10 different random zonings. The tabu and SA (simulated annealing) results are unaffected by the choice of initial random zoning system.

**Table 3.** Compute times (seconds) for zone-size objective-function rezoning.

Number of regions

Algorithm"

AZP AZP-TABU **AZP-SA**

|  |  |  |  |
| --- | --- | --- | --- |
| 10 | 730 | 5169 | 19575 |
| 20 | 1182 | 3306 | 9763 |
| 30 | 1671 | 2915 | 10679 |
| 40 | 1880 | 12828 | 10156 |
| 50 | 1861 | 3877 | 9903 |
| 75 | 2010 | 5684 | 10493 |
| 100 | 2053 | 5169 | 12579 |
| 150 | 2434 | 12876 | 13887 |
| 200 | 2229 | 7875 | 131766 |

Note: Compute times are for Sun-Supersparc 10 model 40 workstation.

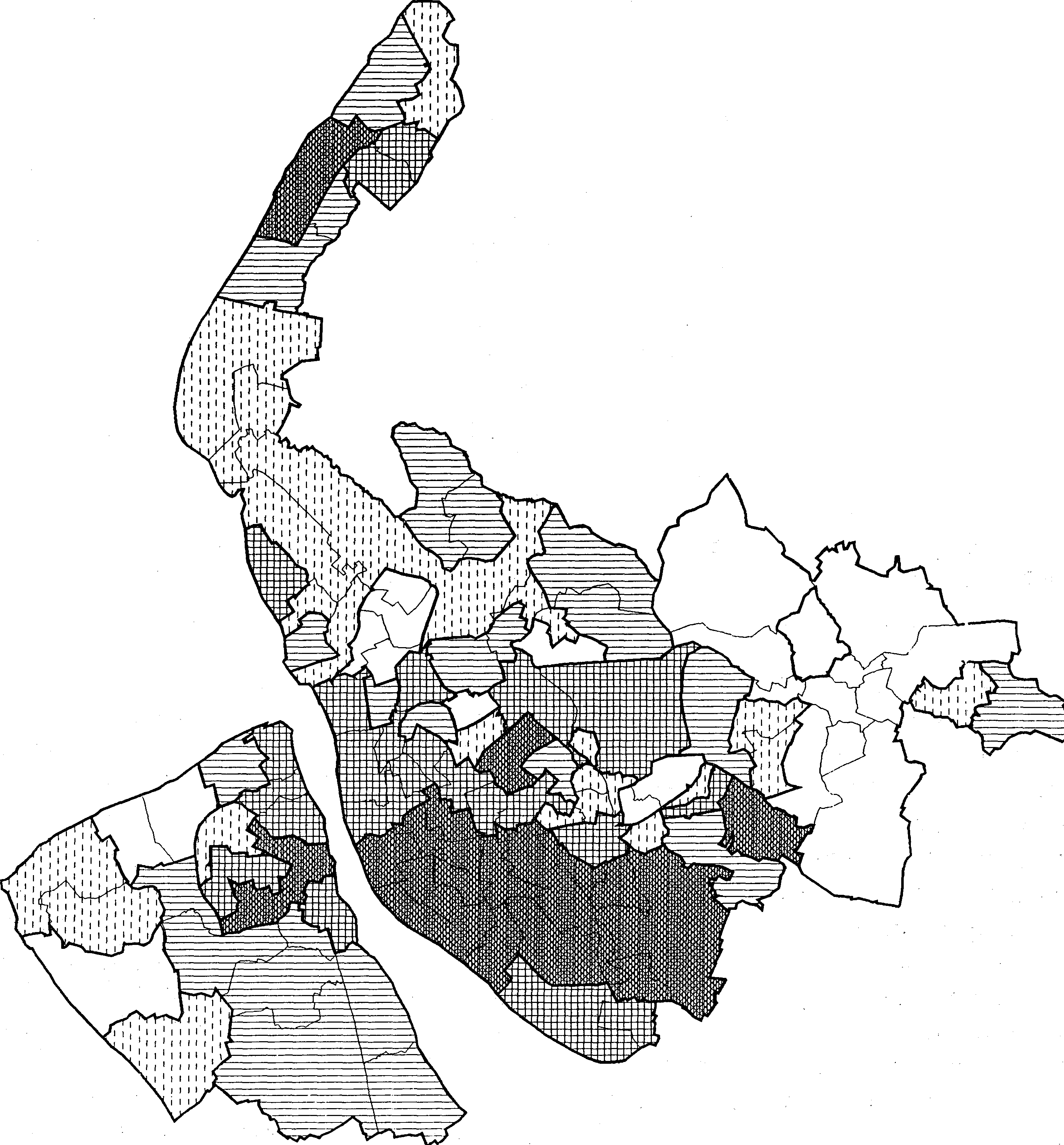
* AZP automatic zoning procedure: SA simulated annealing.
  1. **Using optimal zone design to study ethnic minorities on Merseyside**

A potentially interesting use of the AZP algorithms is to provide a new approach to the map-based analysis of census (and other spatial data). Consider, for example, the study of ethnic minorities on Merseyside at the ward scale; a level of aggrega­ tion commonly used in practice. Figure 2 shows the percentage of population belonging to an ethnic minority for the 119 census wards. The choropleth map is hard to interpret partly because of size variation in the wards, with the result that the map emphasises extreme results typically found in the smallest wards and is, therefore, mainly dominated by variations in ward size both in terms of physical size and population. Figure 3 (see over) presents the same variable but for a 119 zonation for an approximately equal population aggregation of the 2926 census EDs. The patterns are clearer, but the visual impression is quite different. There are no longer such extremes of ethnicity (the maximum percentage is 4.97 compared with 27.7 in figure 2). Moreover, the highest ethnic areas have moved northwards probably because of interaction between the class intervals used in the map and the data generated by the zoning system, illustrating some of Monmonier's (1991) arguments. Clearly these differences require further investigation but they illustrate

how 'unsafe' naive spatial maps of census data can be. Figure 3 also illustrates another aspect of this form of zoning. Indeed, because the 119 zones are almost identical in size, raw counts could have been mapped directly because the denomi­ nators are now virtually the same everywhere.

Equal population zoning is regarded as an extremely useful reaggregation of census data for many purposes because:

1. The areas are of nearly equal size thereby removing the confounding effects of size.
2. The data precision for each new zone is the same so that differences in the mapped patterns reflect the variables being studied rather than either small number effects or sampling variability (more relevant for 10% census data).
3. The areas are size comparable, so they could more readily be ranked or subjected to other forms of spatial analysis and modelling.



□ 0.23 ,;;; *X* < 0.57

OJ] 0.57 ,;;; *X* < 0.80

Quantile class, *x(%)*

§

•

Hfll

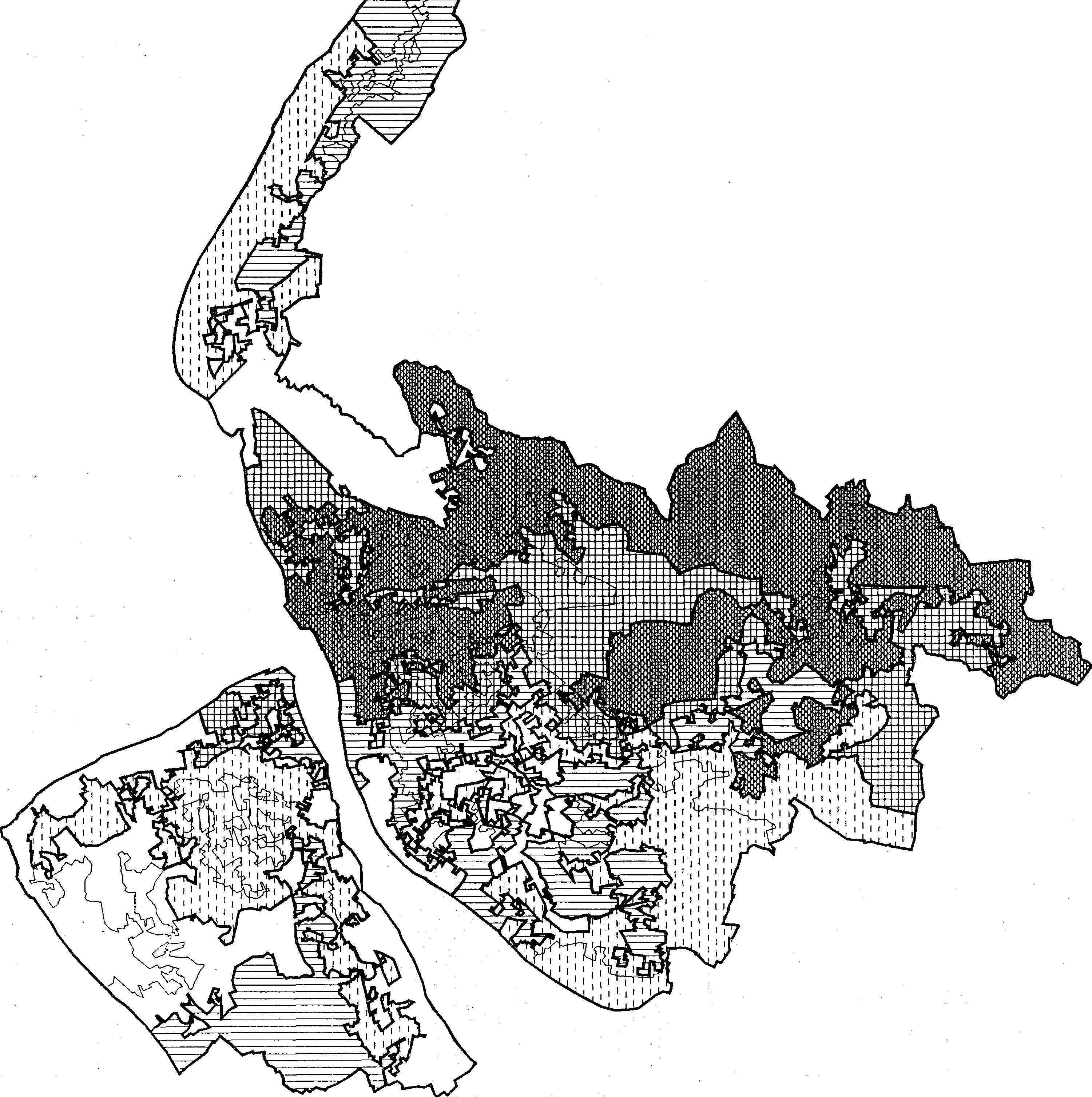
0.80,;;; *X* < 1.17

1.17 ,;;; *X* < 1.85

1.85 ,;;; *X* < 27.77

**Figure 2.** The percentage of the population belonging to an ethnic minority, in 119 wards in Merseyside, 1991.

The principal problem is that the results may be only one of possibly several different but nearly equivalent equal population zonations. This indeterminacy of the results needs to be investigated further. It is also a feature of standard census geography. It is a fact of geographical life and it should not be ignored just because it is hard to handle. In the current example, controlling for denominator size may not be sufficient. It might also be necessary to design the zones such that the regions have similar levels of heterogeneity in terms of a number of other variables; for example, social class. Whether controlling for these effects is a zone-design or a spatial modelling issue, is a matter for further investigation. The objective here is to pro­ vide the tools that will allow users of the 1991 Census to experiment with different zoning systems. At the end of the day it is they, not the census agencies, who should define the zoning systems they wish to use. It is also incumbent upon them to check the robustness of their results to zone-design and aggregation effects.



□ 0.64 ,;; *X* < 0.9

Quantile class, x(%}

GJj 0.9 ,;; *X* < 1.12

8 1.12 ,;; *X* < 1.66

m 1.66 ,;; *X* < 3.33

l!lllill 3.33 ,;; *x* < 4.97

**Figure 3.** The percentage of the population belonging to an ethnic minority, in 119 zones with equal total population.

* 1. **Relationship between unemployed and n,o cars,,,** \_. . ,

A final illustration of the potentiaf u§e:fulness, of 'op'timal zoning as a census analysis

visualisation tool is provided by looking at the correlation between the unemployed and households without cars. At the census ED level, these two variables are highly correlated; *r* = 0.99. The challenge now is to investigate the nature of map-zoning systems that yield very different correlation results. Table 4 shows that very different correlation targets can be easily obtained. Note how the basic AZP algorithm works well when the problem is easier to handle; for instance, when the ratio of *N* to *M* is large. As the number of regions increase so the simulated annealing starts to pro­ duce better results, albeit with a vast increase in compute times. Easy targets such as correlation of 0.0 for 50 regions took AZP 16 seconds. The harder target of

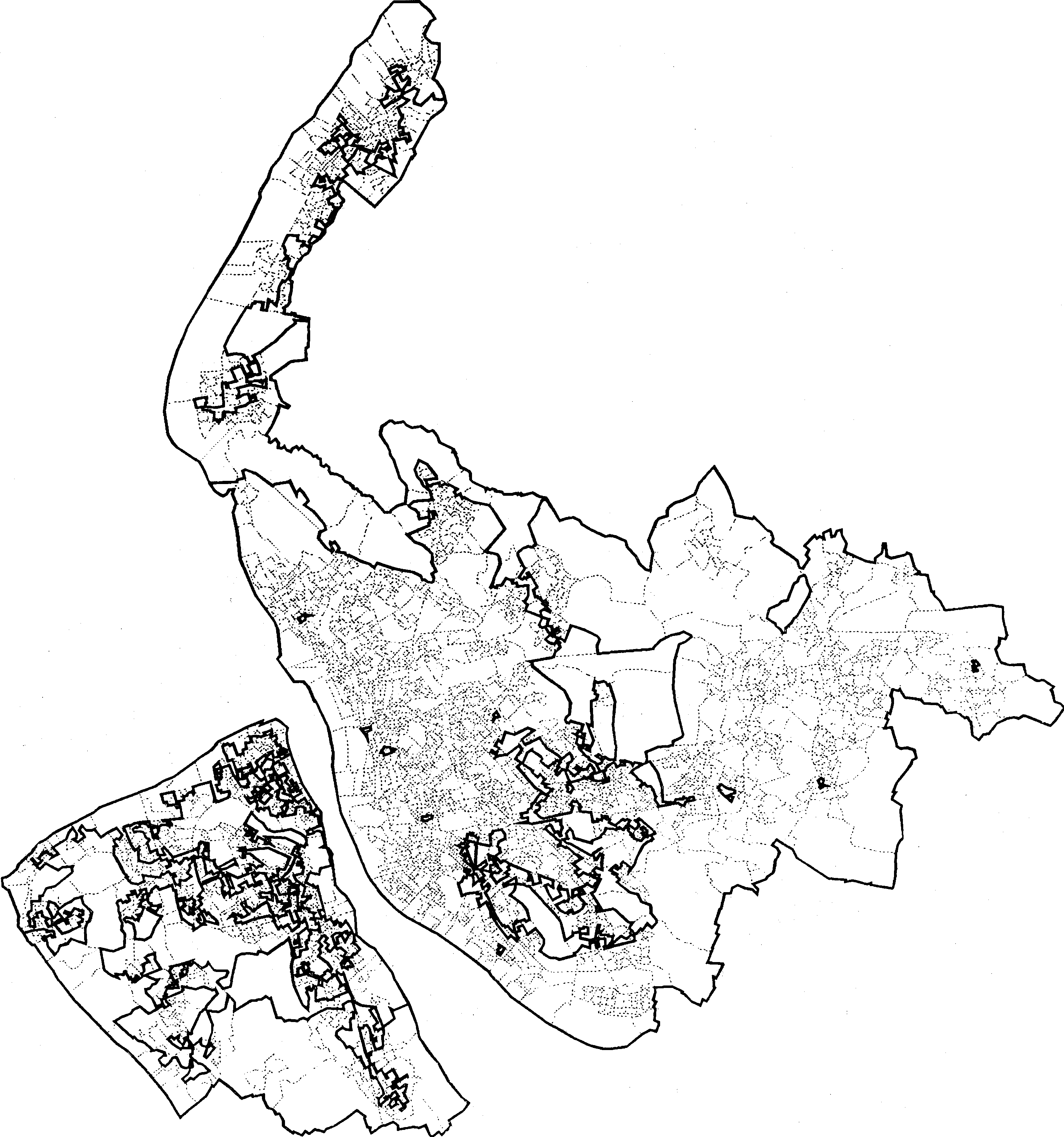
- 1.0 took the simulated annealing version 9011 seconds. The targets of 1.0 were easily attained by the basic algorithm.

**Table 4.** Correlation between unemployed and no cars on Merseyside. Number Target Method

|  |  |  |  |
| --- | --- | --- | --- |
| of regions | correlation | AZP error | **AZP-SA** error |
| 10 | -1 | 0.0019 | 0.0009 |
| 50 , | -1 | 0.025 | 0.0013 |
| 100 | -1 | 0.1745 | 0.0030 |
| 200 | -1 | 0.1998 | 0.0267 |
| 10 | 0 | 0.0000 | not needed |
| 50 | 0 | 0.0000 | not needed |
| 100 | 0 | 0.0000 | not needed |
| 200 | 0 | 0.0000 | not needed |
| 10 | +1 | 0.0520 | 0.0114 |
| 50 | +1 | 0.0217 | 0.0033 |
| 100 | +1 | 0.0066 | 0.0016 |
| 200 | +1 | 0.0095 | 0.0009 |

Fitting a linear model by changing simultaneously both the parameters and the aggregation provides an interesting opportunity to investigate whether or not the zoning systems themselves contain any useful geographic information about the variables under study. In this case, it is seemingly of considerable interest to investigate the nature of the zoning systems able to change a very strong positive correlation at the ED scale into a very strong negative association after aggregation into a much smaller number of regions.

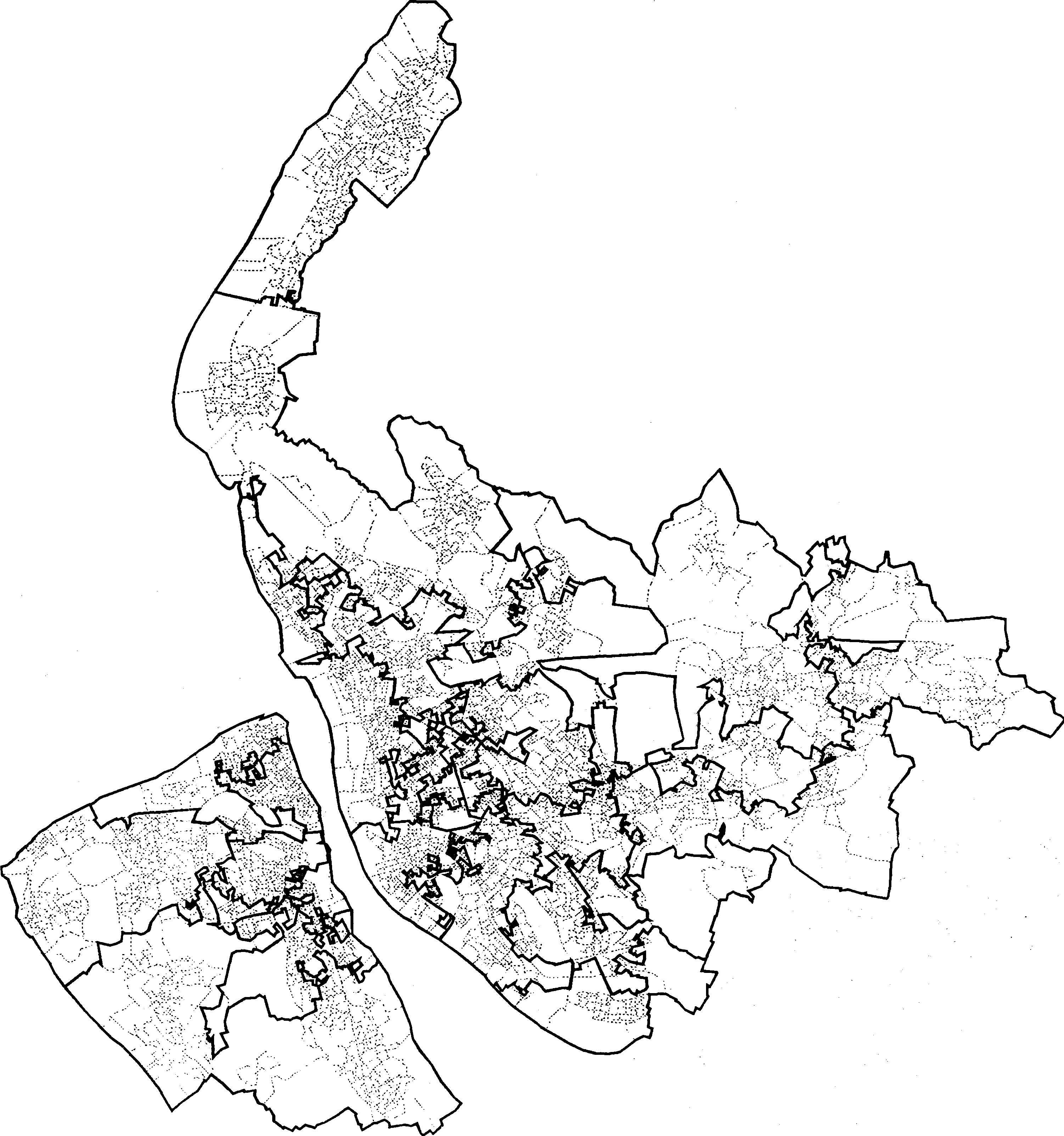
Figures 4, 5, and 6 (see over), show the zoning systems that result when the correlation targets of + 1.0, 0.0, and - 1.0 are specified. It is conjectured that the patterns of these zones contain useful insights about the relationship between scale, aggregation, and the variables of interest. The target of + 1.0 is attained by having some very large and small zones. The target of 0.0 requires zones of almost equal size, with a high degree of compactness. The - 1.0 target requires an extremely irregular set of zones. It would be interesting to explore further what happens as the numbers of output regions change and also to look at scatter plots of the data that the zoning systems create. Clearly ZDES marks the start rather than the end of zone design as a spatial analysis tool.



**Figure 4.** Zoning on correlation between unemployed population and no-car households in Merseyside, 1991; 20 regions and a correlation target of +1.0.

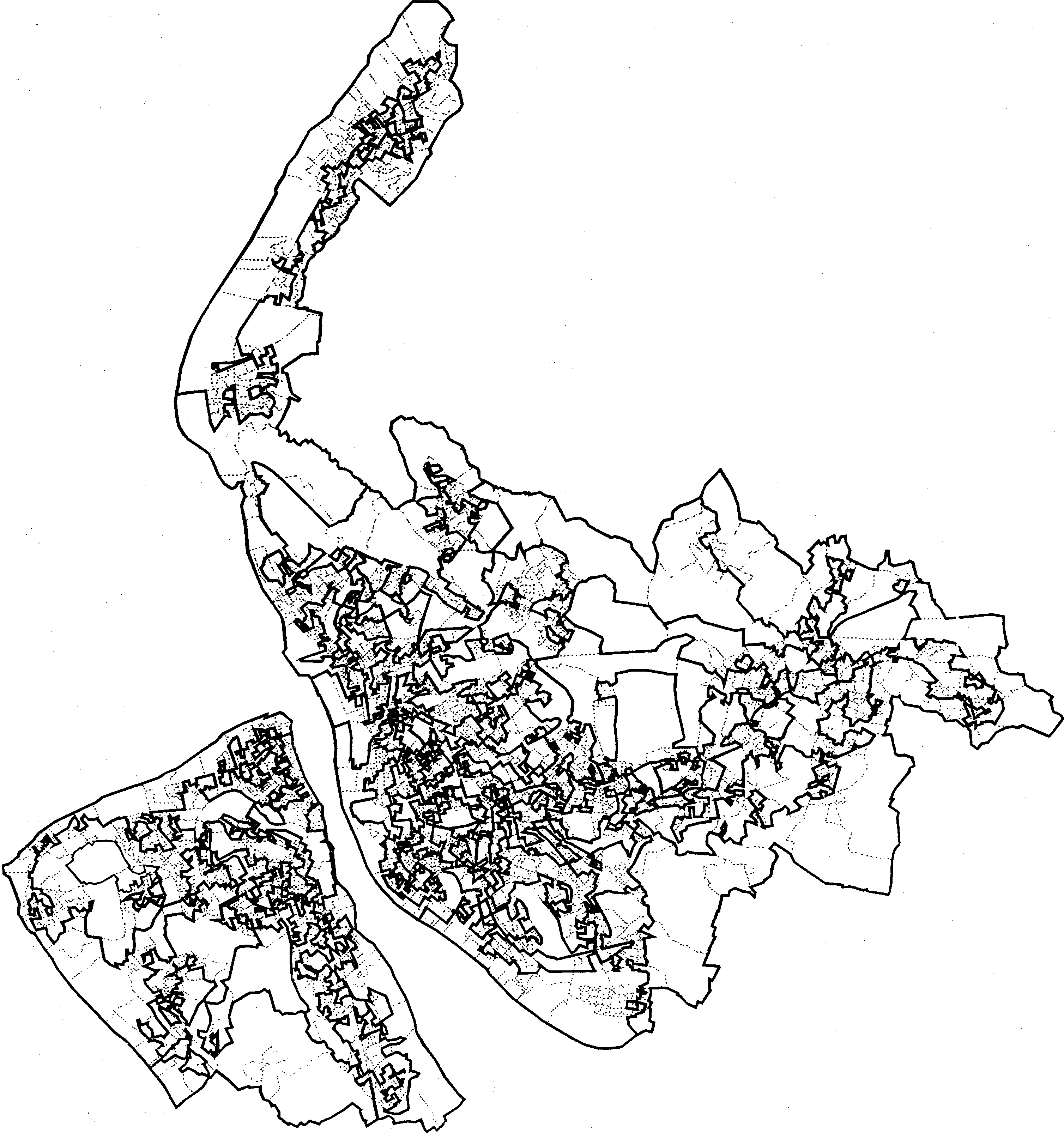
# Conclusions

It is noted that there are two opportunities for user-controlled flexibility in census zone design. The first is when the census agency makes decisions about the spatial framework to be used to report the census; the second is when the user considers what set of areal units best suits a particular analysis objective. GIS frees the need for census data to be reported only for a fixed set of data-collection areas. There is no longer, in principle, any reason why census data for 2001 could not be reported for areal units, different from EDs, but small enough and homogenous enough to be really useful spatial building blocks, whilst also being large enough to be safe in terms of confidentiality. The confidentiality restrictions applied to the 1991 Census data are not based on any scientific understanding of the problem (Openshaw, 1995) and this severely damages both the data and also creates a bureaucratic mind-set that has unnecessarily delayed the introduction of user-defined flexible



**Figure 5.** Zoning on correlation between unemployed population and no-car households in Merseyside, 1991: 20 regions and a correlation target of 0.0.

geographic representations (Openshaw, 1994 ). EDs are really of only internal and administrative relevance to the census agency, and their widespread use in census analysis is a reflection of historic expediency. Even in an ideal world, census users would still want to reengineer their zoning systems; but their task should be eased by the provision of good initial building blocks. Computers are now fast enough to perform the reengineering routinely and long before the next census is due, they will also be able to handle the initial census zone-design task too. Meanwhile, with the 1991 Census, the onus of reengineering census geography to at least reduce some of the perceived problems, rests with the census analyst. However, it is especially important that the move away from fixed, census-agency-defined, de facto zonations to more flexible, user-determined zones, does not result in zoning anarchy. Zone design can be used to destroy or to discredit the results of spatial analysis as well as to provide a powerful new tool able to enrich geographical study. The destruction



**Figure 6.** Zoning on correlation between unemployed population and no-car households in Merseyside, 1991: 20 regions and a correlation target of - 1.0

option only really applies to the naive user who will quite happily analyse data for virtually any zoning system without any great concern about the nature of the units being studied. It is important to move census analysis on from this era and to grasp the problems of zone design. The user needs to bring the uncertainties arising from the MAUP under his or her control. There is unlikely to be an infinity of alternative results, because of the use of an objective function which the zone-design process seeks to optimise. Of course, identifying suitable zone-design functions is not easy, nor is the associated task of knowing how to evaluate alternative zonations. Maybe the answer is to seek continuous space representations; or maybe merely being more explicit about the nature of the zonal entities of spatial study will be sufficient. At least some basic general purpose tools now exist to ease the task of reengineering census geographies. The principal remaining difficulties concern the definition of suitable zone-design functions best for particular purposes.

**Acknowledgements.** The research reported here was funded by the ESRC Census Initiative, under award A507265011. Thanks are due to an anonymous referee, especially for introducing the concept of zoning anarchy. Also this work is based on data provided with the support of the ESRC and JISC and uses boundary material which is copyright of the Crown and the Post Office, and uses Small Area Census Statistics which are copyright of the Crown.

**References**

Aarts E, Korst T, 1989 *Simulated Annealing and Boltzmann Machines* (John Wiley, Chichester, Sussex)

Amrhein C, 1992, "Searching for the elusive aggregation effects using age/sex specific migration data" *The Operational Geographer* **10** 32 - 38

Amrhein C, Flowerdew R, 1992, "The effect of data aggregation on a Poisson regression model of Canadian migration" *Environment and Planning A* **24** 1381-1391

Arbia G, 1989 *Spatial Configuration in Statistical Analysis of Regional and Economic Related Problems* (Kluwer, Dordrecht)

Battiti R, Tecchiolli G, 1992, "Parallel biased search for combinatorial optimization: genetic algorithms and TABU" *Microprocessors and Microsystems* **16** 351-367

Battiti R, Tecchiolli G, 1993, "Training neural nets with the reactive tabu search", unpublished manuscript, obtained by ftp on the internet

Battiti R, Tecchiolli **R,** 1994, "The reactive tabu search" *ORSA Journal on Computing* forth­ coming

Bland J A, Dawson GP, 1991, "Tabu search and design optimisation" *Computer-Aided Design* **23** 195-201

Charlton M E, Carver S, Rao L, 1995, "GIS and the Census", in *Census Users Handbook*

Ed. S Openshaw (Longman, Harlow, Essex)

Cole K, 1993, "The 1991 'Local Base and Small Area Statistics"', in *The 1991 Census User's Guide* Eds A Dale, C Marsh (HMSO, London) pp 201-247

Denham C, 1993, "Census geography: an overview", in *The 1991 Census User's Guide* Eds A­ Dale, C Marsh (HMSO, London) pp 52-69

Dudley G, 1991, "Scale, aggregation and the modifiable areal unit problem" *The Operational Geographer* **9** 28 - 32

Fotheringham S, Wong D, 1991, "The modifiable areal unit problem in multivariate statistical analysis" *Environment and Planning A* **23** 1025-1044

Glover F, 1977, "Heuristics for integer programming using surrogate constraints" *Decision science* **8** 156-166

Glover F, 1986, "Future paths for integer programming and links to artificial intelligence"

*Computers and Operations Research* **13** 533 - 549

Glover F, Laguna M, 1992, "Tabu search", in *Modem Heuristic Techniques for Combinatorial Problems* Ed.CR Reeves (Basil Blackwell, Oxford) pp 70-150

Goodchild M F, 1979, "The aggregation problem in location allocation" *Geographical Analysis* **11** 240 - 255

Hertz A, de Werra D, 1987, "Using tabu search techniques for graph colouring" *Computing*

**39** 345-351

Kendall MG, Yule GU, 1950 *An Introduction to the Theory of Statistics* (Griffin, London) Kirkpatrick S, Gelatt C D, Vecchi MP, 1983, "Optimisation by simulated annealing"

*Science* **220** 671-680

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E, 1953, "Equations for state calculations by fast computing machines" *Journal of Chemical Physics* **21** 1087-1092

Monmomier M, 1991 *How to Lie With Maps* (University of Chicago Press, Chicago, IL)

Morphet C, 1993, "The mapping of small-area census data-a consideration of the role of enumeration district boundaries" *Environment and Planning A* **25** 267 - 2.78

Openshaw S, 1977a, "A geographical solution to scale and aggregation problems in region­ building, partitioning, and spatial modelling" *Transactions of the Institute of British Geographers, New Series* **2** 459-472

Openshaw S, 1977b, "Optimal zoning systems for spatial interaction models" *Environment and Planning A* **9** 169-184

Openshaw S, 1977c, "Algorithm 3: a procedure to generate pseudo random aggregations of

*N* zones into *M* zones where *M* is less than *N' Environment and Planning A* **9**

1423-1428

Openshaw S, 1978a, "An empirical study of some zone design criteria" *Environment and Planning A* **10** 781-794

Openshaw S, 1978b, "An optimal zoning approach to the study of spatially aggregated data", in *Spatial Representation and Spatial Interaction* Eds I Masser, P J B Brown (Martinus Nijhoff, Leiden) pp 93-113

Openshaw S, 1984 *The Modifiable Areal Unit Problem* CATMOG 38 (Geoabstracts, Norwich)

Openshaw S, 1988, "Building an automated modelling system to explore a universe of spatial interaction models" *Geographical Analysis* **20** 31-46

Openshaw S, 1989, "Learning to live with errors in spatial databases", in *The Accuracy of*

*Spatial Databases* Eds M Goodchild, S Gopal (Taylor and Frances, London) pp 263-276 Openshaw S, 1994, "Social costs and benefits of the Census", in *Proceedings of the XVth*

*International Conference of the Data Protection and Privacy Commissioners* Data Protection Registrar, Springfield House, Water Lane, Wilmslow SK9 SAX, pp 89-97

Openshaw S, 1995, "The future of the census", in *The Census Users Handbook* Ed. S Openshaw (Longman, Harlow, Essex) forthcoming

Openshaw S, Gillard A, 1978, "On the stability of a spatial classification of census enumeration district data", in *Theory and Method in Urban and Regional Analysis* Ed.PW J Batey (Pion, London) pp 101-119

Openshaw S, Taylor **P J,** 1979, "A million or so correlation coefficients: three experiments on the modifiable areal unit problem", in *Statistical Applications in the Spatial Sciences* Ed. N Wrigley (Pion, London) pp 127-144

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

Press W H, Flannery B P, Teukolsky S A, Vetterling W T, 1986 *Numerical Recipes*

(Cambridge University Press, Cambridge)

Tobler W, 1989, "Frame independent spatial analysis", in *Accuracy of Spatial Databases*

Eds M Goodchild, S Gopal (Taylor and Francis, London) pp 115 -122

Tobler W, Moellering H, 1972, "The analysis of scale variance" *Geographical Analysis* **4**

34-50

© 1995 a Pion publication printed in Great Britain

***p***