

Automating the administration boundary design process using hierarchical spatial reasoning theory and geographic information systems

Serryn Eagleson

PhD Candidate, Department of Geomatics
The University of Melbourne

Francisco Escobar

Senior Lecturer, Department of Geospatial Science
RMIT University

Ian Williamson

Professor of Surveying and Land Information, Department of Geomatics
The University of Melbourne

Keywords

Geographical information systems (GIS), hierarchical spatial reasoning (HSR), spatial data infrastructure (SDI), administration boundaries, postcode, census collection district (CCD).

Acknowledgements

The authors gratefully acknowledge the support of Land Victoria (LV), the Victorian Government, the Australian Research Council (ARC, Grant No. C00106797), National Mapping Division, Geoscience Australia and the members of the Centre for Spatial Data Infrastructures and Land Administration at the Department of Geomatics, the University of Melbourne, in the preparation of this paper and the associated research. The views expressed in the paper are those of the authors and do not necessarily reflect the views of LV, ARC, National Mapping Division, Geoscience Australia or the Department of Geomatics.

The authors also wish to acknowledge the encouragement and support of Professor Don Grant (former Surveyor General of New South Wales, Australia, and Chairman of the Public Sector Mapping Agencies of Australia) in the investigations into spatial hierarchy problems.

Author for correspondence:

Prof Ian Williamson
Department of Geomatics
The University of Melbourne, VIC 3010, Australia
Telephone +61-3-8344 4431
Fax +61-3-9347 4128
Email: ianpw@unimelb.edu.au

Automating the administration boundary design process using hierarchical spatial reasoning theory and geographic information systems

Abstract

This paper addresses the problems associated with the integration of data between incongruent boundary systems. Currently, the majority of spatial boundaries are designed in an uncoordinated manner with individual organisations generating individual boundaries to meet individual needs. As a result, current technologies for analysing geospatial information, such as geographic information systems (GISs), are not reaching their full potential.

In response to the problem of uncoordinated boundaries, the authors present an algorithm for the hierarchical structuring of administrative boundaries. This algorithm applies hierarchical spatial reasoning (HSR) theory to the automated structuring of polygons. In turn, these structured boundary systems facilitate accurate data integration and analysis whilst meeting the spatial requirements of selected agencies.

The algorithm is presented in two parts. The first part outlines previous research undertaken by the authors into the delineation of administrative boundaries in metropolitan regions. The second part outlines the distinctly different constraints required for administrative-boundary design in rural areas. The formalisation of the algorithm has taken place in a GIS environment utilising Avenue, an object-orientated programming language that operates under ArcView, the desktop software developed and distributed by ESRI.

1 Introduction

Spatial data infrastructure (SDI) is an initiative intended to create an environment for easy and secure access to complete and consistent data sets (Rajabifard et al., 2000). One fundamental problem restricting the objectives of SDI is the fragmentation of data between incongruent boundary systems. Essentially, this problem has occurred because organisations historically hand drafted the majority of boundaries on hardcopy maps. With advances in technology, these hand-drafted maps have been digitised for incorporation into GIS, a technology for which they have not been adequately designed. As a result, administrative boundaries are a product of the era in which they were developed, and change is now required to meet the needs of geospatial information analysts. To achieve this goal, new methodologies for administrative design are required.

In response to the problem of incongruent boundaries, this paper focuses on providing a new generic model, based on hierarchical spatial reasoning (HSR) theory, for the delineation of administrative boundaries. In the past, the authors have focussed on the design of administrative boundaries in metropolitan regions (Eagleson et al., 2002); however, as Haslam-McKenzie (2001) highlights, the problem of incongruent boundaries in rural regions is heightened by the large variation in boundary sizes. For example, small-area census polygons in rural regions are generally much larger than those in urban areas. As a result of the large variation in the physical size and shape of rural boundaries, they are an odd assortment of "building blocks" to make comparisons with non-census spatial units. Therefore this research aims to introduce a hierarchical model to aid in the transparency between data layers — both horizontally between agencies and vertically between layers in the rural administrative-boundary hierarchy.

In presenting the new model for administrative-boundary design in rural regions, the paper expands on both the theory and technical design of administrative-boundary-hierarchy research previously published by the authors (Eagleson et al., 2002). The paper incorporates the properties of zero- and one-dimensional hierarchical structures, embedded within the proposed three-dimensional administrative-boundary hierarchy. This new model must also strike a balance between the increasing geospatial requirements of GIS users and the business rules of selected agencies in the rural landscape.

2 The spatial-hierarchy problem

Administrative boundaries are frequently used for the display and analysis of spatial information. Health, wealth and population distributions are all examples of spatial information commonly attached to administrative polygons. There are numerous advantages in using administrative boundaries for the collection and collation of data. For example, once the administrative boundaries are established, the data is easily collected and efficient to store. Even in the light of technological advancements in using other forms of geographic data, such as address point and line data, these data structures are still relatively expensive to produce, difficult to manipulate and require large amounts of memory to store (Rajabifard and Williamson, 2001). Many organisations are thus utilising established polygon-based administrative boundaries as a base for the collection and collation of spatial data. As the majority of different agency boundaries are incongruent, however, when the boundaries are superimposed they do not coincide. As a result, data remains fragmented between the different administrative units and technologies for analysing the data attributed to the administrative boundaries such as GIS are not able to utilise the data to its full potential.

Spatial hierarchy: the current situation

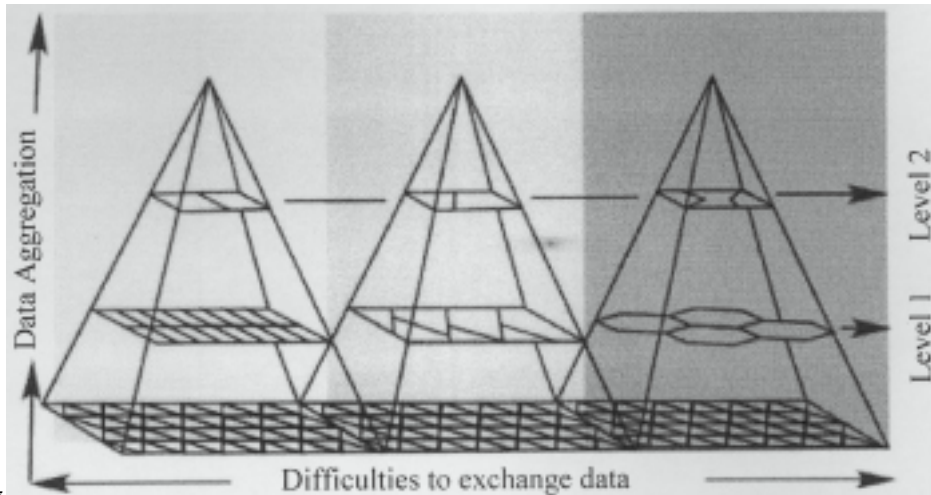


Figure 1. Current hierarchical spatial structures: the problem

Figure 1 illustrates the spatial-hierarchy problem where each agency uses the common baselayer (ie. land parcel) to establish a different size or shape of administrative unit, based on their unique requirements. In turn, each agency aggregates their administrative boundaries in a hierarchical fashion across the state. Data integration is possible for each organisation; however, statistical approximations must be employed to facilitate the cross analysis between the incongruent boundaries of different agencies (Eagleson et al., 2002).

3 Current solutions to facilitate data transfer between incongruent boundary layers

Although incompatible boundary units restrict GIS analysis, GIS technology is proving valuable in the formulation of solutions. As detailed below, researchers have investigated data interpolation, derived boundaries and data aggregation techniques, each aimed at alleviating problems associated with collecting and analysing data attached to incongruent boundary systems.

3.1 The first solution: data interpolation

Interpolation is a widely researched technique involving the transfer of attribute data between incongruent boundary systems. Data interpolation methods often utilise mathematical algorithms within the GIS environment for the transfer of data from the boundaries on a source map to boundaries on a target map (Goodchild, et al. 1993; Martin 1998; Trinidad and Crawford 1996). Although the interpolation process appears to provide a valid solution to the problem, many assumptions are made throughout the process. One, often invalid, assumption is that the distribution of values in the source map is constant (Goodchild, et al. 1993).

In an effort to increase the accuracy of interpolation, and minimise the number of assumptions, supplementary data such as road networks, land-use maps, satellite imagery and administration boundaries are often used as “control” for the interpolation process. Although interpolation techniques are valuable for providing a basis for analysis not currently possible

on a single boundary layer, the errors and assumptions inherent in the techniques result in a less than optimum solution.

3.2 The second solution: derived boundaries

In an attempt to make data readily usable, some organisations have created derived boundaries. Derived boundaries are formed through the aggregation of agency boundaries that approximately nest within more publicly recognisable administrative units. One prime example is the derived postcode, where for operational reasons the Australia Post postcode boundaries do not necessarily match the Australian Bureau of Statistics (ABS), census collection district (CCD) boundaries. In recognition of the separate functions undertaken by these agencies, the ABS aggregates CCDs to approximate the Australia Post postcode boundaries thereby producing ABS derived postal areas. Discrepancies between the boundaries of these two postal 'zones' can easily arise since the two systems are not coordinated. The derived postal areas may be quite different from the actual postcode boundaries both in terms of shape and area (See Figure 2). The two sets of spatial entities (postal zones) are nevertheless given the same identifier by the agencies, consequently leading to the misinterpretation of data by users. If users remain uninformed about the origin of the data boundaries, subsequent decisions will not be well supported. A discussion on this issue can be found in Jones, Eagleson, Escobar and Hunter (in press).

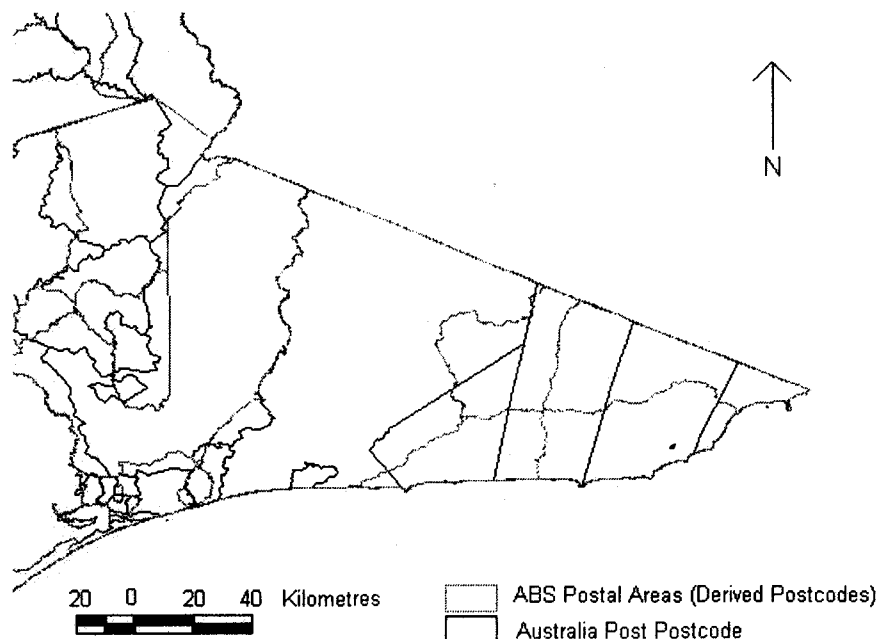


Figure 2. An illustration of two incongruent boundary systems: Australia Post postcodes and ABS postal areas (derived postcodes), (Eagleson et al. 2001).

3.3 The third solution: aggregation

A third method for the dissemination of data sets across incompatible boundary regions is the aggregation of point and polygon data.

Point aggregation

The aggregation of point data requires data to be stored at each parcel or address point and when required aggregated to a different spatial unit. Although the process of aggregation accurately solves the problem, other problems exist. First, this solution is not viable in Australia and many other countries, primarily due to stringent laws protecting confidentiality. For instance, once household data is collected by the ABS, it must be aggregated to the CCD boundaries (approximately 220 households) and the individual household data destroyed (ABS, 1996; Blanchfield, 2001). If confidentiality is not guaranteed, it is probable that people will not complete census forms truthfully, degrading the accuracy and reliability of census information for planning purposes.

Second, a large quantity of storage space is required to store data associated with individual land parcels, and each aggregation of data to new boundaries would be extremely time consuming and costly. Additionally, problems such as differencing exist when data is aggregated to a number of different boundaries. As Duke-Williams and Rees (1998) explain, if polygons containing confidential information are overlapping, in some circumstances it may be possible to subtract one set of polygons from the other to obtain statistics for subthreshold areas thus breaching confidentiality.

Polygon aggregation

The aggregation of polygon data involves the reaggregation of existing units into new spatial units, more suitable for specific analysis techniques. For example, Openshaw (1977) devised the automated-zone-design program (AZP) for investigating the modifiable area unit problem (MAUP). With the introduction of new technology, digital data and improved algorithms during the 1990s, AZP was further refined and extended forming the zone-design system ZDES (Openshaw and Rao, 1995; Openshaw and Alvandies, 1999). These zone design systems allow the data analyst the freedom to start with data at one scale and then reaggregate it to create a new set of regions designed to be suitable for a specific purpose, independent of the collection boundaries used (Openshaw and Rao, 1995). If these initial boundaries are not designed as layers within a hierarchy, however, the problem of data integration between overlapping polygons remains (Eagleson et al, 2002). As an alternative to the above-mentioned interpolation, derived boundary and aggregation approaches — which each have problems associated with confidentiality, accuracy and cost — this paper proposes the adoption of an HSR-based model for the redesign of administrative boundaries.

3.4 The proposed solution: boundary reorganisation

The solution proposed within this research involves the reorganisation of boundaries into a structured system based on HSR theory. The model adopts a common base layer to build individual hierarchical systems based on the properties of HSR. Figure 3 illustrates this approach. Each agency has spatial units that fulfil their individual requirements at the smallest scale, yet when aggregated, one common boundary is formed. Once contained within the second layer, data is easily transferred and cross analysed by each agency. Although this paper only deals with the integration of two agencies, Australia Post and the ABS, it is enough to demonstrate the concept.

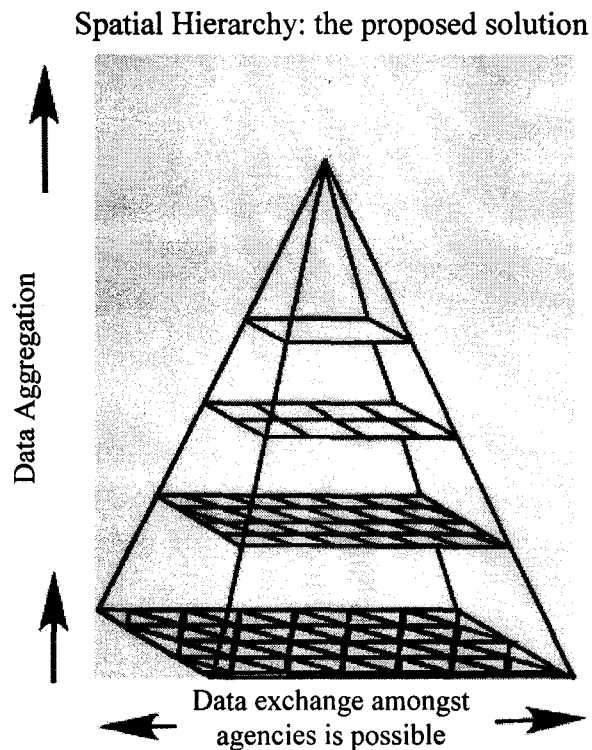
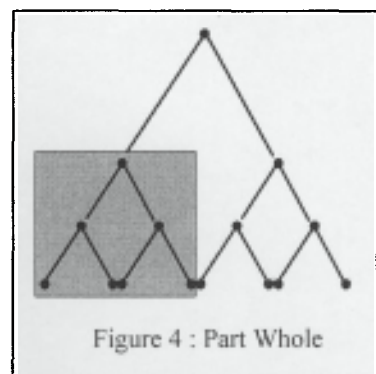


Figure 3. Future hierarchical spatial structures: the solution

3.5 Hierarchical spatial reasoning (HSR) theory

HSR is defined by Car (1997) as the part of spatial information theory that utilises the hierarchical structuring of space for efficient reasoning. It is through the works of Car (1997) for wayfinding, Glasgow (1995) for spatial planning, and Frank and Timpf (1994) devising the intelligent zoom, that this theory has been applied in the spatial industry. Consequently HSR theory has predominantly focused on zero- and one-dimensional structures to model urban systems (as points), road and drainage networks (as lines), and to a certain extent, simple bi-dimensional objects such as square polygons in quadrees (Rajabifard et al 2000). This research paper aims to extend the properties of HSR theory to include the unique relationship between elements and lower-order hierarchies inherent within a three-dimensional administrative-boundary hierarchy.

There are three global properties inherent to hierarchies that make them adaptable to boundary design. The first of these properties is *part-whole* (see Figure 4). This property relates directly to the relationship between elements as each element within the hierarchy forms a part of the elements on the layers above and also constitutes a whole of the elements below (Palmer, 1977). This property is directly related to administrative boundaries as each boundary is formed through the successive aggregation of smaller units to form a whole administrative unit. At the same time the administrative unit only forms a part of the overall system. For example,



provinces are a “whole” made out of localities but it constitutes a “part” if we look at the whole country.

The second property is the *Janus effect*. This property was first introduced by Koestler (1968) and is named after the Roman god with two faces. (See Figure 5) Each level in the hierarchy possesses two faces: one facing the levels below and one facing the levels above. In effect, each administration polygon has two faces: one looking to the smaller unit from which it is formed and a second looking towards the larger administration units it supports. For example, the CCD layer of boundaries has two faces: one facing the cadastral parcels that support it and one facing the statistical local areas (SLAs) that it supports.

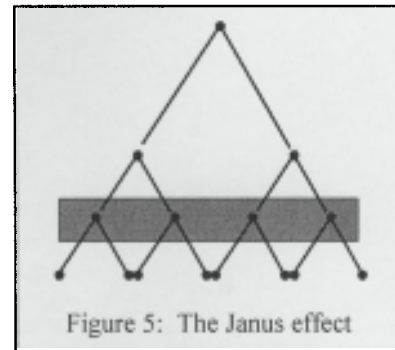


Figure 5: The Janus effect

The third property is *near decomposability*. This property is related to the nesting of systems and is based on the fact that interactions between various kinds of systems decrease as distance increases (Simon, 1973) (see Figure 6). Near decomposability also applies to administrative polygons as administrative polygons are often nested within one another from the parcel base through to the national border. Additionally, the relationship between elements decreases with distance. For example adjoining land parcels are more likely to share the same attributes, such as a common boundary and the same postal delivery service. In contrast it is unlikely that two land parcels located on the opposite sides of the state will share as many common attributes.

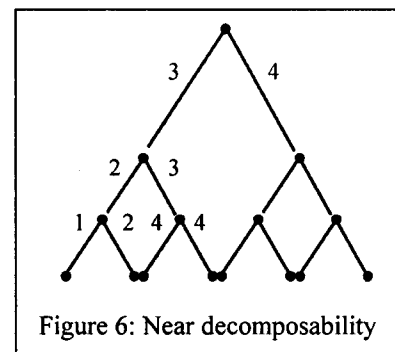


Figure 6: Near decomposability

The fourth property, *embeddedness*, relates to the development of a three-dimensional spatial hierarchy. Embeddedness implies that each lower-dimensional structure is always algebraically and geometrically “embedded” in the next higher-dimensional structure (Feltz, 2001). In the development of administrative boundaries, the model is constructed based on zero- and one-dimensional hierarchical structures, such as the town and road network. Consequently, the properties inherent within the town and road network hierarchies are embedded within the three-dimensional administrative boundary hierarchy.

4 Incorporating HSR theory within a model for the automated design of administrative boundaries

Due to the complexities of spatial entities, there are often many alternatives to the simplest of problems. The objective of HSR theory, therefore, is not to obtain an “optimum” result but one which meets user specifications at each level with minimal effort (Car 1997; Timpf and Frank 1997).

To be consistent with HSR theory and the geospatial requirements of administrative agencies, it is envisaged that this algorithm will need to have the ability to:

- a. automatically subdivide the territory in compliance with the geospatial requirements stipulated by the relevant agency; and

- b. be recursive and re-applicable to the outputs in order to produce new levels of the hierarchy (Eagleson et al., 2002).

5 HSR-based prototype development

In order to effectively integrate HSR theory and GIS technology for the design of administrative boundaries, a model incorporating the requirements of agencies at each layer of the hierarchy must be established. Hugo et al. (1997) have compiled data (see Table 1) detailing criteria for establishing small-unit boundaries in a number of countries.

Table 1. A list of criteria used or recommended by six countries for the delineation of basic spatial units (BSUs) used for the dissemination of statistics

Criteria	Countries
1 BSU must not cross any higher-order boundary	CFNSUZ
2 BSU boundaries must be clearly shown on maps	CFSUZ
3 Newly defined BSU boundaries must be as consistent as possible with those used in previous census	CNSUZ
4 BSU boundaries should separate out urban settlements from rural	CSUZ
5 BSU boundaries should conform to some population range	CNUZ
6 BSU boundaries must be clearly visible on the ground	CUZ
7 BSU areas should not be too large	NUZ
8 Road or communication line should form central artery of BSU, binding it together by giving accessibility	CNS
9 Larger uninhabited areas may/should form zero population BSUs	ZN
10 BSU should be homogeneous as possible in their physical and economic attributes	N
11 BSU should form connected agricultural areas suitable for agricultural planning	N
12 Each BSU must constitute a convenient collector workload	C
13 BSU are to consist of all polygons whose edges are formed by the intersection of visible linear features	U
14 Physical barriers like forests, ridges etc. should be used as BSU boundaries	S
15 BSU boundaries should be chosen to be acceptable to as-many government departments as possible	S

C = Canada, F = Finland, N = Norway, Z = New Zealand, U = USA and S = Switzerland.

(Source: Hugo et al., 1997)

As expected, the criteria vary between countries depending on their level of geocoding of individual parcel information and confidentiality restrictions (Hugo et al. 1997). Additionally, it has been concluded that within the study site of Victoria, Australia, metropolitan and rural landscapes are different, and as a result, so are the business rules of the selected agencies operating within each of these regions. Thus the following section of the paper highlights the

differences between these two landscapes, and the business rules of the administrative agencies operating in these regions.

6 A summary of the metropolitan boundary project

Previously, research has been undertaken by the authors into the delineation of administrative boundaries within metropolitan regions (Eagleson et al., 2002). This research involved the formalisation of business rules established by two administrative agencies (namely, Australia Post and the ABS) within a prototype for the automated delineation of administrative boundaries.

Due to the uniform density of the population in metropolitan regions, road centrelines were used to form meshblocks* for the aggregation of boundaries according to a range of agency constraints. These constraints were based on the requirements of Australia Post and the ABS and include:

- Topographic barriers are to be preserved. Examples of barriers include large rivers and roads that may obstruct delivery mechanisms. Additionally, these boundaries often divide different community groups. Because preserving the difference between community groups is imperative to many planning activities, topographic features, which appear as barriers, are preserved within the prototype.
- In order to protect confidentiality, the ABS states that each CCD unit must cover approximately 200 households (ABS, 1996).
- To facilitate delivery mechanisms, it is important that the boundaries are in alignment with the road network and are identifiable on the ground. Using the meshblock for the aggregation assures this constraint is met.
- To ensure uniformity across the area, it is important that the boundaries are contiguous and provide complete coverage across the area without gaps or overlaps.
- Although there is no formal definition of boundary shape, it was decided that the boundaries should be constructed in a manner that enabled them to be compact. To ensure the boundaries established are compact, the model tests each boundary based on the circularity index devised by Tomlin (1992).

(Eagleson, et al., 2001).

In undertaking this initial research, it became apparent that it is possible to design a spatial hierarchy of administrative boundaries within the metropolitan landscape; however, metropolitan and rural landscapes are very different. In direct contrast to the metropolitan landscape, the rural landscape is large, parcel sizes vary and roads often unite rural communities instead of dividing them. Table 2 highlights some of the differences in the infrastructure of rural and metropolitan landscapes in Victoria.

* Meshblock: a spatial unit formed through the intersection of road centreline data.

Table 2. Differences between the rural and urban landscape

	Urban	Rural
Address point database:	Complete	Incomplete
Cadastral parcel size:	Typically small	Highly variable
Large topographic features:	Divide communities	Unite and divide communities
Administrative boundary: accuracy	High	Low
Meshblock:	Uniform unit size	Variable unit size
Roads:	Divide Communities	Unite Communities

7 The function of administrative agencies in rural environments

7.1 Australian Bureau of Statistics census collector district boundaries

In Australia, some organisations are currently utilising hierarchically ordered units for the collection and dissemination of data. The ABS has defined a unique hierarchy for the collection of demographic data called the Australian Standard Geographic Classification (ASGC). The ASGC divides Australia into numerous hierarchical levels to facilitate statistical functions. These levels are based on six interrelated classification structures. These are:

- main structure;
- local government area structure;
- statistical district structure;
- statistical region structure;
- urban centre/locality structure; and
- section-of-the-state structure. (ABS, 1996)

The main, statistical region and section-of-the-state structures each cover the whole of Australia with out gaps or overlaps. The remaining structures cover only part of Australia. In the formation of the ASGC, the smallest of the spatially defined units is the CCD. Consequently, the CCD has been used in the aggregation of the six classification structures outlined above. Due to the importance of the CCD in underpinning the ASGC hierarchy, this research incorporates this unit and the first aggregated layer of the hierarchy.

In designing CCD boundaries, there are a number of specific criteria that are imperative to their effectiveness — both operationally and as display units for the analysis of statistics such as population, income, health and culture. To be effective as CCDs, it must be possible for census collectors to cover the entire area of the CCD within a two-day period. To guarantee confidentiality, each CCD must contain an aggregation of approximately 220 households (Blanchfield, 2001). To reach this number of dwellings in the rural environment is often a practical impossibility. However, in these instances, the ABS imposes confidentiality at the processing stage by suppressing any cell that has less than three contributors (Blanchfield, 2001). For example, in a table of income by occupation, if there were only two teachers within the CCD the income for this group would be suppressed.

In addition to maintaining confidentiality for the CCD boundaries, to be effective for the display of demographic related data, the boundary design must ensure the heterogeneity of the rural population can be represented. One way to achieve this is through the appropriate use of road centrelines for delineation. In the past, road centrelines were used in the delineation of CCDs, dividing rural communities of similar constitution and combining them with the diverse outer rural regions. As a result, the overall aggregation of population statistics to the demographic boundaries reveals homogeneity between units, when in fact they are very different. Therefore the design of CCD boundaries must aim to reflect the heterogeneity of rural communities whilst ensuring confidentiality restrictions are maintained.

7.2 Australia Post postcode boundaries

Australia Post has derived postcode boundaries to facilitate the delivery of mail. The initial allocation of postcode boundaries was on a somewhat *ad hoc* basis, and the allocation method varied between states. Today postcodes are (where possible) being brought into alignment with suburb and locality boundaries. Postcodes are extremely important administrative units as they are often used as a common identifier for the aggregation and storage of different information types. For example: "...with postcodes you can locate people and see the hows, wheres and whys of markets, customers and prospects, competitors, prices, suppliers, routes and profits. Postcodes neatly define convenient demographic zones and are familiar to everyone." (AUSLIG, 2000).

7.3 Summary of rural administrative boundary constraints

When defining spatial boundaries it is important that the administration units created are not only functional, but may also be used for the display and analysis of a wide number of social and economic characteristics. In an attempt to meet these requirements, in rural areas the following constraints have been established:

- The primary importance to each agency is the preservation of town boundaries. Towns shall therefore be delineated according to *population density*. Alternatively, if town or locality boundaries exist, these will be used as input into the model.
- Individual catchment areas around each *town* are important within rural regions. These catchments often link the populations outside of towns with the town and necessary services. For planning purposes, therefore, it is important to consider the requirements of the surrounding regions.
- As each administrative agency uses the boundaries to facilitate the distribution of resources, the *distance* travelled along the road network within a unit will be crucial for determining functional administrative boundaries. The ABS specifies that for CCDs to be effective as collection districts it must be possible for census collectors to cover the entire area of the CCD within a two-day period (ABS, 1996; Blanchfield, 2001).
- Large topographic features that form *natural barriers* between communities should be preserved. As Morphet (1993) explains, major topographic features not only present barriers that limit routing, but they often segment demographic classes. It is thus important to ensure that major topographical barriers are preserved in order to facilitate accurate statistical analysis.
- As outlined in section 7.1 road centrelines are important in the delineation of rural administrative boundaries. In the past, road centerlines were used as dividing features in the delineation of CCDs, often dividing communities and combining them with the outer rural regions. Consequently, the overall aggregation of population statistics to the

boundaries reflected a demographic homogeneity between units when in fact they were very different. This research project utilises road centrelines as *uniting features* in the rural landscape. The additional benefit of incorporating road centreline data as uniting features in administrative boundary design, is the implicit representation of major non-traversable topographical barriers such as mountains and rivers (Zoltners and Sinha, 1983).

- To ensure *complete coverage* across the region, the newly formed administrative boundary layer shall be contiguous across the state without gaps or overlaps.

8 Prototype development

8.1 Algorithm mechanics

The implementation of a spatial hierarchy requires a technical solution. This section of the paper details the boundary-allocation prototype that has been developed. The workings of the prototype are hidden from the user. An algorithm runs through a series of routines ensuring that the requirements of agencies are met. Steps 1–3 outline the datasets required as input to the prototype. Steps 4–6 detail each of the decisions and processes operating behind the prototype.

Step 1: Rural regions (input 1)

To obtain the rural data it is important to filter out any regions that are classified as urban. In this research, predefined locality boundaries are used to filter out the urban areas. The locality boundaries have been obtained from Land Victoria, which maintains a representation of the State's suburb, town and rural-district boundaries. The dataset contains representations of suburb, town and rural-district names and boundaries as approved by the Registrar of Geographic Names and published in the Victoria Government Gazette (Land Victoria, 2001). (See Figure 7.) To prepare the data as input into the prototype the road layer is clipped based on these locality boundaries. As a result, the metropolitan model (Eagleson et al., 2002) can be used to design boundaries within the metropolitan regions, whilst the remaining rural regions can be segmented based on the rural model as detailed in the prototype.

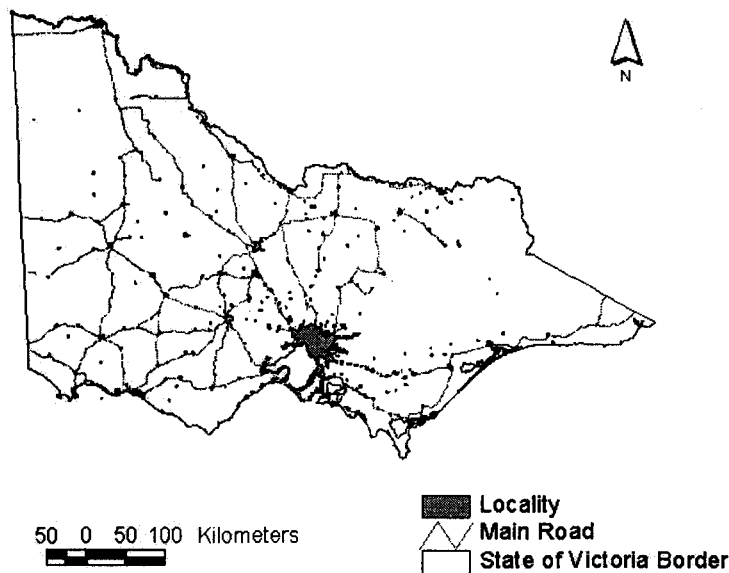


Figure 7. Locality boundaries in Victoria

Step 2: The road network (input 2)

As discussed in section 7.3, the road network is fundamental to the design of rural administrative boundaries. Within the prototype, the road network is converted to a grid, and each cell within the grid is assigned an impedance value. This value is a unit measurement that depicts the cost involved in moving through each cell. The value of each cell in the cost grid is assumed to represent the cost per unit of distance of passing through the cell, where a unit distance corresponds to the cell width (ESRI, 2001). Within the model, the impedance value is assigned based on the road hierarchy, where highways are 1, major roads are 2 and minor roads are 3.

Step 3: Cadastre (input 3)

Within this research paper the cadastre is used as the base layer for all administrative-boundary design. The following points highlight the importance of the cadastre as an input layer into the boundary allocation prototype:

- A spatial hierarchy can only be broken down to the smallest unit from which it is composed. From an administrative-boundary perspective, this smallest unit is the cadastral land parcel.
- Dale and McLaughlin (1988) argue: “Due to the fact that almost all activities take place within a land parcel, it makes sense that this parcel should form the basic spatial unit for social, economic, administration and other boundaries”.
- It would be impractical for any parcel to be dissected by an administration boundary as this would render the parcel with two postcodes or two collector-district numbers. Hugo et al. (1997) reinforce the usage of the cadastre as the primary unit on the basis that cadastral boundaries are already unequivocally determined on the ground. The ability to identify the boundaries on the ground is an important factor as boundaries will be used to facilitate the collection of data including display and dissemination. To ensure maximum

efficiency of the cadastre within the model, it is thus important that the cadastre is accurately maintained and that each parcel is structured as a polygon.

Once the input data has been prepared as detailed above, the boundary-allocation algorithm can be run. This algorithm is detailed in steps 4–6 below.

Step 4: Cost–distance allocation

The boundary allocation algorithm firstly initiates the ESRI-developed cost–distance subroutine. This subroutine calculates the least-accumulative-cost distance to a designated source for each cell.

The subroutine requires two input grids. The first input grid is a source grid. In this instance the source grid is the town; however, agencies could use their distribution points. The second input grid is a cost grid. For the purpose of this application, the cost grid is based on the road network incorporating the costs associated with travelling along the road network.

Using both the source and cost grids, the algorithm is then able to calculate the least-accumulative-cost distance, assigning each element of the road network to a town. To determine the accumulative cost of moving from cells in the road network to cells in the source grid, the algorithm utilises a node-and-link cell representation. In the node-and-link representation, each centre of a cell is considered a node, and each node is connected by links to its adjacent nodes. (See Figure 8.)

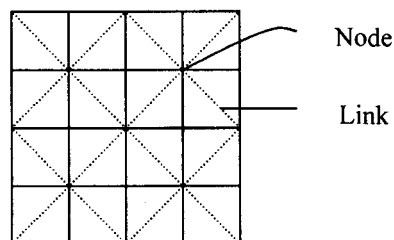


Figure 8. Nodes and links: a view of a grid through graph theory (ESRI, 2001)

Creating an accumulative cost-distance grid using graph theory is an iterative process. In the first iteration, the source cells are identified and assigned the value zero, since there is no accumulative cost to return to themselves. Next, all the source cell's neighbours are activated, and a cost is assigned to the links between the source cell's nodes and the neighbourhood cells' nodes based on accumulative-cost formulas. (See ESRI, 2001.) Each of these neighbourhood cells are then assigned a value based on the accumulative cost to reach a source. The accumulative values are arranged in a list from the lowest accumulative cost to the highest. The result of this process is a grid identifying which cells will be allocated to which source on the basis of the lowest accumulative cost to reach a source (ESRI, 2001).

By taking this approach, each segment of the road network is allocated to a town based on the most cost-effective route along the road network.

Step 5: Road proximity boundaries

Following the assignment of road segments to distribution points in step 4, a road proximity grid is established. Each cell is assigned the value of the nearest road network feature to create

the grid. Once the grid is established, the boundary-allocation algorithm converts the grid from a raster to a vector shapefile (Figure 9).

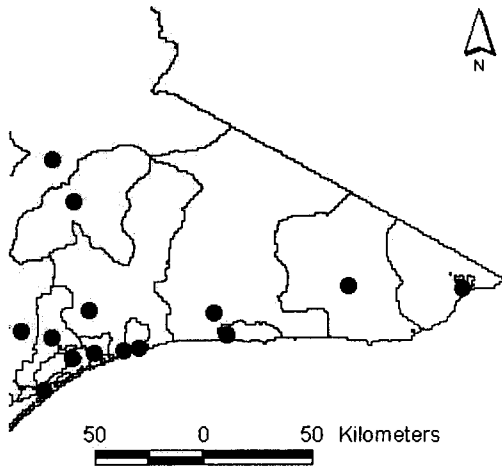


Figure 9. Boundaries are established for each distribution point.

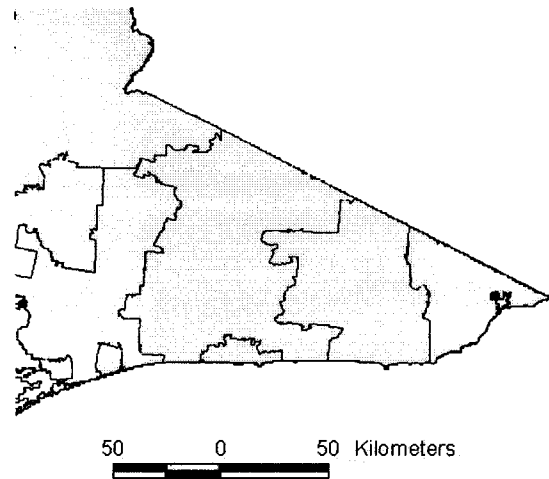


Figure 10. Boundaries formed in step 5 aligned with the cadastre.

Step 6: Boundary Alignment

To finalise the boundaries so they are practical and well defined on the ground, the boundary-alignment subroutine aligns the boundaries designed in step 5 with the cadastre. This process facilitates the alignment of boundaries from the smallest spatial unit, in this instance the cadastre, to the newly established boundaries. (See Figure 10.)

Once the boundary-allocation algorithm has been run, the result is a new administrative boundary layer. This layer can be used as an input layer for the next administrative boundary layer, as required by the operator. At this point, the operator may add additional data sets or criteria to the boundary-allocation process, meeting the next set of requirements at the next layer of the hierarchy.

8.2 Advantages of automated boundary design within an HSR-based model.

Hierarchical structures are imposed on space to facilitate efficient spatial reasoning. The approach adopted within this research was to utilise and expand upon current HSR theory for the purpose of administrative-boundary design. As discussed in section 3.5, in the past, HSR theory has predominantly focused on zero- and one-dimensional structures. The design of an administrative-boundary hierarchy is three dimensional, however, and must incorporate the properties of the zero- and one-dimensional structures (such as the town regions and road network structures), which through the process of aggregation, become embedded within each subsequent layer of the hierarchy. It is important to note that from the data-analysis perspective the behaviour of a layer within the hierarchy is often a reflection of the relationship between elements and lower-dimensional hierarchies that are embedded within it.

In order to comply with HSR theory, the prototype has been derived using a model through which a flat non-hierarchical system (i.e. cadastre) can be transformed using business rules in the development of new administration boundaries. The development of the rural prototype demonstrates the benefits of automated boundary design as each new boundary unit is created

in a co-ordinated systematic and rigorous manner. Additionally, the process is repeatable and can be used to generate higher-order boundary layers depending upon agency constraints.

In comparison to the metropolitan model derived in earlier research, which was based on the iterative aggregation of units, this algorithm, derived for the allocation of rural administrative boundaries, is unique. The model is based on a raster data structure, which is fast to operate, and takes into account many of the constraints applicable only in sparsely populated rural landscapes. As a result, the algorithm aims to complement previous research and facilitate effective data management so that the full potential of geospatial data collected within rural environments can be realised.

8.3 The Modifiable Area Unit Problem

The MAUP is classic problem associated with the design and display of boundaries. The MAUP is “a form of ecological fallacy associated with the aggregation of data into areal units for geographical analysis. This aggregated data is then treated as individuals in analysis” (Openshaw and Taylor, 1981). An example of this process is census data, which is collected from every household but released only at CCD boundaries. When values are averaged through the process of aggregation, variability in the dataset is lost, and values of statistics computed at various resolutions will be different. This is called the *scale effect*. Additionally, the data analyst gets different results depending on how the spatial aggregation occurs. This is called the *zoning effect*. This problem is integral to the display of demographic data as the information displayed is a product of the size, shape and scale of the administrative boundaries used in the data collection (Fotheringham and Wong, 1991). In the past, because boundaries were assumed to be fixed, researchers had to use whatever boundaries were available (Openshaw et al., 1998). Consequently, the user has little, if any, control over the MAUP.

It is expected that the proposed hierarchical-reorganisation model will allow GIS analysts to have greater flexibility over the level of aggregation and zoning configuration of spatial units used in analysis. Using a common base layer, the model allows analysts to start from the smallest divisions available, or the smallest they can process and aggregate these in a fashion relevant to their investigation (Openshaw and Taylor, 1979). This freedom to re-aggregate spatial units gives the GIS analyst greater control over the MAUP, which has not been possible in the past.

8.4 Implementation Issues

The aim of this research project has been to develop new methods through which space can be divided into administrative boundaries in a structured manner. In undertaking this research, it has become clear that it is possible to align administrative boundaries based on HSR theory. This was further supported by the development of prototypes for the automated delineation of administration boundaries using GIS in both urban and rural environments. Whilst the theoretical framework is strong, there are some technical and institutional issues requiring further investigation to ensure the model can be effectively implemented.

Technical Issues

As established in section 3.4, the boundary-allocation algorithm for the rural application is required to meet the criteria of the two agencies under investigation; i.e. Australia Post and the ABS. From a practical perspective, the boundary-allocation program meets these criteria. The town catchment areas are incorporated within the cost-distance algorithm (step 4). The

boundaries are contiguous and complete, without gaps or overlaps, and once clipped to the cadastre, they are identifiable on the ground.

From a mathematical perspective, however, the boundary allocation program does not fulfil the criteria of distance to be travelled by an agency. The reason for this is that each boundary is established based on the position of the town. Practically, the solution to this would be for the program to assign a new point source of distribution for the agency. Due to the sparseness of the rural landscape, however, it is not practical for agencies to set up new distribution points in sparsely populated regions; and therefore, it is more practical for the agency to be flexible with this criterion.

The model derived in this research is highly dependent upon accurate and current data sets as input for data analysis; however, it is recognised that this level of data sophistication may not always be available for boundary design. Alternative methods of boundary design may thus be required to automate the boundary-design process in regions where accurate spatial data does not exist. Similarly, these models must be flexible so that, as data does become available, it can be incorporated within the model.

Technology is impacting the way agencies do business. For example the Internet has been suggested as a future tool to conduct censuses (Mobbs, 1998). If the Internet does become the primary source of data collection, the boundary-delineation criteria adopted for establishing boundaries will no longer need to consider the distance and time taken by census collectors. The method established for boundary design will thus need to contain a mechanism for recording boundary change over time, allowing analysts to track changes and accomplish more accurate time-series analysis than currently available.

Additionally, future research is required to understand the social relationships between rural communities and their spatial boundaries. As a result of this research it is anticipated that the model will be refined to incorporate social issues in boundary delineation, such as the identity of place and spatial cognition.

Institutional Issues

Developing a technical solution alone cannot ensure the development of a statewide hierarchy of administrative boundaries. Unless institutional and political issues are addressed, administrative boundaries will continue to be developed by individual agencies, further compounding the spatial-hierarchy problem. In order to begin addressing these issues, a better understanding of the complex nature of SDIs is required together with an extension of their current function so as to promote the implementation of new methods for designing administrative boundaries.

Additionally, it is proposed that incentives for agencies to participate in the hierarchical design framework need to be established. These incentives may include the accreditation of agencies establishing boundaries within the hierarchy standard and/or benchmarking administrative-boundary hierarchies to assess the comparative effectiveness of the systems.

9 Conclusion

This paper identifies current problems associated with incongruent boundary systems. In response to these problems the primary objective of this research has been to develop a new

method through which space can be divided into administrative boundaries in a structured manner. In contrast to metropolitan regions, the attributes of the rural landscape are vastly different. As a result, the model developed aims to incorporate the differences in the rural landscape along with agency requirements for the effective development of a rural spatial hierarchy.

The proposed solution involves the reorganisation of the spatial environment using the properties of HSR theory. To achieve an administrative boundary hierarchy, the properties of HSR theory have been extended to incorporate the complexities of zero- and one-dimensional point and line data structures within a three-dimensional administrative boundary hierarchy.

This model has been formalised within the GIS environment by automating the process of administrative-boundary design. By using this approach, it is anticipated that administrative, political and commercial marketing boundaries can be formed through the aggregation of smaller units, where the smallest spatial unit is the land parcel. Importantly, this system would enable rapid and efficient cross analysis between data sets.

10 References

AUSLIG, 2001, Australian Spatial Data Infrastructure, *AUSLIG*.

<http://www.auslig.gov.au/asdi/> (17/10/2001, 2001).

Australian Bureau of Statistics, 1996, *Statistical Geography Volume 2 Census Geographic Areas Australia*, Commonwealth of Australia.

Blanchfield, F., 2001, AURISA paper *Personal Communication* (4/12/2001)

Car, A., 1997, *Hierarchical Spatial Reasoning: Theoretical Consideration and its Application to Modeling Wayfinding*, PhD Thesis, Department of Geoinformation, Technical University, Vienna.

Dale, P.D. and McLaughlin, J.D., 1988, *Land Information Management*, Clarendon Press, Oxford.

Duke-Williams, O. and Rees, P., 1998, Can Census Offices publish statistics for more than one small area geography? An analysis of the differencing problem in statistical disclosure. *International Journal Geographical Information Science*, 12,6, 579-605.

Eagleson, S., Escobar, F. and Williamson, I.P., 2001, Developing a theoretical framework for the delineation of administrative boundaries within a rural context, In *Proceedings of AURISA 2001 - The Grand Hyatt, Melbourne, VIC, 19-23 November (CD-ROM)*.

Eagleson, S., Escobar, F. and Williamson, I.P. 2002, Hierarchical Spatial Reasoning Theory and GIS Technology Applied to the Automated Delineation of Administrative Boundaries. *Computers, Environment and Urban Systems*, 26, 185-200.

ESRI., 2001, Cost Distance Modeling Discussion, *ArcView Online Help*, Environmental Systems Research Institute, Inc.

- Feltz M., 2001, Essay #1: A Hierarchy of Structures - Circle/Sphere/Hypersphere (<http://www.geocities.com/CapeCanaveral/7997/essay1.html>)
- Fotheringham, A. and Wong, D. (1991), 'The Modifiable Area Unit Problem', *Environment and Planning A*, vol. 23, pp. 1025-1044.
- Frank, A.U. and Timpf, S. 1994, 'Multiple Representations for cartographic objects in a multi-scale tree - an intelligent graphical zoom'. In *Computers and Graphics; Special Issue on Modelling and Visualization of Spatial Data in GIS*, 18, 6, 823-829.
- Glasgow, J., 1995, A Formalism for Model-Based Spatial Planning. In *Spatial Information Theory-A Theoretical basis for GIS*, edited by, A. Frank and W. Kuhn, Berlin-Heilderberg-New York: Springer. pp. 501-518.
- Goodchild, M., Anselin, L. and Deichmann, U., 1993, A framework for the areal interpolation of socioeconomic data. *Environment and Planning A*, 25, pp. 383-397.
- Haslam-McKenzie, F. M., 2001, Statistical boundaries: a means by which the realities of rural decline in the Western Australian wheat-belt has been hidden. <http://www.csu.edu.au/research/crsr/conferen/abstracts/haslam2.htm> (Date of Access 22/06/2001)
- Hugo, G., Griffith, D., Rees, P., Smailes, P., Badcock, B. and Stimson, R., 1997, Rethinking the ASGC: Some Conceptual and Practical Issues. In *Final Report for Review of the Australian Standard Geographical Classification Project*, National Key Centre for Teaching and Research in Social Applications of Geographical Information Systems, University of Adelaide.
- Jones, S. D., Eagleson, S., Escobar, F. and Hunter, G. J., in press, Lost in the mail: the inherent errors of mapping Australia Post postcodes to ABS derived postal areas, *Australian Geographical Studies*. In Press.
- Koestler, A., 1968, *Das Gespant in der Maschine (A ghost in the machine)*, Wien-Munchen-Zurich, Verlag Fritz Molden.
- Land Victoria., 2001, *LGA, Locality and Suburb Boundaries*. (<http://www.giconnections.vic.gov.au/content/digital/lgabdy/lga.htm>), (Date of Access 13/11/2001).
- Martin, D. 1998, Optimizing census geography: the separation of collection and output geographies. *International Journal of Geographical Information Science* 12, 7, pp. 673-685.
- Mobbs, J., 1998, Australia comes to its census. *FIG 7*, Melbourne Australia (web access <http://www.sli.unimelb.edu.au/fig7/Brighton98/Comm7Papers/SS34-Mobbs.html>).
- Morphet, C., 1993, The mapping of small area census data – a consideration of the effects of enumeration district boundaries, In *Environment and Planning A*, 25, 9, pp.1267-1277.
- Openshaw, S. and Alvandies, S., 1999, Applying geocomputation to the analysis of spatial distributions. In *Geographic Information Systems: Principles and Technical Issues*.

- Longley P, Goodchild M, Maguire D, Rhind, D (eds), (New York: John Wiley and Sons Inc).
- Openshaw, S., Alvanides, S. and Whalley, S., 1998, Some further experiments with designing output areas for the 2001 UK census Paper presented at the 4th of the ESRC/JISC supported workshops.
- Openshaw, S., 1977, A geographical solution to scale and aggregation problems in region-building, partitioning, and spatial modelling. In *Trans. Inst of British Geographers New Series 2*, pp.459-472.
- Openshaw, S. and Rao, L., 1995, Algorithms for reengineering 1991 Census Geography, In *Environment and Planning A. 27*, pp.425-446.
- Openshaw, S., and 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*, N.Wrigley, (ed), (Pion: London)
- Openshaw, S., and Taylor, P.J., 1981, The modifiable areal unit problem, in *Quantitative Geography: A British View*, N. Wrigley and R. J. Bennett, (eds), (Routledge: London).
- Palmer, E. S., 1977, Hierarchical Structure in Perceptual Representation. In *Cognitive Psychology*, 9, pp.441-474.
- Rajabifard, A., Williamson, I.P., Holland, P. and Johstone, G., 2000, From Local to Global SDI initiatives: a pyramid building blocks, In *Proceedings of the 4th GSDI Conference, Cape Town South Africa*, (<http://www.gsd.org/docs.html>, Date of Access 26/04/2000).
- Rajabifard, A. and Williamson, I. P. (2001), Regional Administrative Boundaries Pilot Projec', Report Presented at *7th Meeting of the Permanent Committee on GIS Infrastructure for Asia and the Pacific (PCGIAP)*, 23-27 April 2001, Tsukuba, Japan.
- Simon, H. A., 1973, The organisation of complex systems. In *Hierarchy Theory*, H. Pattee (ed), (New York : Braziller), pp.1-27.
- Timpf, S., and Frank, A. U., 1997, Using hierarchical spatial data structures for hierarchical spatial reasoning. In *Spatial Information Theory - A Theoretical Basis for GIS (Int. Conference COSIT'97)*, Hirtle, S.C., and Frank, A.U., (eds), Lecture Notes in Computer Science 1329. Heidelberg: Springer-Verlag, pp.69-83.
- Tomlin, C., 1992, *Geographic Information Systems and Cartographic Modeling* (Prentice Hall: Englewood Cliffs).
- Trinidad, G. and Crawford, J., 1996, A tool for transferring attributes across thematic maps. In *AM/FM*, pp.127-137.
- Zoltners, A. and Sinha P., 1983, Sales Territory Alignment a Review and Model. *Management Science* 29, 11, pp.1237-12457.