The Visualisation of Uncertainty in Spatially-Referenced Attribute Data using TRUSTworthy Data Structures

Julian Kardos[1,2], Antoni Moore[1] & George Benwell[1]

[1]Spatial Information Research Centre
University of Otago. Dunedin, New Zealand
Phone: +64 3 479-7391 Fax: +64 3 479-8311
Email: jkardos@infoscience.otago.ac.nz

[2] Present Address: IAE,
Université Jean Moulin Lyon 3,
69365 Lyon Cedex 08, France.
Phone: +33 6 23 48 04 80

Presented at SIRC 2003 – The 15th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
December 1-2nd 2003

ABSTRACT

This paper presents the use of hierarchical spatial data structures to visualise attribute and spatial uncertainty when using spatial information systems. This is demonstrated using spatially-referenced data from the New Zealand 2001 Census. Firstly, selected current spatial visualisations created to show uncertainty were assessed in an Internet survey, revealing that overall there is not many *usable* techniques and also users want to see the original information display as well as a display of uncertainty. This forms a background to the other purpose of this paper, to discuss hierarchical tree structures as a potential visualisation-of-uncertainty technique for socioeconomic data. The major question is: can an uncertainty be effectively communicated using data structures whilst simultaneously displaying the attribute information? Two such structures were compared and assessed in another survey: the region quadtree and the Hexagonal or Rhombus (HoR) quadtree, both variable resolution structures. These structures work in the following way: an area where attribute data is uncertain will show less resolution through the data structure than an area that is not, exemplifying a level of detail metaphor. Data structures can disseminate aggregated census data and standardise spatial units, thus reducing subjectiveness in boundary definitions for socioeconomic data. The principles behind possible implementations are presented through a demonstration program, TRUST (The Representation of Uncertainty using Scale-unspecific Tessellations). The results of the second survey revealed that the region quadtree and the HoR quadtree can display uncertainty information to a user, and also that the region quadtree is more visually complicated. The HoR tessellation ranked similar to common visualisation-of-uncertainty techniques like overlay and blinking pixels.

Keywords and phrases**:** Visualisation-of-Uncertainty, Attribute, Spatial, Tessellation, Decision making.

1.0 INTRODUCTION

Spatial and attribute information have an inherent associated uncertainty that should be expressed to a potential user. Decisions are made using information displays riddled with data generalisation and inaccuracies; this is more prevalent when using socioeconomic data (see Martin 1996). Moreover, the information shown on a map is represented as being precise, and few methods exist for displaying uncertainty (Goodchild 2000). Inaccuracies arise from temporal, spatial and attribute uncertainty. Census data for example, is only collected every five years in New Zealand, therefore temporal uncertainty will be prominent and will increase with elapsed time from the collection date. Temporal uncertainty is normally ignored due to the expense of running localised surveys (Agumya & Hunter 2002). Spatial uncertainty arises when data is aggregated into census mesh blocks to preserve confidentiality. These mesh blocks are created for ease of enumeration rather than to display socioeconomic patterns over the geography, therefore the patterns become more a function of the zone boundaries (Martin 1996). Attribute uncertainty can propagate when census participants are missed; they can input wrong data or their data can be collected twice. It was only in the early 1990's that GIS users as a whole started to take notice of spatial and attribute uncertainty. As GIS has evolved, uncertainty issues have become more apparent to users and uncertainty is now a prominent topic in the GIS field. A recent special issue of Cartography and Geographic Information Science focused on the research agenda of 'Geovisualisation' (MacEachren & Kraak 2001). In this special issue Fairbairn et al. (2001) proposed that a representation of uncertainty was a key research challenge, and could be characterised as a new component of data. Fairbairn et al. (2001 p.20) continues, "*Making information available about data uncertainty, [and ensuring] the suitability of a representation for a particular task is essential, if users are to make informed decisions and we are to extend the visualization toolkit. A comprehensive program of research is needed to ensure the development and test (in a variety of circumstances) efficiency of quality or uncertainty indicators for new methods of representation.*"

This key research challenge was a driving force for this research, which introduces a new method to represent attribute and spatial data uncertainty. Taking the example of census data stored in mesh blocks and displayed as choropleth maps, hierarchical data structures such as quadtrees can be used to represent uncertainty as a transparent overlay. This overlay would appear on top of the attribute information choropleth display with only outlines of the quadtree cells showing. An area where attribute data is uncertain will show less resolution through the data structure than an area that is not. This will create a multi-resolution image, expressing uncertain areas through less resolution. Some of the first literature by Rosenfeld and Samet (1985) looking into tree structures for region representation expressed them as "variable-resolution arrays" representing detail only when it is available. Current geographic information systems have vast amounts of data and detail available, but the accuracy of these datasets can be limited. This research intends to revisit tree structures and use them in a hitherto unexplored mode to express uncertainty in GIS, by showing uncertainty at the resolution applicable to the spatial and attribute error possessed by the actual displayed information. The rest of this paper is organised as follows: Section 2 explains uncertainty in spatial systems, uncertainty models and how an uncertainty measure was derived for this research. Section 3 presents some current visualisation-of-uncertainty techniques and results from a survey about their usability and overall effectiveness. Section 4 attempts to define the metaphoric reasoning behind using variable-resolution tessellations to view uncertainty in spatial information. Section 5 proposes a new approach to visualising attribute uncertainty information using tree structures showing examples from The Representation of Uncertainty using Scale-unspecific Tessellations (TRUST) software suite. Section 6 provides results from another survey to assess the usability and effectiveness of the proposed visualisation-of-uncertainty techniques and then concludes the paper.

2.0 UNCERTAINTY IN SPATIAL SYSTEMS

The real world is very intricate with large amounts of information and phenomena at differing spatial complexities. Spatial data can be collected from the real world and modelled in a computer environment in an abstracted form. It is almost impossible for the computer model to be an exact representation of the real world; therefore uncertainty will exist in the computer model. The user of digitally modelled spatial information can be completely unaware of any underlying uncertainty, and therefore make clear-cut decisions based on the information display resulting in undesirable consequences (Agumya & Hunter 2002). Defining and then expressing this uncertainty can be a way of providing the user with more complete representations of the spatial phenomena being modelled (Mahoney 1999).

Uncertainty as expressed by Zhang and Goodchild (2002) can be categorised under three main terms; error, randomness and vagueness. Error is the imprecision in values where there is an assumption that true values are obtainable; but it would be very difficult to collect and input the full truth into spatial systems in practice. Randomness occurs in spatial information and data collection technologies. Spatial data accumulates randomness through the dissimilarity and dependence inherent in the data. Technologies can help collect some spatial data but other spatial data can be harder to obtain at the same accuracy level, i.e. aerial photographs of urban environments can be categorised easily (buildings, roads, parks, etc), but aerial photographs of rural areas are harder to define (landforms, land-cover, etc). Vagueness exists in many spatial studies and applications; this is reflected through the complexity of the world and the subjectivity of using the spatial data to produce spatial information, i.e. this area is good for crop growing. *Good* crop growing is a subjective concept and is a vague application in spatial studies, conventionally modelled by fuzzy logic.

2.1 Uncertainty Models

Uncertainty models can be utilised to help determine the error associated with data, ideally by providing an uncertainty measure. These uncertainty measures can then be visualised in a coherent way applicable to the context and use of the spatial data. Different attribute data uncertainty models are used for different spatial data. Some models are ideal when using soil data; like fuzzy set theory (Zadeh 1965) to express vagueness in soil type boundaries (Goodchild 1994); or rough sets (Pawlak 1982) and their applications (e.g. deriving an ontology of imperfection in retail geography -Duckham et al., 2001). Other models like Monte Carlo can be used to express propagation of error (Fisher 1991; Goodchild, Guoqing & Shiren 1992; Davis & Keller 1997).

2.2 Uncertainty for Census Data

2.2.1 Spatial Uncertainty

Census data is commonly viewed as a choropleth display (Martin 1996). There are a number of spatial disadvantages of using choropleth mapping. Martin (1989) discussed spatial problems mapping population data in choropleth form. Martin explains that the uncertainty arises when aggregated data is assigned to areal units that are designed for ease of enumeration and therefore, are not designed to be data driven. This means that the data for an areal unit is basically a function of boundary definitions, rather than the underlying data distribution. This uncertainty in the choropleth map was coined the modifiable area unit problem (Openshaw & Taylor 1981). This means using imposed zone boundaries will apply a single geography to all variables rather than an appropriate geographic settlement pattern. Moreover, in Martin (1996, p.144) there is reference to the 'ecological fallacy' when using choropleth mapping. This term is used to represent the issues associated with the aggregation level for a dataset. A representation could be statistically correct but different patterns and trends will appear in the data depending on the aggregation level used (i.e. a shift from mesh block boundaries to district boundaries).

Since spatial uncertainty is implicitly added to census data and the irregular mesh block boundaries are designed to maintain privacy with some accuracy, there is an assumption that changing these boundaries will not add any more uncertainty. In fact Zhang and Goodchild (2002) discuss that adding precise boundaries to maps is not what users want to see, but rather definitions between occurring spatial patterns (i.e. transitions between rich and poor suburbia). Therefore, zone boundaries don't enhance classification; rather they introduce subjectivity. They continue to state that a good approach would be to split or group classified locations into regular areal units or tessellations. This would allow the exploitation of spatial adjacency, spatial autocorrelation and other contextual properties, creating more spatially coherent and meaningful zones. Also, Martin (1996) discussed a number of potential population representations when using aggregated socioeconomic data. A number of points were raised justifying the use of tessellations in socioeconomic data. The first point (p.169) concerns using existing census zones; '…there is little justification for the use of such boundaries as representations of the socioeconomic data.' Also, after a review of vector and raster population representations (p.176) when introducing modelled representations the following was proposed: 'The remaining alternative is to attempt to model the distribution of the underlying phenomena, regardless of the collection zones'.

2.2.2 Attribute Uncertainty

Monte Carlo was selected as the process to generate uncertainty values in attributes for this research. A Monte Carlo simulation would be a good approach to analysing error in census data because of its straight statistical methodology and the use of a Gaussian distribution is an ideal start when dealing with a large number of random factors contributing to error (Longley, Goodchild, Maguire et al. 2001; Zhang & Goodchild 2002). The use of Monte Carlo is well established in the GIS community (Burrough & McDonnell 1998; Heuvelink 1998; Zhang & Goodchild 2002).

Monte Carlo and the New Zealand 2001 Census

An uncertainty value for each feature in the New Zealand 2001 census dataset was derived. This was done with the help using data from the 2001 post enumeration survey (Statistics New Zealand. 2002), as well as running Monte Carlo simulations. The post survey was employed by Statistics New Zealand to test the quality of the census data, and provide documentation of the census accuracy to users. The survey showed an attribute error for census undercount and geographic location which was stochastically modelled through Monte Carlo, then propagated. The provided geographic attribute uncertainty data was specific to three regions covering New Zealand (South Island, Northern North Island and Southern North Island).

Procedure

Firstly a Monte Carlo simulation was run on the undercount data using the sample undercount errors from the post enumeration survey. After a user specified amount of simulations, mean and standard deviation layers were created for the undercount data. Next, uncertainty for the census data was obtained using a separate Monte Carlo simulation. The error value used in this simulation varied depending on the geographic location of the current feature in the simulation. There were three possible error values corresponding to the geographic areas specified above. The simulation used a normal distribution of random values with a mean of zero and a derived variance for each feature. After performing the iterations for each feature, mean and standard deviation layers were produced. The mean and standard deviation layers from the undercount and census simulations were combined. By dividing the standard deviation layer by the mean layer, an overall layer of relative error was produced and ready to be visualised.

3.0 REVIEW OF VISUALISATION-OF-UNCERTAINTY METHODS

A number of visualisation-of-uncertainty techniques have been developed over the past decade or so to visualise uncertainty measures. Part of this research instigated a survey to assess *effectiveness* of current uncertainty techniques and also to determine which techniques users prefer. The survey assessed the visual appeal, speed of comprehension, and the overall effectiveness of nine visualisation-of-uncertainty techniques. The study group consisted of 41 participants throughout New Zealand that had an understanding about uncertainty in GIS. These participants subscribed to a New Zealand GIS user group email list (Eagle Technology Group); their GIS experience was high, with 39% experts, 48% advanced, and 13% beginners (self-appointed). More than half of the respondents were in the GIS profession or used GIS in government. The techniques involved in the survey included:

Static Techniques

- *Adjacent maps*: Two value-by-area maps can be used to show the uncertainty, one to show the actual information and another to show the uncertainty (MacEachren 1992).
- *Overlay:* A single choropleth map can be used to show the attribute information with an overlay of the uncertain information shown as textures on top (MacEachren 1992; MacEachren 1998).
- *Blurring:* The clarity of an area boundary is used to define the uncertainty of the spatial data. A sharp pattern would indicate certain information, whilst a more approximate pattern definition would indicate uncertain information (MacEachren 1992).
- *Fog:* Uncertain parts on a map became partially hidden making the areas unclear to see. The thicker the fog the more uncertainty is in that part of the map The fact that fog obscures data from viewing is not an issue since such uncertain data is of no inherent use and therefore not deserving of being seen by the user (MacEachren 1992).
- *Pixel Mixture:* Commonly used for fuzzy data between categories. Fuzzy data can have multiple categories in a single location (i.e. different soil classes); therefore, by using fuzzy membership values many categories can be represented for a location. Pixels are divided into sub-pixels and an appropriate class value is given to each sub-pixel proportional to the membership function calculation (De Gruijter, Walvoort & van Gaans 1997).
- *Saturation of Colour:* Saturation of colour is used to visualise uncertainty. The more saturated (richer) a colour representing a particular class, the more certain the information is on that part of the map. Hengl et al. (2002) created a multi-dimensional uncertainty chart using fuzziness between classes and uncertainty in the data represented by gradual changes in hue and saturation respectively.
- *Sound:* This provided a level of uncertainty at a particular location on a map through a variable pitch. A low pitch sound depicted low uncertainty and a high pitch sound for large uncertainty – this was cursor driven (Fisher 1994; Krygier 1994).

Dynamic Techniques

- *Blinking Pixels:* Information in the spatial display was manipulated causing it to blink, hence highlighting those uncertain areas to the viewer (through different colours) (Fisher 1993; Monmonier & Gluck 1994; Evans 1997).
- *Animation:* A movie clip of map realisations (generated from a Monte Carlo simulation) can highlight areas where data is considered to be uncertain. Ehlschlaeger et al. (1997) state that if there is little change between realisations then one can be fairly convinced about the extent of the uncertainty.

3.1 Survey Results

The results of the first survey had two aspects; first if an uncertainty technique was deemed to be *usable or not* (Figure 1). Second, if a technique was deemed to be usable then the visual appeal, speed of comprehension and overall effectiveness of that technique was rated; in essence, 1 was ineffective through to 5 being excellently effective. Survey results showing overall *effectiveness* for usable techniques are shown in Figure 2. If a technique was deemed *not usable*, then a separate rating scheme was provided to assess why the technique was not usable (Figure 3).

The survey results from Figure 1 portray a message of limited usefulness of current visualisation-of-uncertainty techniques available. Only three visualisation-of-uncertainty techniques; blinking pixels, map adjacency and map overlay had a higher proportion of participants rating the technique more usable than not.

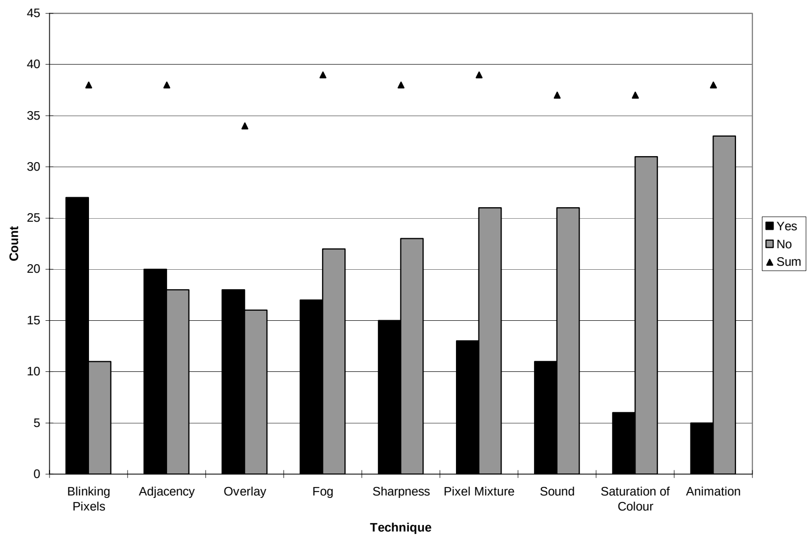


Figure 1: Survey results showing if an uncertainty technique is ***usable or not****; ordered by most usable (sum indicates participant response)*

Figure 2 shows that even though blinking pixels was the highest ranked, overall it rated weaker than adjacency on the 1-5 scale. Errors in the data collection for the overlay method have only produced a few results where the technique was deemed to be usable; therefore this research can only assume that the rankings will be somewhere in-between the adjacency and fog bar results from Figure 2. A high proportion of comments were raised about the need to see an unobstructed view of the original as well as the uncertainty information. This is consistent with the results from Figure 1; all the top three techniques allow an unobstructed view of the original data (apart from a textured overlay where some parts are unobstructed).

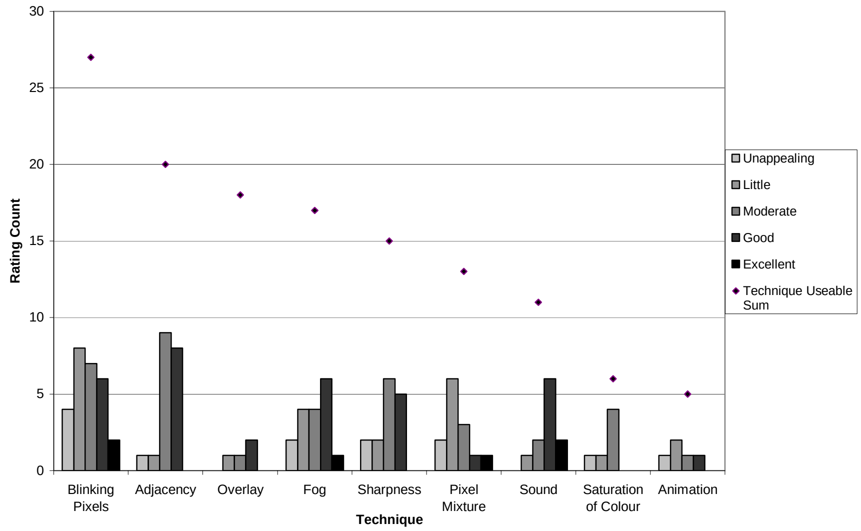


*Figure 2: Survey results showing **effectiveness** of visualisation techniques; ordered by most effective (sum indicates participant response)*

Figure 3 shows reasons why a particular technique was deemed to be not usable; and this provides some vital information defining pitfalls in current techniques. Some techniques had similar comments such as animation, blinking pixels and sound cannot be displayed on a printed map. Comments about the fog technique explained the disadvantage of obscuring the original data. Trends are noticeable in the graph; such as the inability for image sharpness and pixel mixture to convey uncertainty information and also how confusing an animation of uncertainty information can be. Sound ranked highly on the 'other' input dialog, due to many users not having the necessary hardware to perform an assessment.

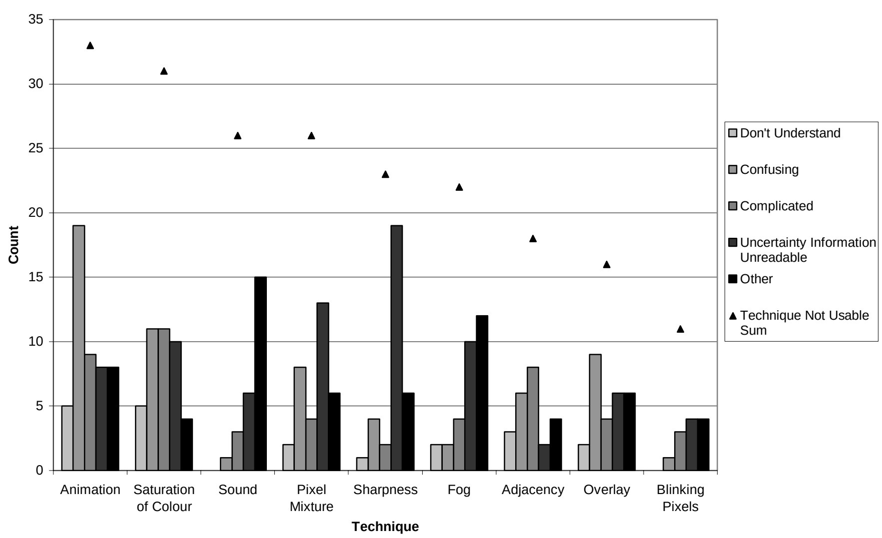


*Figure 3: Survey results showing why a technique is deemed **not usable**; ordered by least usable (sum indicates participant response)*

A lot of general comments were raised about the need for some form of visualisation-of-uncertainty information. This is consistent with GIS literature (MacEachren 1992; Mahoney 1999; Goodchild 2000; Agumya & Hunter

2002; Zhang & Goodchild 2002). This led to the question: what is a usable and effective technique to visualise attribute and spatial uncertainty information?

4.0 METAPHORS AND THE VISUALISATION-OF-UNCERTAINTY

Metaphors are commonly used in spatial information displays; this is because a metaphor has the power to explain reality better than reality itself. Peuquet (2002 p.123) defines the metaphor as "giving the thing a name that belongs to something else". Peuquet continues to state that metaphors are fundamental in human thought, and can map across different knowledge domains using a mental representation.

Zhang and Goodchild (2002) express that using metaphors to visualise uncertainty can provide a user-friendly interface when working with spatial data, and makes information more accessible to the end user community. Metaphors are commonly used when viewing spatial information. Fairbairn et al. (2001) write that the map is a metaphor to present and represent spatial and non-spatial data. They continue to discuss using map metaphor representations in terms of abstraction. Where data is realistic, a metaphor is chosen to represent a low degree of abstraction. On the other hand, when data is abstract a metaphor that represents a high degree of abstraction is chosen. Although a property of spatial and attribute data, quantified uncertainty can be regarded as an attribute in itself, representing an entity that has no physical form; it is created when data is collected, stored, handled, manipulated, and output. This would lead uncertain data towards the abstract end of the representation spectrum. Dent (1993) discusses using metaphors in abstract representation generally through geometric shapes such as circles, squares and triangles. Part of this paper will discuss using squares as an abstraction to represent spatial and attribute uncertainty. Furthermore, MacEachren (1992) demonstrates the use of visual metaphors to show uncertainty. Some metaphors used include: fog cover to hide the uncertain map parts, using resolution to express data at an applicable granularity level and the blurring of uncertain areas. Table 1 shows some uncertainty techniques and their metaphors in relation to data uncertainty.

Technique **Metaphor of:**	**Fog** – Detail	**Blur** – Focus	**Blinking Pixels** – Stability	**Colour Mix** – Clarity	**Pixel Mix** – Fuzziness	**Resolution** – Detail
Certain data **Metaphor of:**	*Clear* – Revealing detail	*Sharp Focus* – Focused	*No Blinking* – No movement, therefore more stable	*High Saturation* – High clarity, less recessive	*Single Hue* – Low fuzziness	*Fine Resolution* – Revealing more detail
Uncertain data **Metaphor of:**	*Foggy* – Hiding detail	*Blurry* – Unfocused and merging	*Blinking over areas* – Less stable, more unsettling	*Low Saturation* – Low clarity, more recessive	*Multiple Hues* – High fuzziness	*Coarse Resolution* – Hiding detail

Table 1. Spatial data metaphors used for visualising uncertainty in data

5.0 PROPOSED TREE STRUCTURE

For the remainder of this paper, when discussing the use of spatial data structures for visualisation, it is referring to the output from a spatial structure. Spatial data structures can be used as a way to visually represent spatial data as a metaphor. When representing spatial data, the output of the spatial data structure could be used as a metaphor for uncertainty showing levels of detail. In particular, the quadtree could use variable pixel size to show varying levels of detail; this uses detail as a metaphor. This is similar to the resolution metaphor, but instead of changing the detail the spatial data structure is a transparent layer on top of it. It is a multi-resolution display because more resolution is gained the more the tree structure divides. A display of uncertainty could be achieved using a tree structure to enable spatial visualisation using a tessellation of varying resolutions. This structure provides the ability to impose a coarse tessellation where the uncertainty is high and a finer tessellation where the uncertainty is low. MacEachren (1992) discusses using resolution as a way to visualise uncertain information. MacEachren states that the resolution of a map can be adjusted to represent graphic detail to correspond with the resolution of the attributes. Also, work by Chen and Tobler (1986) used quadtrees to index trend surfaces of digital terrain. A mathematical equation was used to approximate a terrain surface. If all of the pixels in a given node were less than a specified error range, then the mathematical function can be stored in the leaf node of a quadtree and used to represent the terrain surface. If a node is outside the error range, the node is divided into quadrates. This process is repeated until all nodes are within the error range. This produces a surface patch quadtree representation (Jones 1997) with quadrates of varying sizes according to the terrain surface. There are similarities between the patch quadtree technique Chen and Tobler developed and the proposed tree structure described. Both of the techniques use the quadtree to show a varying resolution based on an uncertainty or error derived from input data.

5.1 The Region Quadtree

The region quadtree is a hierarchical data structure. It is characterised by the recursive decomposition of space, by subdividing a bounded image array into four equal sized quadrates (Samet 1990). The region quadtree derives from a single root node and is divided into four ordered nodes representing four equal quadrate regions. These nodes can be also divided in turn if the quadrate region is non-homogeneous and continues to be divided until each node has homogeneity. Once each node has homogeneity then that quadrate is stored as a leaf node with its corresponding attributes (Worboys 1995). The region quadtree takes advantage of homogenous spatial regions in raster data and has a variable resolution.

5.1.1 Using the Region Quadtree to Visualise Uncertainty

Samet (1990) exemplifies the region quadtree as decomposing a bounded image array. In the context of this research the region quadtree will be used to decompose census mesh block polygons. The uncertainty of attributes associated with polygons can be conveyed through resolution of data; the less resolution the more uncertain the data is relating to that polygon. The resolution of the decomposition can be controlled by the uncertainty property of the input data.

The Representation of Uncertainty using Scale-unspecific Tessellations (TRUST)

A demonstration program, TRUST was created to view the uncertainty of spatial information using tree structures. TRUST v1.0 used the region quadtree as the tessellation to represent spatial uncertainty. The programmed steps in TRUST v1.0 to create a region quadtree showing accuracy information are:

- A Monte Carlo simulation is run to generate uncertainty values for the census data, using the undercount, geographic location and systematic error values for the selected area.
- The uncertainty data is scaled between 0 – (accurate) and 255 – (uncertain) for easy dissemination of uncertain features.
- An accuracy threshold limit is initialised using a number between 0 and 254, the smaller the number the more the structure will divide and when the limit is reached the division ceases.
- The accuracy threshold starts at 255 and decreases after each division using the following equation: Threashold = Threashold - (Threashold * 0.1) or 10% less than the previous value.
- Decomposition takes place using the quadtree structure only if the there is an accurate feature (i.e. one with an attribute error value lower than the threshold) at least 50% inside the quadrate, this function also starts after a user specified amount of tree structure divisions (else the tree structure might not start dividing and also if a small sliver of an accurate area is inside a quadrate but the remainder of the quadrate is filled with inaccurate areas, division should not take place; therefore, a percentile function is utilised).

Figure 4 shows a screenshot of TRUST v1.0 using the region quadtree to visualise attribute uncertainty whilst simultaneously displaying the attribute information. The quadtree can also convey spatial uncertainty by breaking up the rigid census mesh block boundaries into regular tessellations reducing subjectivity in boundary design and also promoting spatial relationships like adjacency and spatial autocorrelation (Zhang & Goodchild 2002). Also, where data is uncertain large tessellations will be present; therefore, large boundaries between the uncertain tessellations will exist. These large boundary definitions reduce in size in a stepwise approach when progressing towards accurate areas. This stepwise progression from high towards low uncertainty areas could be described as spatial autocorrelation in the data. It is more likely that socioeconomic data close together is going to be a similar accuracy, than data large distances apart (Martin 1996). The white figures (displayed for theoretical purposes) in Figure 4 represent the scaled uncertainty, which is proportional to the size of the quadrates. The greater the uncertainty of a feature, the larger the overlaid quadrate becomes.

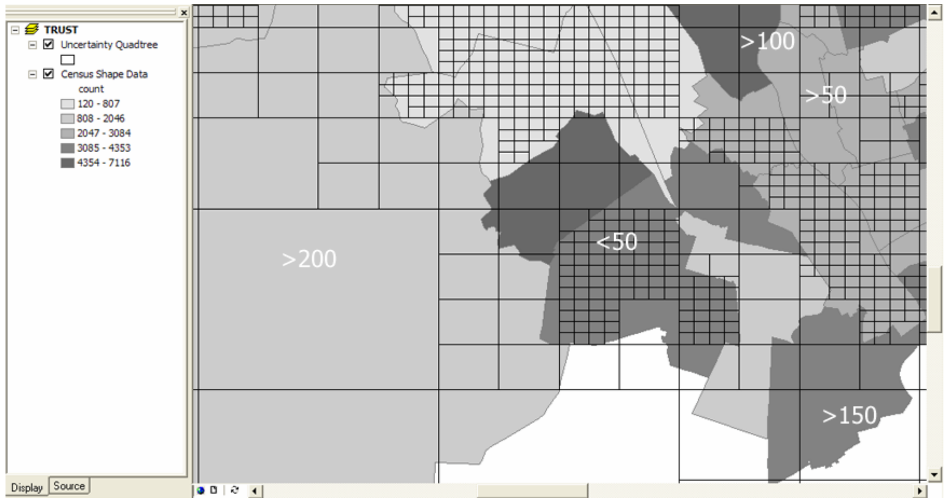


Figure 4: Screenshot of TRUST v1.0 showing attribute and spatial uncertainty in New Zealand 2001 Census data. The choropleth representation shows value by area census data, whilst the quadtree tessellation represents uncertainty in the data. Smaller tessellations represent accurate areas and visa versa.

5.2 Other Potential Data Structures

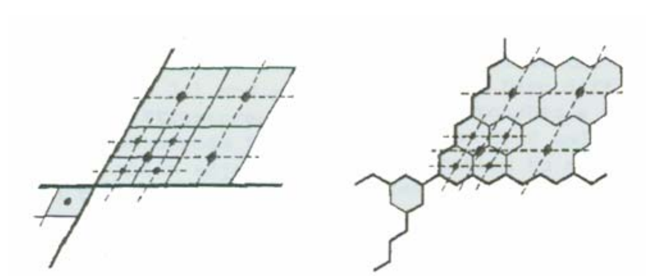
For geovisualisation, Carr et al. (1992) state that hexagons have two advantages over squares, visual appeal and representational accuracy. Also, the horizontal and vertical visual lines created by square tessellations can have a strong effect on humans due to our internal sense of gravitational balance. Therefore it is suggested that the horizontal and vertical lines generated by square tessellations in standard orientation are difficult to view and are less preferred. The hexagon tessellation is defined as being more compact and has the added ability to divide into a triangular tessellation at every level, which could lead to future research (White, Kimerling & Overton 1992; White 2000). Moreover, in White et al. (1999, p133) it is conversed that, ‘A hexagon tessellation minimizes spatial distortion and provides an equal area sample…’. Therefore, instead of using the square tessellation, the Lucas septree or the HoR quadtree could be possible tree structures to use in geovisualisation. Their advantages and pitfalls will now be discussed.

Lucas Septree

Bell et al. (1983) discuss a number of different tilings that can occur over a surface. One such hexagonal tiling is the septree or otherwise known as the 7-shape. It is a regular tiling where all sides are of equal length, as are the interior angles (Samet 1990). The septree is also a limited tiling, the tiling hierarchy becomes dissimilar; a partition cannot be infinitely decomposable into smaller patterns. Other limitations come to light when the conventional septree is divided, as each recursive child is rotated by an unjustified angle in relation to its parent (Bell, Diaz & Holroyd 1987; Samet 1990; Holroyd & Bell 1992). This can produce visual confusion with too many angles making the tessellation harder to create and index.

HoR (Hexagonal or Rhombus) Quadtree

The HoR structure (Figure 5) was designed by Bell et al. (1987) to remove the dividing limitation of a standard hexagon. Also the HoR quadtree does not have any rotational differences between child and parent shapes. After amalgamating the centroids of 4 hexagons or 4 rhombi, the HoR quadtree forms an identical centroid pattern. The rhombi pattern is unlimited, and is infinitely decomposable, but the hexagonal pattern is limited when at the lowest level, as shown in Figure 5b.



a. *b.*

Figure 5: The HoR structure in rhombic form (a), is an unlimited tiling whereas in hexagonal form (b), it is a limited tiling at the lowest level (Bell, Diaz & Holroyd 1987)

5.3 HoR Quadtree and TRUST

Both the septree and HoR quadtree have geovisualisation advantages over the square shape, but the HoR quadtree has the benefit of being an unlimited tiling with less visual confusion and therefore was chosen as another technique for the TRUST suite. Using the same steps to create a visualisation defined in section 5.1.1, the HoR quadtree tessellation was implemented into TRUST v1.1. The HoR tessellation in rhombic form has the down side of producing an optical illusion, known as Orbison's illusion (Robinson 1972). The bounding area around the tessellation seems to be distorted, the top of the map appears at an angle to the top edge of the image (see Kardos et al. 2003); therefore, it was not included in the TRUST suite. Figure 6 shows an example of the HoR quadtree using hexagonal tessellations to visualise spatial and attribute uncertainty whilst simultaneously displaying the attribute information. The white figures displayed for theoretical purposes represent the scaled uncertainty values, which is proportional to the size of the quadrates. The greater the uncertainty of a feature, the larger the quadrate size becomes.

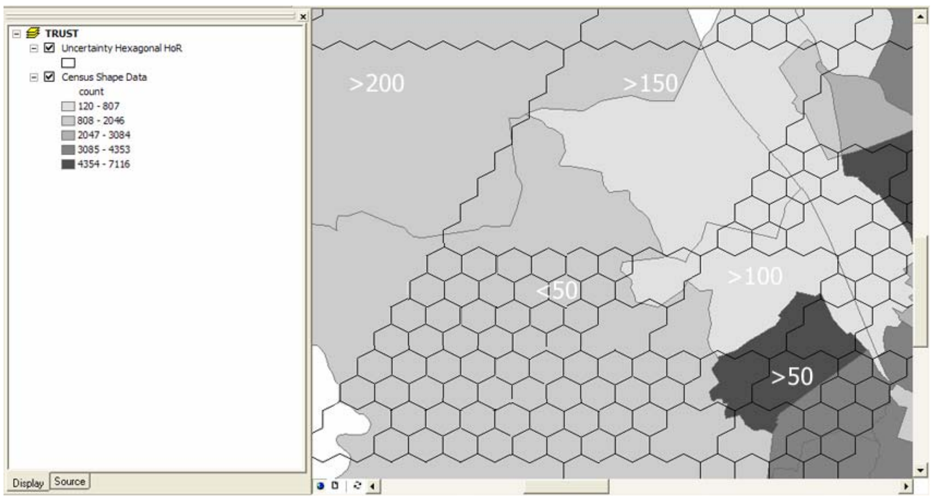


Figure 6: Screenshot of TRUST v1.1 showing attribute and spatial uncertainty in New Zealand 2001 Census data. The choropleth representation shows value by area census data, whilst the HoR tessellation in hexagonal form represents uncertainty in the data. Larger tessellations represent uncertain areas and visa versa.

6.0 RESULTS AND CONCLUSION

6.1 Survey Results

Another survey was designed to mimic the first survey but incorporate the new uncertainty visualisations discussed here, namely the quadtree representation and the HoR quadtree in hexagonal mode. These visualisations were created in the same format as the previous survey and used the same attribute uncertainty measures derived for each spatial feature. The study group consisted of 44 participants throughout the world that had an understanding about uncertainty in GIS. These participants all subscribed to numerous GIS user group email lists including Mapinfo, Geomedia, and Inmap (contact author for complete list); their GIS experience was high, with 46% experts, 32% advanced, and 19% intermediate (self-appointed). Close to half of the respondents were in the GIS profession or used GIS in government which was consistent with the previous survey. The results of the second survey were consistent with the results from the previous survey (shown in Section 3.1). Again, two main aspects were collected: if an uncertainty technique was deemed to be *usable or not* (Figure 7); and if it was then the visual appeal, speed of comprehension and overall effectiveness of that technique was rated using the same 1 (ineffective) to 5 (effective) scales. Survey results showing overall *effectiveness* rankings for usable techniques are shown in Figure 8. If the technique was not deemed usable then the separate rating scheme was provided to assess why the technique was *not usable*, shown in Figure 9.

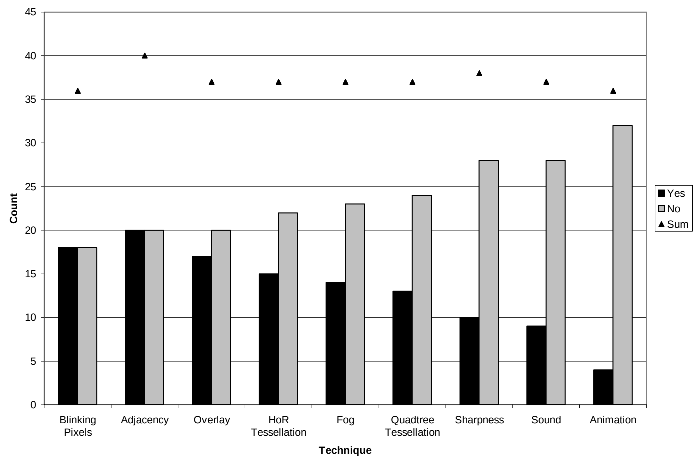


*Figure 7: Survey results showing if an uncertainty technique is **usable or not**; ordered by most usable (sum indicates participant response)*

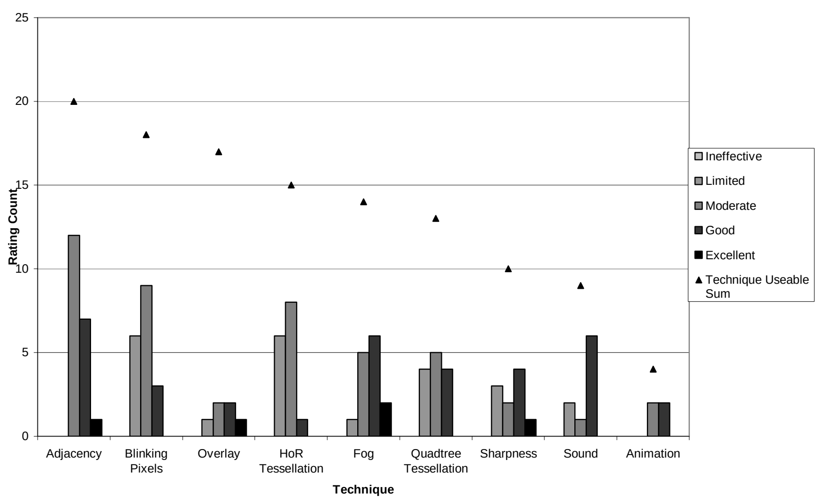


*Figure 8: Survey results showing **effectiveness** of usable uncertainty techniques; ordered by most effective (sum indicates participant response)*

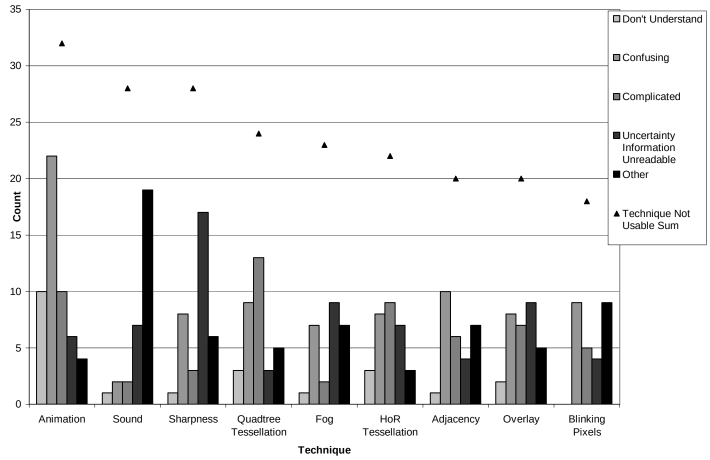


Figure 9: Survey results showing why uncertainty techniques are ***not usable****; ordered by least usable (sum indicates participant response)*

The results in terms of the proposed tree structures were a bit disappointing as neither technique rated high on the usability assessment, but then no technique was evaluated as being more *usable than not* as Figure 7 shows. After assessing the overall effectiveness results (Figure 8), it was clear that the blinking pixel technique was one of the most highly ranked techniques, but also the HoR quadtree produced a similar bar chart on the *effectiveness* scale. This is a good result considering that blinking pixels was easily the highest ranked visualisation technique in the first survey (Section 3.1). Again, errors in the data collection for the overlay method produced only a few results; therefore this research can only assume that the rankings will be somewhere in-between the blinking pixels and HoR tessellation bar graphs. Adjacency maps showing value-by-area uncertainty information ranked at the top, with 20 participants stating it is usable, and all ranked it 3 or above on the overall effectiveness scale.

After reviewing the reasons why a visualisation-of-uncertainty technique was *not usable* (Figure 9), some clear conclusions could be drawn. Firstly as previously suggested, the quadtree tessellation was found to be visually complicated, in fact the most visually complex. Other results were consistent with the first survey, animation of uncertainty information was found to be confusing, and a large pitfall with sound was the lack of a medium to obtain the information through. Also, many found the uncertainty information unreadable when the sharpness of an image became distorted. Other comments were directed towards different approaches to visualise uncertainty such as overlaying dots, symbols, numerical values and pie charts. Again a majority of comments expressed the benefits of an overlay, adjacent map or blinking pixel technique.

6.2 Conclusion

Current literature in GIS (Fairbairn, Andrienko, Andrienko et al. 2001; MacEachren & Kraak 2001; Zhang & Goodchild 2002), has expressed the need to visualise uncertainty, whether it be a supplement to existing data or a visualisation in its own right. Before visualising uncertainty data a measure of uncertainty is needed. A propagated uncertainty measure of census data can be derived using the Monte Carlo simulation statistical method. This can provide a good overall approach to analysing error in census data because of its straight statistical methodology. Also the use of a Gaussian distribution is ideal when dealing with a large number of random factors contributing to error, a property that census data can have.

The focus of this research has been to use hierarchical spatial data structures such as the region quadtree to visualise spatial and attribute uncertainty in socioeconomic data. The region quadtree provides a potential approach to disseminate aggregated census data and standardise spatial units, thus reducing subjectiveness in boundary definitions; moreover, attribute uncertainty can be expressed by exploiting the tessellation's ability to show variable resolution. Even the finest resolution cell in the uncertainty quadtree looks coarse when compared with the relatively precise choropleth boundaries that it overlays. These crisp-looking boundaries are misleading, as they imply that there is a stepwise progression of attribute value across the boundary line, when a smoother function is normally the case. Therefore, the coarse boundaries implicit in the quadtree which divide neighbouring polygon uncertainties are entirely appropriate. The use of hierarchical data structures to express both attribute and spatial uncertainty in this way is discussed in Kardos et al (in review).

When expressing attribute uncertainty; the more accurate an area, the more the data structure will divide, and therefore the smaller the resolution. In the converse case it could be argued that the tessellations could be made sufficiently dense so as to hide the attribute data, working in a similar way to the fog metaphor – this scenario will be explored. Other potential tessellation shapes were explored including the septree and HoR quadtree. As the literature and the survey results indicate, these do have visualisation advantages over the conventional region quadtree.

Finally, Fairbairn et al (2001) has also included multi-scale visualisation on the geovisualisation research agenda. With this in mind, there is an irony in using a multi-resolution structure such as a quadtree for something other than spatial location, effectively suppressing the multi-scale modelling capacity of the quadtree. Using a quadtree cell, not as a representation of geographical area, but as being proportional to uncertainty, is essentially the definition of a cartogram (see Kardos et al. in review).

ACKNOWLEDGEMENTS

This work would not be possible without funding from the School of Business at the University of Otago.

REFERENCES

Agumya, A. & G. J. Hunter (2002) Responding to the consequences of uncertainty in geographical data. *International journal of Geographical Information Science*, 16:5, pp. 405-417.

Bell, S. B. M., B. M. Diaz & F. C. Holroyd 1987, 'The HoR quadtree: An optimal structure based on a non-square 4-shape.', in *Mathematics in Remote Sensing,* Ed S. R. Brooks, Clarendon Press, Oxford, Chelmsford.

Bell, S. B. M., B. M. Diaz, F. C. Holroyd & M. Jackson (1983) Spatially referenced methods of processing raster and vector data. *Image and Vision Computing,* 1:4, pp. 211-220.

Burrough, P. A. & R. McDonnell (1998) *Principles of geographical information systems,* Oxford University Press, Oxford ; New York pp. xiii, 333.

Carr, D. B., A. R. Olsen & D. White (1992) Hexagon Mosaic Maps for Display of Univariate and Bivariate Geographical Data. *Cartography and Geographic Information Systems,* 19:4, pp. 228-236.

Chen, Z.-T. & W. R. Tobler (1986) Quadtree Representation of Digital Terrain. *Auto-Carto 6 Proceedings:*, London, pp. 475-484.

Davis, T. J. & C. P. Keller (1997) Modelling and Visualizing Multiple Spatial Uncertainties. *Computers and Geosciences,* 23:4, pp. 397-408.

De Gruijter, J. J., D. J. J. Walvoort & P. F. M. van Gaans (1997) Continuous soil maps -- a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. *Geoderma,* 77:2-4, pp. 169-195.

Dent, B. D. (1993) *Cartography : thematic map design,* Wm. C. Brown Publishers, 3rd, Dubuque, Iowa pp. xvii, 427.

Duckham, M., K. Mason, J. Stell & M. Worboys (2001) A formal approach to imperfection in geographic information. *Computers, Environment and Urban Systems,* 25: pp. 89-103.

Ehlschlaeger, C. R., A.M. Shortridge & M. F. Goodchild (1997) Visualizing spatial data uncertainty using animation. *Computers and Geosciences,* 23:4, pp. 387-395.

Evans, B. J. (1997) Dynamic display of spatial data reliability: does it benefit the Map user? *Computers and Geosciences,* 23:4, pp. 409-422.

Fairbairn, D., G. Andrienko, N. Andrienko, G. Buziek & J. Dykes (2001) Representation and its Relationship with Cartographic Visualization. *Cartography and Geographic Information Science,* 28:1, pp. 13-28.

Fisher, P. (1993) Visualizing uncertainty in soil maps by animation. *Cartographica,* 30:2-3, pp. 20-27.

Fisher, P. (1994) Hearing the Reliability in Classified Remotely Sensed Images. *Cartography and Geographical Information Systems,* 21:1, pp. 31-36.

Fisher, P. F. (1991) Modelling soil map-unit inclusions by Monte Carlo simulation. *International Journal of Geographical Information Systems,* 5: pp. 193–208.

Goodchild, M. F. (1994) GIS error models and visualization techniques for spatial variability in soils. *Proceedings 15th International Congress of Soil Science*, Acapulco, July 10-16, 1994, pp. 683-698.

Goodchild, M. F. (2000) Communicating geographic information in a digital age. *Annals of the Association of American Geographers,* 90:2, pp. 344-355.

Goodchild, M. F., S. Guoqing & Y. Shiren (1992) Development and test of an error model for catagorical data. *International Journal of Geographical Information Systems,* 6: pp. 87–104.

Heuvelink, G. B. M. (1998) *Error propagation in environmental modelling with GIS*, Taylor & Francis, London ; Bristol, PA pp. xviii, 127.

Holroyd, F. C. & S. B. M. Bell (1992) Raster GIS: Models of Raster Encoding. *Computers and Geosciences,* 1:4, pp. 419-426.

Jones, C. B. (1997) *Geographical information systems and computer cartography,* Longman, Harlow pp. xv, 319 , [318] of plates.

Kardos, J., G. Benwell & A. Moore (in review) The Visualisation of Uncertainty for Spatially Referenced Census Data using Hierarchical Tessellations. *Transactions in GIS.*

Kardos, J., A. Moore & G. L. Benwell (2003) Visualising Uncertainty in Spatially-Referenced Attribute Data using Hierarchical Spatial Data Structures. *GeoComputation 2003*, University of Southampton, England.

Krygier, J. B. 1994, 'Sound and Geographic Visualization', in *Visualization in the Modern Cartography*, Eds A. MacEachren & D. R. F. Taylor, Pergamon, New York, pp. 149-166.

Longley, P., M. F. Goodchild, D. J. Maguire & D. W. Rhind (2001) *Geographic information systems and science,* Wiley, Chichester ; New York pp. xviii, 454.

MacEachren, A. & M.-J. Kraak (2001) Research Challenges in Geovisualisation. *Cartography and Geographic Information Science,* 28:1, pp. 3-12.

MacEachren, A. M. (1992) Visualizing Uncertain Information. *Cartographic Perspective,* 13:Fall, pp. 10-19.

Mahoney, D. P. (1999) the picture of uncertainty. *Computer Graphics World,* Nov: pp. 44-50.

Martin, D. (1989) Mapping population data from zone centroid locations. *Trans Institute British Geographers,* 14:1, pp. 90-97.

Martin, D. (1996) *Geographic information systems: socioeconomic applications,* Routledge, Second Edition, London.

Monmonier, M. & M. Gluck (1994) Focus groups for design improvement in dynamic cartography. *Cartography and Geographical Information Systems,* 21:1, pp. 37-47.

Openshaw, S. & P. J. Taylor 1981, 'The modifiable areal unit problem', in *Quantitative Geography: a British view*, Eds N. Wrigley & R. J. Bennett, Routledge and Kegan Paul, London.

Pawlak, Z. (1982) Rough set theory and its applications to data analysis. *International Journal of Computer and Information Sciences,* 11: pp. 341-356.

Peuquet, D. J. (2002) *Representations of space and time*, Guilford Press, New York pp. xii, 380.

Robinson, J. (1972) *The psychology of visual illusion*, Hutchinson, London, pp. 288, [283].

Rosenfeld, A. & H. Samet (1985) Tree Structures for Region Representation. *Auto-Carto 7 : proceedings : digital representations of spatial knowledge : Washington, D.C., March 11-14, 1985*, Falls Church, Va., xvii, 599.

Samet, H. (1990) *The design and analysis of spatial data structures,* Addison-Wesley, Reading, Mass. pp. xvii, 493.

Statistics New Zealand. 2002, *A Report on the Post Enumeration Survey 2001*,, Statistics New Zealand, City, pp. 1-41.

White, D. (2000) Global grids from recursive diamond subdivisions of the surface of an octahedron or icosahedron. *Environmental Monitoring and Assessment,* 64:1, pp. 93-103.

White, D., A. J. Kimerling & W. S. Overton (1992) Cartographic and geometric components of a global sampling design for environmental monitoring. *Cartography and Geographic Information Systems,* 19:1, pp. 5-22.

White, D., E. M. Preston, K. E. Freemark & A. R. Kiester 1999, 'A hierarchical framework for conserving biodiversity.', in *Landscape ecological analysis: issues and applications.*, Eds Klopatek JM & G. RH, Springer-Verlag, New York, pp. 127-153.

Worboys, M. (1995) *GIS : a computing perspective,* Taylor & Francis, London ; Bristol, Pa. pp. xiv, 376.

Zadeh, L. (1965) Fuzzy sets. *Information and Control,* 8: pp. 338-353.

Zhang, J. & M. F. Goodchild (2002) *Uncertainty in geographical information,* Taylor & Francis, London ; New York pp. x, 266.