22nd Australasian Transport Research Forum
Sydney, September 1998

A graph partitioning approach towards the automatic generation of transport analysis zones

Jessica Guo
Division of Building, Construction and Engineering
CSIRO

Abstract:

In spatial analysis, it is often necessary to use aggregated data, either because the data is available only for predetermined spatial units, or to make the problem manageable. Zoning is a process where M input zones are grouped into N output zones so that various constraints are satisfied (e.g. internal connectivity of the resulting zones, $N < M$). For the purpose of the zoning problem, the exact configuration of the boundaries of the input zones is not important. One possible abstraction is to represent the input zones as nodes of a graph, where each node in the graph represents a zone and each link represents the adjacency relation between zones. Using the graph representation as a convenient framework for solving the zoning problem a zoning algorithm based on a modified graph partitioning problem (GPP) is proposed.

Contact author:

Jessica Y Guo
Division of Building, Construction and Engineering
CSIRO
PO Box 56 (Graham Road)
HIGHETT VIC 3190

Telephone: (03) 9252 6000 Fax: (03) 9252 6092
Email: Jessica.Guo@dbce.csiro.au

Introduction

The general subject of spatial modelling and analysis relates to the phenomena that take place in a two-dimensional space. For example, travel behaviours which transportation models attempt to simulate are in essence vehicular movements from one geographical location to another. Consequently, the way in which space is represented in these studies is of great importance. Unless there are needs and available methodologies to represent space as continuous as it is, however, in almost all spatial modelling studies and analyses a discrete representation is used. This is usually achieved by dividing the study region, such as a metropolis, into smaller contiguous areal units, the collection of which is often referred to as a zoning system

While data inputs for spatial studies are aggregated over the areal units, the process of aggregation is constrained by the chosen zoning system In fact, the level of perception of the spatial model, and hence the outcome of the study, is dependent on the zoning system used Different zoning systems in terms of number or configuration of zones yield different results. This modifiable areal unit problem (MAUP) is endemic to all zonal data Its impact is evident in various types of spatial analysis and modelling studies (Arbia 1989; Batty and Sammons 1978; Ding 1994; Fotheringham and Wong 1991; Openshaw 1977; 1984; Putman and Chung 1989).

The presence of MAUP raises the question of how to obtain the most suitable zoning system, if any, for a given spatial study This is referred to as the zoning problem in this paper, which intends to provide insight into the complexity of the issue in the context of transportation modelling. Criteria involved in zoning system design are discussed both qualitatively and quantitatively, with mathematical formulations to aid the understanding of the problem. A brief description of existing algorithms for zoning is also presented, and a new method, namely, the graph partitioning approach, is described together with its implementation and some empirical results

Use of transport analysis zones in transportation modelling

In conventional aggregate transportation modelling, the boundary of the study region, namely, the cordon line, is defined prior to the beginning of the forecasting process The study region is then divided into areal units known as transport analysis zones (TAZs) A system of TAZs thus represents the underlying space in which the travel activities are to be modelled Centroids of these zones are then used to represent traffic loading nodes in the simplified transportation network of the model. In other words, the TAZs serve as point representations of all possible trip origins/destinations within the study region Theoretically, finer TAZs could yield greater modelling accuracy. In the extreme case, each individual household can be treated as a zone. However, the problem of stability over time renders a super-fine zoning system impractical as it is almost impossible to forecast at that level of detail all the changes that would affect transport demand (Ortuzar and Willumsen 1994). Therefore, the use of a zoning system is required to reduce the level of detail and hence improve the stability of the spatial environment that the individual zone is representing

Another role of the TAZs is to serve as unified spatial units for spatial data. In transportation modelling, data of such a spatial nature includes land-use information, socio-economic information and origin–destination information Depending on the original units in which these data are collected or released, and also depending on the data type, different aggregation procedures are required to aggregate them and link them to the TAZs. In this regard, the use of a zoning system is inevitable so as to make the amount of information manageable.

The TAZs are often defined by grouping together smaller spatial units such as the census collection districts (CCDs). Variation in both the scale and the configuration in which the spatial units are grouped into zones leads to a large number of possible zoning systems, each of which represents the underlying space and data in a different way. Although the presence and the impact of the MAUP has been acknowledged since the 1970s, the use of fixed TAZs, often developed manually in a haphazard manner, is widespread in practice (O'Neil 1991; Batty and Xie 1994; Ding 1994; Openshaw and Alvanides 1996). Not only is the suitability of the fixed zoning system with regard to the particular study at hand not questioned, but also the changes that have taken place over time in land use and/or socio-economic characteristics of the population are also ignored (O'Neill 1991).

The use of fixed TAZs has probably been the result of two problems. The first problem lies in the difficulty in grouping the areal units in a large number of ways to generate the zoning system of the desired characteristics. The process is lengthy and cumbersome, with the degree of the difficulty escalating according to the number of units to be grouped. This immense computational problem has partly led to the second problem: our limited understanding of the interweaving relationship between the underlying data, the zoning system and the transportation model itself. The crux of the problem is the difficulty in measuring the errors produced by the discrete representation of space. Is there a generic definition for the optimal zoning system to use? Or, should the zoning system be designed according to the input data and the specific model at hand? In order to answer these questions, the problem of computational difficulty has to be overcome first. Nevertheless, before an approach to solving the computational problem is presented, an examination of the various issues involved in the current practice of zone design is deemed necessary to provide a better understanding of the zoning process

Design of transport analysis zones

A survey of the literature reveals the absence of absolute guidelines for designing optimal zoning systems Opinions on what design criteria to consider, which have been put forward by practitioners and researchers who are aware of the significance of zoning effects are diverse. For example, Oppenheim (1995) gives the general principle that the number of zones should be as many as possible, while at the same time maximising the internal homogeneity of the resultant zones and maximising the differences between them. In the online document *Introduction to Urban Travel Demand Forecasting*, it is stated that while zones should attempt to bound homogeneous urban activities, they should also consider natural boundaries and census designations Zone size should also be considered in terms of the density or nature of urban development, with the transportation network forming the boundaries of the zones. Ortuzar and Willumsen (1994) also suggest that homogeneity and compatibility with other administrative divisions should be taken into account. They believe that the use of main roads as zone boundaries should be avoided. They further argue that the role of zone centroids and centroid connectors in the modelled transportation network should be recognised and used to help defining zone boundaries. Zone shape should allow easy determination of centroid connectors so as to represent the main costs to access them. Zone size should be of similar dimension in terms of travel time rather than areal size

The various sets of zoning criteria described above are largely determined by what the researchers and modellers interpret as meaningful. The differences among them are most likely rooted in the perspective of the modellers as a result of the different purposes of their studies. They also reflect the dependency of zoning criteria on the data available to the modellers The level of details of the network, the land use and the social characteristic information are all fundamental constraints to the degree of flexibility in zoning system design.

One characteristic common to these divergent guidelines is their multiplicity of criteria, many of which may be intrinsically conflicting with each other. For example, while the desirable zone size is pursued, the homogeneity of zones might be jeopardised. Difficulties in handling multiple zoning criteria simultaneously have been noted in the works of O'Neill (1991) and Ding (1994). In an attempt to develop optimal TAZs using GIS, O'Neill (1991) proposed to take into account homogeneous socio-economic characteristics, minimal intra-zonal trips, utilisation of physical, historical and political boundaries, avoidance of doughnuts (no zones contained entirely within other zones), equality in size of population, household, or area per zone, and confirmation to census track boundaries where possible. However, in her actual implementation, only a subset of these criteria was used. Ding *et al.* (1993) also attempted to provide a GIS-based approach to delineate TAZ boundaries. The factors considered include homogeneity, contiguity, consistency (no island and a zone should not be separated) and equality in terms of trip generation. Upon realising the difficulty in meeting all these criteria, Ding (1994) also only took into account a subset of them in a subsequent paper investigating the impact of spatial data aggregation.

In addition to demonstrating the escalating computational difficulty due to the multiplicity of criteria, the experiences of both researchers also led to the need to re-examine the practice of using as many zoning criteria as nominated. As a reference, the most commonly used zoning criteria as proposed by the various researchers are listed below:

- minimum variation in spatial characteristics of zones (i.e. area, size, shape);
- minimum variation in socio-economic characteristics of zones (i.e. population, number of households);
- maximum spatial interaction (i.e. number of trips);
- contiguity and compactness of zones;
- compatibility with natural and physical barriers to travel;
- equal trip generation/attraction;
- traffic load points as centroids;
- minimum intra-zonal trips;
- compatibility with other existing spatial units (i.e. administrative divisions, previous zoning systems, census tract boundaries, political boundaries); and
- decision maker's preference in number of zones

However, even if a subjectively compromised solution can be found despite computational difficulty, there is no way to guarantee that such a zoning system is 'better' or more 'meaningful' than all other alternative systems to the modelling task at hand. In this regard, the role of an efficient zoning tool, which is capable of taking into account any nominated combinations of criteria, has become very important. The tool will greatly facilitate the investigation of the effects of various combinations of zoning criteria on modelling outcomes and help the modellers to decide on the most suitable one to use.

The mathematics of the zoning problem

Previous attempts at devising a zoning tool are all variations of the same approach of treating the problem as a combinatorial optimisation problem. The **M** basic areal units are grouped into **N** zones such that the target set of zoning criteria is optimally satisfied. These attempts differ mainly in the zoning criteria being considered, and hence in the objective functions and sets of constraints being used during the process of optimisation. The key feature common to these attempts is to start the search process with an initial zoning system formed by growing zones from those CCDs selected as zone centres. CCDs are then swapped from one zone to another in an effort to improve the performance of the solution in terms of the chosen criteria. The design of the swapping procedure and the degree of difficulty in reaching an optimal solution vary according to the nature and the number of objective functions and constraints considered. One major drawback of these methods, however, is the lack of the flexibility to account for various multiple zoning criteria.

The mathematical complexity of the zoning problem can be best reflected by an integer programming formulation as follows.

Let X be a M by N matrix of variables representing the assignment of an areal unit i and a resultant zone j. That is, for $i = 1 \ldots M$ and $j = 1 \ldots N$:

$$X_{ij} = \begin{cases} 1, \text{ if areal unit } i \text{ is assigned to zone } j \\ 0, \text{ otherwise} \end{cases}$$

For a solution of X to be feasible, that is to be a proper representation of a zoning system, the solution has to satisfy certain conditions including:

$$\sum_{j=1}^{N} = X_{ij} = 1, \text{ for } i = 1 \ldots M \tag{1}$$

$$\sum_{i=1}^{M} = X_{ij} \geq 1, \text{ for } j = 1 \ldots N \tag{2}$$

While condition (1) ensures that every areal unit is assigned to one and only one zone, condition (2) ensures that every zone contains at least one areal unit.

Another condition imposed on solutions to the problem is that zones have to be internally connected. In other words, every areal unit needs to be reachable via units in the same zone to every other unit in that zone. The formulation of this condition relies on the use of the adjacency matrix, A, defined over the areal units such that:

$$A_{ik} = \begin{cases} 1, & \text{if areal units } i \text{ is adjacent to areal unit } k \\ 0, & \text{otherwise} \end{cases}$$

where $i = 1 \ldots M$ and $k = 1 \ldots M$. Thus, the adjacency matrix for a zone j can be calculated by:

$$A_{jik} = A_{ik} X_{ij} X_{kj}$$

yielding:

$$A_{jik} = \begin{cases} 1, & \text{if areal units } i \text{ and } k \text{ are both in zone } j \text{ and are immediately} \\ & \text{adjacent to each other} \\ 0, & \text{otherwise} \end{cases}$$

A series of matrices can then be defined for *1*-neighbour adjacency, *2*-neighbour adjacency etc., where n-neighbour adjacency means that one areal unit can be reached from another unit via at most n neighbouring units:

$$A^1_{jik} = \sum_{n=1}^{M} A_{jin} A_{jnk} \qquad \text{1-neighbour adjacency matrix}$$

$$A^2_{jik} = \sum_{n=1}^{M} A^1_{jin} A_{jnk} \qquad \text{2-neighbour adjacency matrix}$$

$$A^3_{jik} = \sum_{n=1}^{M} A^2_{jin} A_{jnk} \qquad \text{3-neighbour adjacency matrix}$$

$$A^{M-2}{}_{jik} = \sum_{n=1}^{M} A^{M-3}{}_{jin} A_{jnk} \qquad \text{(M-2)-neighbour adjacency matrix}$$

It follows for the contiguity condition that, for any two areal units i and k of zone j, unit i is reachable from unit k can be formulated as:

$$A^{M-2}{}_{jik} \geq 1, \text{ for all } i \text{ and } j \text{ such that } X_{ij} X_{kj} = 1$$

Rewriting the above expression and extending the condition to every zone in the system yields condition (3) for a solution being feasible:

$$A^{M-2}{}_{jik} - X'_{ij} X_{kj} \geq 0, \text{ for all } i, j \text{ and } k \qquad (3)$$

Following the above formulation, it is found that almost all the other zoning criteria can be quantified and expressed as quadratic functions of X_{ij}. For example, if zones of equal population size are pursued, the criterion can be expressed as to minimise the total squared deviations of individual zonal population from the mean population size:

$$\text{Min} \sum_{j=1}^{N} \left(\sum_{i=1}^{M} X_{ij} \times W_i - \frac{\sum_{i=1}^{M} W_i}{N} \right)^2$$

where W_i denotes the population size of areal unit i. In this fashion, the zoning problem becomes mathematically a quadratic 0–1 integer programming problem. However, as discussed by Macmillan and Pierce (1994), the high complexity of condition (3) results in a non-convex solution space for the problem. This implies that the problem cannot be solved by any standard integer programming routines. Thus, the mathematical model described above is theoretically valid, but impractical for solving the problem.

Zoning as graph partitioning

Another mathematical model, based on graph theory, is developed by the author to provide a different perspective for examining the zoning problem. As illustrated in Figure 1(a), a map of areal units, each of which is defined digitally by a series of polylines, typically forms the input to the zoning process The idea is to reformulate the problem so that the units themselves disappear from consideration. Instead, each areal unit, or polygon, can be represented by a node which is associated with some data Wherever two units are adjacent, the neighbouring relationship is represented by a link. The nodes and links thus comprise a mathematical graph, as shown in Figure 1(b). The problem of grouping the areal units to form the zones can be rephrased as partitioning the nodes in the graph into disjoint subsets.

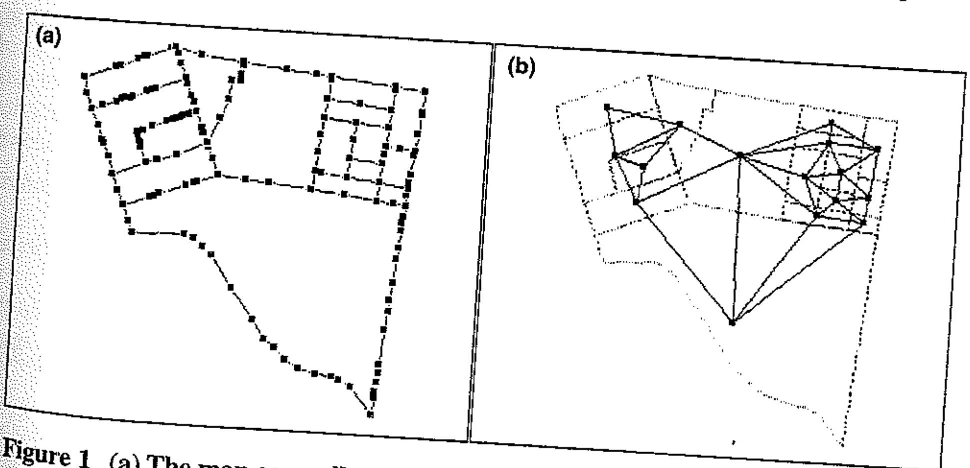

Figure 1 **(a) The map as a collection of areal units defined by polylines; and (b) the corresponding graph representation, in which each areal unit is denoted by a node, and a link is drawn between every pair of adjacent units**

The advantage of the graph formulation is its conceptual simplicity. As observed in the previous section of zoning system design, data involved in the zoning problem falls into two categories. The first category is data describing the characteristics of individual areal units (e.g. population, area, number of households). The second category is data describing the interaction between the areal units (e.g. number of trips, level of adjacency). The proposed formulation serves as a convenient framework to capture these two types of data as attributes of nodes or links. While the integer programming model treats the problem as an assignment matrix, the graph model provides better visual representation and treats the problem as a graph partitioning problem.

Another advantage of the graph formulation is the possible adaptation of existing algorithms for problems of a similar nature in other fields. An already established graph partitioning problem (GPP), though not immediately equivalent to our zoning problem under the graph formulation, is a problem of bisecting the nodes of a graph into roughly equal partitions such that the number of links connecting nodes in the two partitions is minimised. When more than two partitions are required, the problem becomes the k-way graph partitioning problem (k-GPP). The algorithm can be naturally extended to graphs that have weights associated with the nodes and the links of the graph. In this case, the goal is to partition the nodes into k disjoint subsets such that:

- the sum of the node weights in each subset is the same; and
- the sum of the link weights whose incident nodes belong to different subsets is minimised.

The GPP has found applications in many areas including parallel scientific computing, task scheduling and VLSI design. Some examples are domain decomposition for minimum communication mapping in the parallel execution of sparse linear system solvers, mapping of spatially related data items in large geographical information systems on disk to minimise disk I/O requests, and mapping of task graphs to parallel processors (Ahmad and Dhodhi 1994; Bui and Moon 1996).

The zoning problem of our interest resembles the k-GPP in many ways. Not only do both problems aim at producing disjoint subsets of nodes, but also the objectives of zoning, derived from the zoning criteria, can often be formulated as one of the two objectives of the k-GPP. This is possible because the graph representation has captured the relevant data as weight vectors associated with the nodes and the links. Depending on the characteristics of the criteria selected for consideration, the objectives of zoning can be handled as follows:

- criteria involving 'equal attributes' (e.g. area, population, number of households) conform to the first objective of k-GPP;
- criteria involving maximising (or minimising) inter-zonal (or intra-zonal) attributes conform to the second objective of the k-GPP;
- compactness of a zone can be measured in terms of the adjacency level between the composing areal units. Better compactness can be achieved by maximising the intra-zonal adjacency levels; and
- in cases where existing political or natural boundaries are to be respected, one can first remove from the graph representation the links corresponding to such 'uncrossable' boundaries. This criterion thus involves preprocessing the graph rather than being treated as a partitioning objective.

As zoning may involve a combination of some or all of these criteria, the problem differs from the k-GPP in that weight vector rather than single weight variable are considered. This multiplicity of variables leads to an extended version of the k-GPP developed by the authors, the multi-criteria k-GPP.

Proposed zoning algorithm

The proposed algorithm for solving the multi-criteria k-GPP, and consequently the zoning problem, has evolved from existing algorithms for solving the k-GPP. The k-GPP has been proven to be NP-complete, which means that the computational effort required for finding the optimal partitioning grows exponentially as the value of k and the number of nodes, n, increase. Hence, for problems involving large k and n, heuristics are needed to give approximate solutions. A large proportion of the existing algorithms developed for solving the k-GPP is based on recursive bisection. That is, the algorithms first generate a 2-way partitioning of the graph, and then recursively generate further 2-way partitioning of each resulting partition. After $\log_2 k$ phases, the graph is partitioned into k partitions. Thus, the problem of performing a k-way partitioning is reduced to that of performing a sequence of bisections. Recently, many such algorithms have been combined with the *multi-level technique*, and have been shown to be highly effective (Karypis and Kumar 1997).

The algorithm proposed for the multi-criteria k-GPP is based on the multi-level bisection algorithms. Figure 2 demonstrates how a 5×5 grid graph is partitioned into 4 groups using this algorithm. It consists of the following procedures:

- *Coarsening* – The original graph is transformed into a sequence of smaller graphs by merging pairs of nodes together. The reduction in graph size significantly reduces the computational efforts required for partitioning
- *Uncoarsening* – The smaller graphs are successively partitioned to yield the desired number of partitions, while the partitioning is eventually projected back to the original graph.
- *Partitioning* – At a given level, partitioning is performed by bisecting each of the existing partitions. Bisection aims to replace the original partition with two smaller partitions, of which the sums of the total node-weight vectors are almost equal. Thus, for example, 2 levels of partitioning are required to yield 4 partitions.
- *Refinement* – As the graph is uncoarsened, a higher degree of freedom is gained. That is, the partitions now consist of more nodes and, consequently, there are more ways to move the nodes between partitions. This regrouping of nodes takes place whenever the partitioning from a coarser level is projected to the finer level, so that the sum of the weight vectors of the cutting links is minimised.

Experimental results

To illustrate the performance of the multi-level graph partitioning approach, the authors have written a program using the C++ language incorporating the proposed multi-level k-GPP algorithm. Two series of zoning problems are solved for a study region comprising three LGAs in Victoria: Kingston, Bayside and Glen Eira. The first series takes into

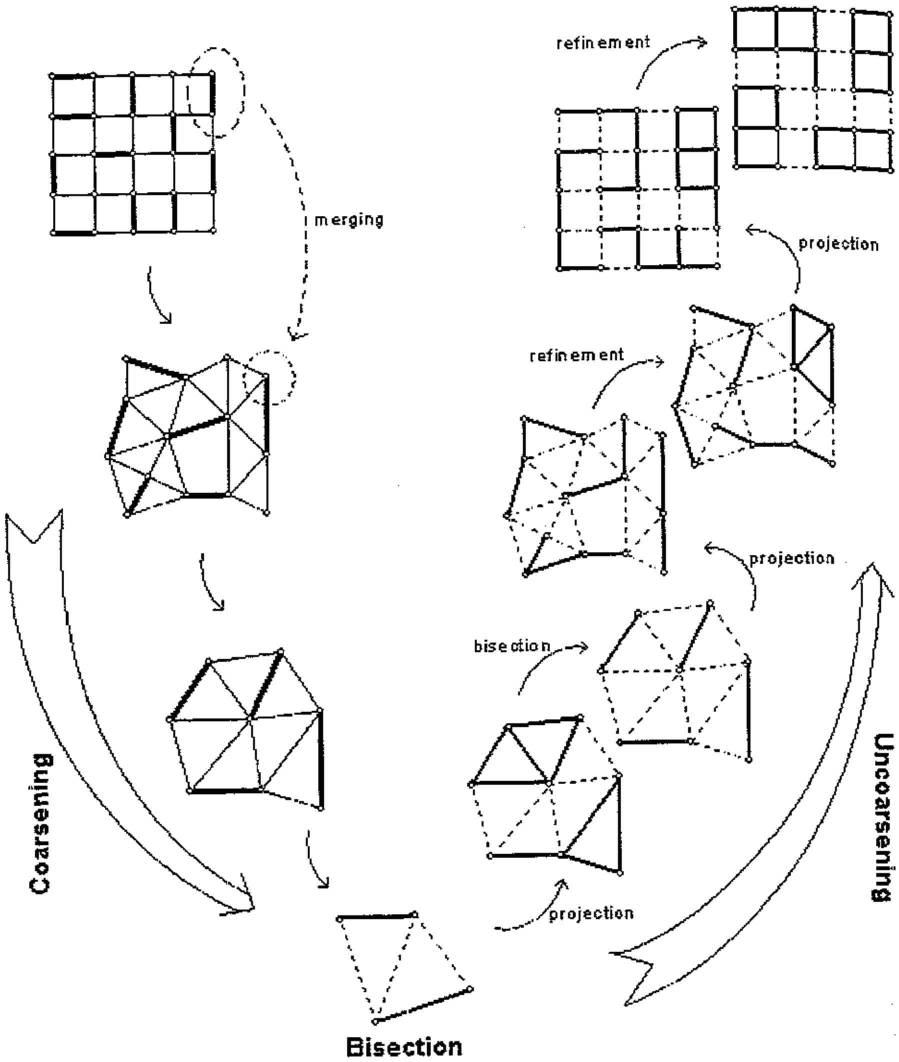

Figure 2 The proposed multi-level bisection algorithm for the zoning problem

account one single criterion of equality in population size. The second series aim is to achieve both equity in population and compactness in shape as the design criteria. The digital boundaries of the 577 CCDs and the population data for the area are extracted from the 1996 Australian census data as input for zoning. The population sizes of the CCDs form the node weights of the underlying graph. The digital boundaries are used to calculate the level of adjacency between every pair of CCDs to yield the link weights. That is, for every pair of adjacent CCDs, the averaged ratio of the length of their common boundary to the boundary of the two individual CCDs is computed. This calculation is performed external to the zoning procedure by the MapBasic program running on MapInfo, though the linkage between the two programs can be made hidden from the user.

For each of the test series, 10 levels of zoning, from the lowest scale of 10 zones to the highest of 100 zones, are performed. At each zoning scale, multiple runs of the program have been undertaken and five sets of the outcome are presented below. Results for the first test problem is evaluated with respect to only the first objective, namely, minimising the population deviations (see Figure 3) In addition to the same first objective, results for the second test problem is evaluated also with respect to the second objective of minimising link weights, namely, the level of adjacency between zones (see Figures 4 and 5). From these figures, a general improvement in performance is observed as the number of resultant partitions falls. This is obviously due to the concurrent increase in the number of possible ways to refine the partitions. Figures 6(a) and 6(b) show examples of the zoning systems produced for the two problems. For the same number of 20 resultant zones, the inclusion of compactness as the second zoning criterion evidently yields zones of 'better' shapes, as reflected in Figure 6(b)

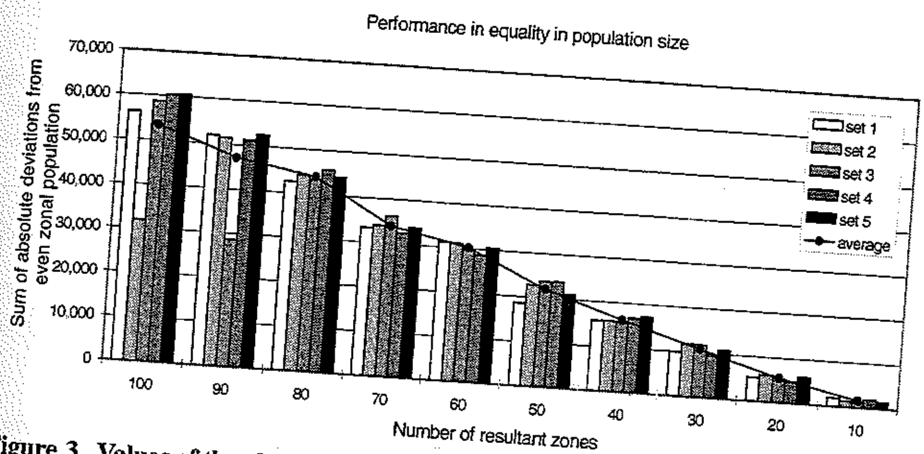

Figure 3 Values of the objective function for the first test problem of generating zones of equal population

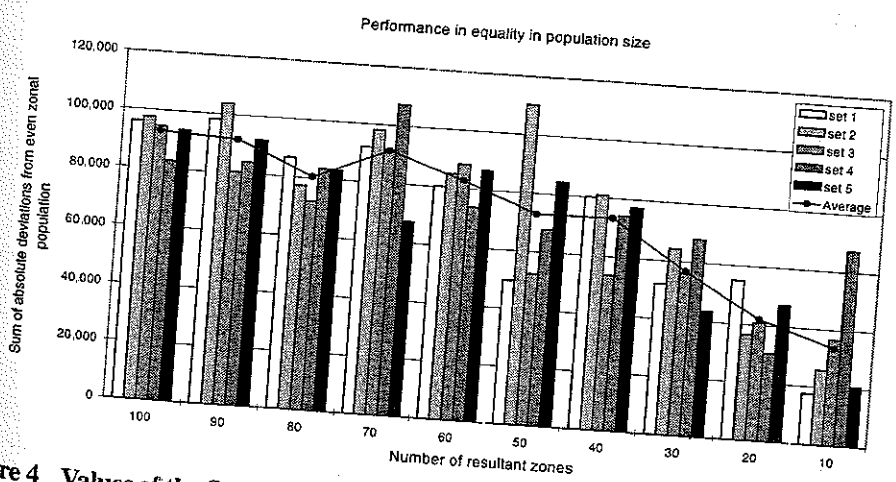

Figure 4 Values of the first objective function for the second test problem of generating zones of equal population and maximum degree of compactness

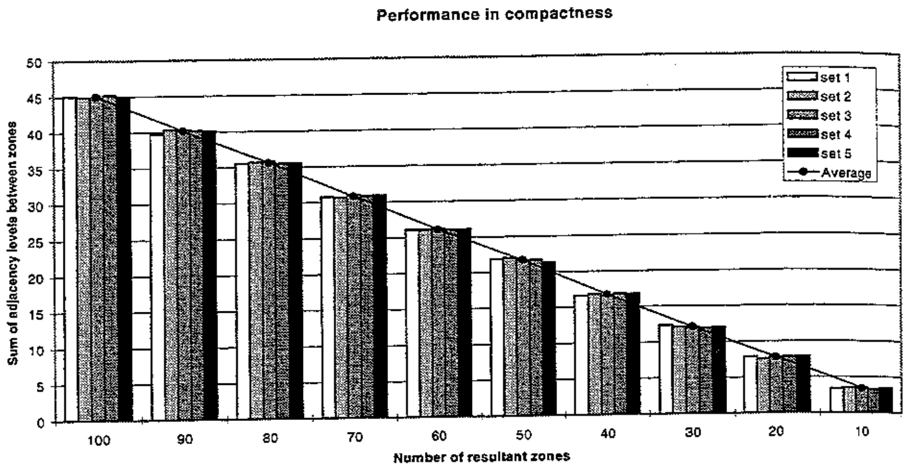

Figure 5 Values of the second objective function for the second test problem of generating zones of equal population and maximum degree of compactness

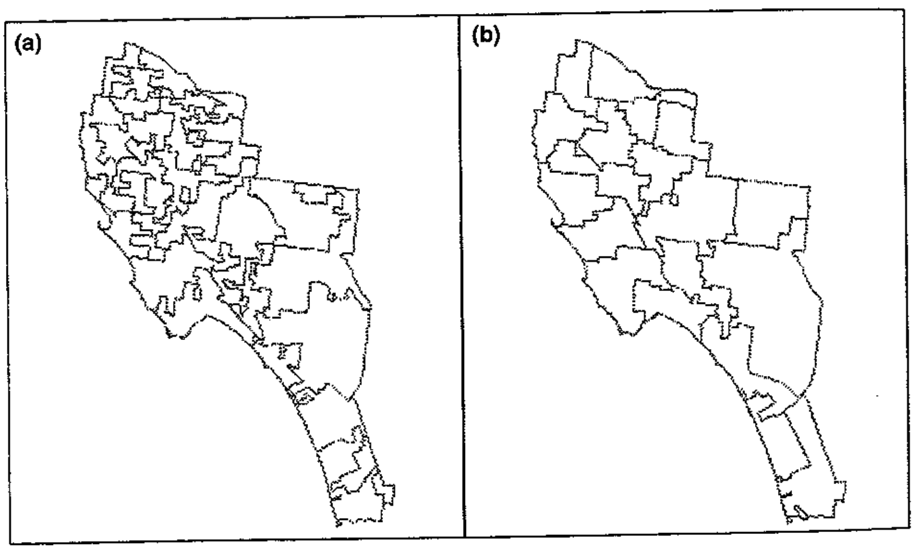

Figure 6 Zoning systems of 20 zones as solutions to the two test problems: (a) equal population zoning; and (b) compactness and equal population zoning

Conclusion and directions for future research

The problem of zoning system design is ultimately based on how space is represented intrinsically in spatial models. It is of paramount importance not only in transport modelling but also in other spatial studies. Yet little theory on the subject has been developed. The multi-criteria *k*-GPP algorithm, as described in this paper, aims to serve as a practical solution for generating meaningful zoning systems. It is also envisaged as a handy tool to aid future empirical studies to examine the impact of interaction among diverse zoning criteria on zoning system design.

It has been demonstrated how the graph representation provides a suitable conceptual framework for generating zoning systems Formulated as an extended *k*-GPP, the multi-criteria *k*-GPP, the zoning problem can be partly solved by applying an algorithm evolved from existing multi-level bisection algorithms. Currently the algorithm can successfully handle two zoning criteria at the same time in addition to the intrinsic criterion of connectivity. In order to handle more than two criteria, one can always resort to combining multiple objectives into single weighted function. Nevertheless, the author is exploring the use of more sophisticated techniques drawn from the field of multi-objective optimisation to tackle the problem. The ultimate goal is to allow user interaction with the zoning tool so that the user is able to explore and modify his or her preferences on zoning criteria easily for evaluation as zoning systems are generated automatically in a relatively efficient manner. In this case, the zoning tool serves as a decision support system which can be further linked to spatial models and geographical information systems to enhance the accuracy and validity of spatial analysis.

Acknowledgments

The author wishes to thank Gerardo Trinidad, Meigu Guan and Nariida Smith for their valuable comments and generous support.

References

Ahmad, I and Dhodhi, M K (1994) On the m-way graph partitioning problem *The Computer Journal 38*(3), 237–244

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

Batty, M and Sammons, R (1978) On searching for the most informative spatial pattern *Environment and Planning A 10*, 747–779

Batty, M and Xie, Y (1994) Modelling inside GIS: Part 1 – Model structures, exploratory spatial data analysis and aggregation *International Journal of Geographical Information Systems 8*, 291–307

Bui, T N and Moon, B R (1996) Genetic algorithm and graph partitioning *IEEE Trans. on Computers 45*(7), 841–855

Ding, C (1994) Impact analysis of spatial data aggregation on transportation forecasted demand: a GIS approach, pp 362–375 of *URISA Proceedings 1994*

Ding, C Choi, K Kim, T J (1993) GIS-based traffic analysis zone design, presented to 3rd International Conference on Computers in Urban Planning and Urban Management, 23–25 July, Atlanta, GA

Fotheringham, A S and Wong, D W S (1991) The modifiable areal unit problem in multivariate statistical analysis *Environment and Planning A 23*, 1025–1044

Karypis, G and Kumar, V (1997) A fast and high quality multilevel scheme for partitioning irregular graphs *Technical Report*, Department of Computer Science, University of Minnesota <http://www.cs.umn.edu/~karypis>

Macmillan, W and Pierce, T (1994) Optimization modelling in a GIS framework: the problem of political redistricting, chapter 11 of Fotheringham, S and Rogerson, P (eds) *Spatial Analysis and GIS* London: Taylor & Francis

O'Neill, W A (1991) Developing optimal transportation analysis zones using GIS *ITE Journal 12*, 33–36

Openshaw, S (1977) Optimal zoning systems for spatial interaction models *Environment and Planning A 9*, 169–184

Openshaw, S (1984) *Concepts and Techniques in Modern Geography, Number 38 – The Modifiable Areal Unit Problem* Norwick: Geo Books

Openshaw, S and Alvanides, S (1996) Designing zoning systems for representation of socio-economic data <http://www.geog.leeds.ac.uk/pgrads/s alvanides/papers/designin.html>

Oppenheim, N (1995) *Urban Travel Demand Modelling* New York: Wiley

Ortuzar, J deD and Willumsen, L G (1994) *Modelling Transport* (2nd ed) New York: Wiley

Putman, S H and Chung, S-H (1989) Effects of spatial system design on spatial interaction models: 1 – The spatial definition problem *Environment and Planning A 21*, 27–46