Knowledge-Based Generalisation of Road Networks

Kazemi, S.,[1*] and Forghani, A.,[2]
[1]Faculty of Education, Science, Technology and Mathematics, University of Canberra (UC), Bruce ACT 2601 Australia, E-mail: sharon_kazemi@hotmail.com, sharon.kazemi@canberra.edu.au
[2]School of Natural and Built Environment, University of South Australia, Adelaide, South Australia, Australia, E-mail: alan.forghani@unisa.edu.au; alanforghani@yahoo.com.au
***Corresponding Author*

Abstract
This paper focuses on applying a systematic approach to the generalisation of a road network database using a dynamic generalisation approach through an evaluation of DynaGen™ *generalisation capabilities. It is used as a testbed based on the principles of generalisation. It then compares the results derived from* DynaGen™ *with an earlier work with ArcGIS™ road generalisation results. It also includes details of the functionality of each process which, combined with human interaction, involves resetting the rules and numerical parameters to achieve the desired results.*

1. Introduction

The problem of deriving a range of map scales from a common database in today's map production environment is considered a challenging area of research and development (e.g. Arnold and Wright, 2005 and Hardy *et al.,* 2004). Technological developments in dynamic cartographic generalisation applications (e.g. mobile mapping, vehicle navigation) call for a robust, practical and cost effective way of fulfilling automated generalisation needs. The trends toward automated generalisation require a powerful set of software tools to replace traditional methods for cartographic map production. A review of the literature on automated generalisation has found a need for intensive experiments with the existing tools developed in the industry rather than the development of new generalisation systems (Lecordix *et al.,* 1997 and Kazemi and Lim, 2007a). Therefore this research has focused on the application of the generalisation tools to the development of a detailed generalisation framework for deriving multiscale road data. Much work has been carried out in the last decade on the development of various generalisation algorithms, but the need to evaluate and validate existing generalisation tools has been overlooked. It was revealed that an assessment of these existing systems will facilitate the development of a detailed generalisation framework for deriving multi scale data and map products from a single high resolution database. For example, Karsznia (2008) tested road network and settlement generalization using DynaGen from the point of view of its formalization as well as further development of a knowledge base concerning small-scale spatial data generalization in commercial software such as DynaGen by Intergraph. More particularly, the research by Kazemi *et al.,* (2005b) and Kazemi and Lim (2007a) discussed utilising the generalisation operators provided by ESRI ArcGIS™ and Intergraph DynaGen™, combined with the skills of a trained cartographer, to generalise a road network database from 1:250,000 national topographic data and produce smaller scale maps at 1:500,000 and 1:1,000,000. The method was tested (Kazemi *et al.,* 2009) and the results show that the derived maps are satisfactory when compared with the existing small scale road maps such as the Global Map at scale of 1:1,000,000. The *Global Map* Australia 1:1,000,000 (2000) is a digital spatial dataset covering the Australian continent and island territories which consists of eight spatial layers in the vector form (administrative boundaries, drainage, transportation and elevation as well as several raster forms (elevation, vegetation, land cover and land use). This was created from Geoscience Australia's (GA) 1:250,000 national topographic data. The generalisation operators of ArcGIS™ have been tested (Kazemi *et al.,* 2005a) for the generalisation of roads by employing the conceptual generalisation framework for derivative mapping, but the results showed that ArcGIS™ algorithms of Douglas and Peucker (1973) and *Bendsimplify* do not support a dynamic generalisation. The way to work around this problem was to use Intergraph™ generalisation software which enables the production of a multiscale database from a master database. As noted in earlier (Section 1), a multipurpose spatial corporate database offers capabilities to derive different maps at different scales from objects (e.g. topographic objects), at scales ranging from, say, 1:250,000 to 1:1,000,000. This capability is referred to as a 'derivative mapping' capability. This paper focuses on the principles of generalisation that were researched using Intergraph DynaGen™ as a testbed. It then compares the results derived from DynaGen™ with earlier work using ArcGIS™ road generalisation. Intergraph's DynaGen™ software is one of the most detailed generalisation software products. The components of this technology together with detailed functionalities were described in Section 2. According to existing literature and Intergraph, this generalisation system has been tested for various generalisation tasks in several countries but not in Australia. Iwaniak and Paluszynski (2001) investigated the application of cartographic skills, together with Intergraph MGE Map Generaliser software tools, to automate the generalisation of topographic maps of urban areas from 1:10,000 to 1:50,000 scale. The system uses the MGE Map Generalise™ and a rule-based system implemented in C Language Integrated Production System (CLIPS), developed by Purdue University for controlling the generalisation process via a knowledge-based expert system that generates results similar to those obtained with the manual procedure. One of the key features of this system is its efficient handling of conflict resolution among objects. The expert system enables the integration of generalisation operations, generalisation rules and manual intervention. The authors suggested that this approach warranted further research. Chybicka et al., (2004) applied the same approach to the generalisation of the Polish topographic database at a scale of 1:10,000 to derive 1:50,000 databases. Several researchers (e.g. Meng, 1997, McMaster and Shea 1989 and Mackaness *et al.,* 2007) have highlighted the questions of 'how', 'when' and 'why' to generalise. Building an automated generalisation presents an immense challenge when handling national/global spatial coverage at widely varying levels of detail, as digital and paper products while maintaining currency of spatial information products; it therefore requires practical solutions to the aforementioned questions. Mackaness *et al.,* (2007) noted that recent developments in cartometric analysis techniques are able to support high levels of automation among multiscale derivation techniques within existing and emerging technologies such as mobile mapping. For example, which generalisation algorithm(s) is (are) the most suitable for point generalisation for real time execution, and how it (they) can affect map reading tasks (in terms of efficiency and accuracy) with minimum operator intervention (Bereuter and Weibel, 2010). In relation to an integrated approach, Yang and Gold (2004) proposed a system which combines database generalisation and object generalisation to overcome generalisation problems (e.g. feature displacement and smoothing) which have been widely reported in the literature (e.g. Ruas, 1998 and Stoter *et al.,* 2004). They observed that, since current generalisation practices involve human interaction rather than dynamic generalisation, there is a need to develop a generalisation framework for the generalisation of spatial databases. They documented the development of the generalisation of topographic data from a purely manual process to interactive generalisation using LAMPS2™ software and ArcGIS™ for the generalisation of buildings. It was suggested that future research should concentrate on the development of a robust core data model and evaluation of generalisation systems. Since an effective data model stores special relationships among features, by designing a good data model relationships such as adjacency and connectivity can be established so that generalisation operations (e.g. aggregation) will be more effectively defined through topological relationships between features. In this regard ESRI geo-database technology provides a framework for objects to maintain geometric attributes, spatial references, relationships, domain and validation rules, topology and custom behaviours (Zeiler, 1999). For instance on a very large scale map roads appear with detailed multiple lanes, and at medium range scale lanes are formed as two-way traffic directions and at small scales appear as a single road. This study will therefore contribute to addressing this issue by combining cartographers' knowledge with different generalisation algorithms through an evaluation of a variety of software systems in order to achieve cartographically acceptable results. Kazemi et al., (2005b) noted that, for the automation of the map generalisation process it is necessary to integrate generalisation algorithms with cartographer's intuition and skills within a GIS, as this approach leads to more desirable outputs (Joao, 1998). To date there have been no systematic attempts to undertake a comprehensive evaluation of generalisation systems and their performance. This paper discusses the DynaGen™ generalisation capabilities using road datasets. In the next section a brief overview of the study area, the Intergraph™ software, and the research dataset is provided. Section 2 describes the developed generalisation methodology using the software cartographic generalisation capability in a 'dynamic mode'.

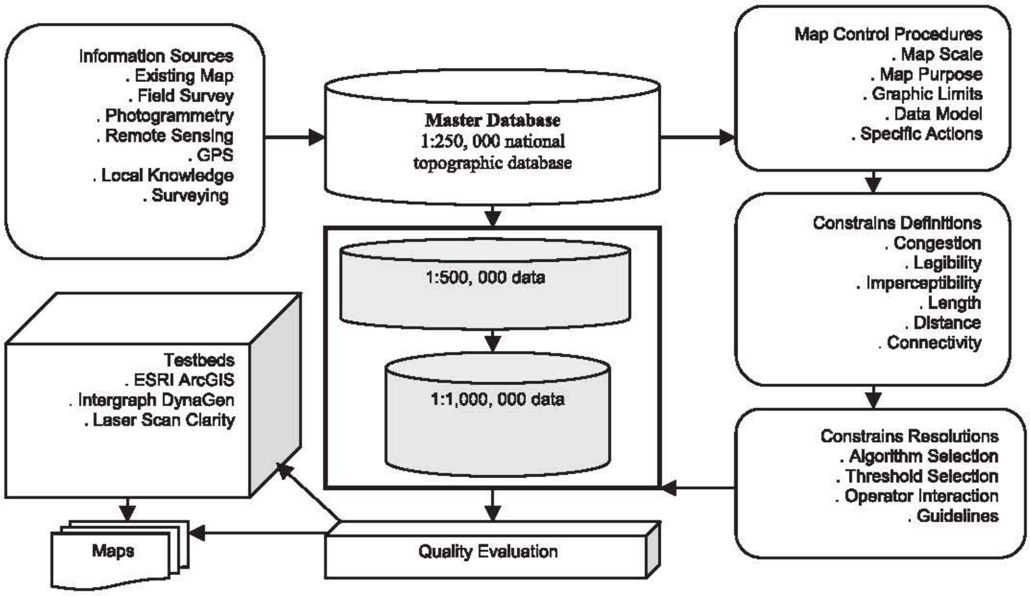

Figure 1a: Road network generalisation workflow using DynaGen™ (Adapted from Kazemi and Lim, 2005b)

In particular, this section elaborates on the application of different generalisation algorithms used by software and their performance. Then, generalisation performance for the target small scales (1:500,000 and 1:1,000,000) is briefly discussed. Finally, conclusions and recommendations for future research are given. Relevant examples of research studies for spatial data generalisation include: Bart et al., (2013), Bereuter et al., (2013), Lüscher and Weibel, (2013), Benz and Weibel, (2014), Venkateswaran *et al.,* (2014), Weiss and Weibel (2015), and Technitis *et al.,* (2015).

2. Methodology

A schematic representation of the research methodology is presented in Figure 1. The motive for this research is to develop a workflow specifically for derivation of multiple scale maps from a master road network database over a study area in Australia. A large portion of this approach is similar to the previous study by the author in which ArcGIS™ generalisation capabilities were tested. However, all the processes in the flowchart are considered specifically for road generalisation using the DynaGen™ system. The difference between these two, approaches is in the underlying principles of the two generalisation systems and the authors' earlier work with ArcGIS™ (Kazemi *et al.,* 2005b). The study area covers approximately 23,630 km² of Australia's capital city, Canberra Australia (Figure 1b). The area coverage of the chosen sites are in the longitude and latitude ranges 148° 42′ 7″ to 149° 25′ 56″ and 35° 55′ 35″ to 35° 7′ 22″ respectively, with an elevation of 550-700m above sea level.

The study site was chosen because it provided a mix of different roads with a reasonable amount of feature density. The selection of the area was based on testing the generalisation concepts over an urban environment, since the density of roads was a determinant factor, and for ease of access to the datasets. The National Mapping and Information Group of Geoscience Australia (GA) produce 1:250,000 seamless topographic data and associated cartographic databases across Australia that is used in this research. Roads in the 1:250,000 national topographic data have varying widths and types (GA, 1999). In the classification, six road classes can be distinguished in the study:

- Dual Carriageway (DC) – divided highway, freeway (e.g. Federal Highway), tollway, and other major roads with separated carriageways in opposite directions.
- Principal road (PR) – highways, major through routes and major connecting roads as described by the Australian Automobile Association (AAA) (both sealed and unsealed).
- Secondary Road (SR) – connecting roads (both sealed and unsealed) that provide a connection between major through routes and major connecting roads or between regional centres. All principal and secondary roads are shown including those in built-up areas.
- Minor Road (MR) – all other roads (both sealed and unsealed) that form part of the public roads

system between principal roads and secondary roads.

- Vehicle Track (VT) – public or private roadways of minimum or no construction which are not necessarily maintained.
- Foot Tracks (FT) – these tracks are for pedestrian traffic only. Tracks with a length of less than1.25 kilometres have not been captured.

Major roads (e.g. Northbourne Avenue) are located in the north and south of the study area. Because roads in the study area have different widths, the road vector coverage was overlaid on the fused ETM+ imagery (15m colour), and every road segment was labelled with a specific distance for its width based on visual assessment.

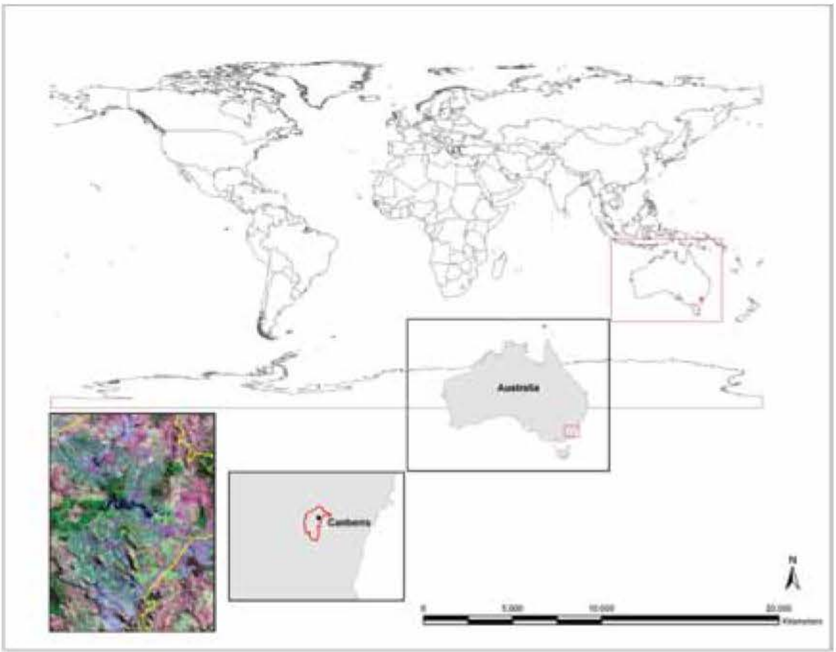

Figure 1b: Location of study area, Canberra, Australian Capital Territory (ACT) Australia; Study site is shown in Landsat ETM+ image (composite bands 6, 4, 2), courtesy of Geoscience Australia © 2001. The Landsat image is not to the same scale

Figure 1c: Examples of roads within the study area (a) Corner of Hindmarsh Drive and Jerrabomberra Avenue, Symonston (Dual carriageway), (b) Commonwealth Avenue, Barton (Principal road), (c) Intersection of Melrose Drive and Yarra Glen Drive, Woden (Secondary road) and (d) Intersection of Leichhardt Street and Dawes Street, Kingston (Minor road).

The roads were classified into eight categories with buffer distances ranging from 1 to 6 metres. Principal roads (e.g. urban main roads) can be seen extending in two directions, starting from the north and extending to the southeast and to the southwest. The width of these is between 5 to 10 metres. Minor roads are concentrated mainly in the central part of the study area. The width of these streets is 4 to 8 metres. Minor roads are located mostly in the recreational areas. These roads include rural roads, access roads, and unpaved tracks. The width of these roads varies from 2.5 to 5 metres. Kazemi et al., (2009b) used Landsat ETM+ for an effort in developing a framework for segmentation and generalisation of raster data known as *'Interactive Automated Segmentation and Raster Generalisation Framework'* (IASRGF). She applied three generalisation techniques (supervised classification, thematic generalisation and spatial aggregation) in order to build the IASRGF for segmentation and generalisation of satellite imagery. Given that ETM data was one of the fundamental datasets used for Kazemi's doctoral dissertation research. It should be noted that the use of other high resolution satellite imagery was not the focus of that study in order to apply an improved spatial resolution of satellite data such those imagery used by Google Earth, Microsoft Bing, Nokia, etc). Figure 1c illustrates the different roads in the study area. The process of developing GIS data for this study is the most important step in developing a guideline for derivative mapping using software applications to generalise 1:250,000 national topographic roads to produce 1:500,000 and 1:1,000,000 roads spatial data and maps. This process of road network generalisation is detailed in Section 3. The process of developing GIS data is an important step in developing a guideline for derivative mapping using software applications to generalise 1:250,000 national topographic roads to produce 1:500,000 and 1:1,000,000 roads spatial data and maps.This process has been discussed in detail by Kazemi *et al.,* (2013). Their paper presented a generalisation methodology to derive multiscale spatial data through an evaluation of ESRI ArcToolbox Generalise™ software. That study focused on integration and utilisation of generalisation operators in order to generalise a road network database and produce small scale maps at 1:500,000 and 1:5,000,000 from 1:250,000 national topographic data.

2.1 System Architecture

The Intergraph GIS system consists of three major components, namely GeoMedia™, GeoMedia WebMap™, and DynaGen™. GeoMedia™ enables appropriate spatial data integration including the capability for incorporating data from disparate sources and multiple formats into a master GIS for viewing, analysis (e.g. query, buffering, overlaying) and presentation. Intergraph's web-based map visualisation and analysis system offers real time access to a data warehouse to publish data on the Web. DynaGen™ is a subsystem of Intergraph Dynamo™ which allows the use of a graphical environment, and topological functions, and the data models of Dynamo™. DynaGen™ software is a cartographic suite of tools combined with Data Editor (DIDE) for creating, or connecting to, a database for further processing. Generalisation processes in DynaGen™ can be very detailed; users can take control of the data and monitor the whole process. Intergraph's DynaGen™ automates the production of multiscale maps from a master database. It operates using two modes of generalisation: the batch mode (automatic) and the cartographer-assisted mode. Methods discussed in this paper are more related to vertex reduction using line simplification and smoothing. Various algorithms, such as the Douglas and Peucker (1973) simplification (Section 3.3) and Brophy smoothing algorithm (Section 3.2), have been tested with a variety of input parameters. Brophy smoothing is mathematically described in Brophy (1973).

Automated (Batch) Operations: generalisation operators are either based on mathematical formulae or by unequivocal procedure descriptions (algorithm). This transformation is named the generalisation step. DynaGen™ offers a number of generalisation operators, such as simplify, smooth, and feature elimination that have been applied to road generalisation in this paper. DynaGen™ automated generalisation uses a sequence of such transformations through selection and incorporation of parameters in such a way as to maintain certain characteristics and to establish relationships between generalised objects. The cartographer also has access to a number of operators including simplify, smooth, aggregate, change the presentation of objects, extend borders, select representative objects, angle straightening and amalgamation of objects.

Dynamic (Interactive) Mode: DynaGen™ enables 'dynamic' generalisation since the operator is able to change any parameter value using sliders, and visually inspect the dynamically changing generalisation results in real time. This allows the cartographer to fine-tune the process when selecting individual features or a group of features during the generalisation process. The main engine of DynaGen™ is Dynamo, which allows the user to control the changes in the database. This useful capability enables the cartographer to visualise the results even before running the process, which could lead to an overall cost saving compared to the automated mode since the cartographer is able to visually assess, validate, and accept the generalised output 'on the fly' (Iwaniak and Paluszynski, 2001).

2.2 Data Pre-Processing

Because data could come from disparate sources and often the user has limited knowledge of the source, its capture method, and accuracy of GIS data in terms of completeness and currency, it is essential to perform some validation and testing of data prior to starting generalisation. Topology validation, geometric feature validation and feature attribute quality checks were performed on the 1:250,000 national topographic roads data used in this research. This resulted in an improvement of quality of the research data. The improvements include fixing attribute and topological (under-shoot and over-shoot) errors. Overshoots occurs when a line (arc) does not end at its termination point on another line (arc) and goes beyond it is called as overshoot. Undershoots occur when a line (arc) finishes before connecting / intersecting to another line (arc) on desired location.

3. Building Knowledge Base

DynaGen™ validates the generalisation output against the operator's generalisation knowledge, ensuring that changes do not violate topological relationships, and it calculates and updates feature attribute information for changed or recently generalised features (Watson and Smith, 2004). It is used to construct a knowledge base that conforms to two principles; applying generalisation rules executed in the automated mode, and using fundamental generalisation processes executed and monitored by cartographers. The DynaGen™ knowledge-base uses specific generalisation steps, namely: 1) the name of the generalised object, 2) the operator, 3) the algorithm, 4) the values of the algorithm parameters, 5) the names and values of attributes referring to objects created as a result of the generalisation, 6) a condition implementing a particular approach, and (7) prescribing prohibited topological changes in terms of spatial relations between generalised objects. The DynaGen™ system incorporates two sets of principles: a) those containing rules executed in an automatic mode and in a restricted sequence, serving as preliminary data preparation, and b) those describing fundamental generalisation processes executed interactively and managed by the cartographer.

3.1 Feature Elimination

This function enables the user to define feature class and attribute conditions of features that are not needed and to remove these from the object space. This is especially useful for enforcing map specifications by removing features that do not meet product requirements. For an efficient feature elimination function, the database needs to be made more 'intelligent' by incorporating object-oriented technology. DynaGen™ therefore allows the user to define spatial feature relationships that may impact on the process of feature elimination. An example of the application of this operation is to inspect the Initialisation Data Editor (IDE) tool, allowing the user to build parameter sets containing information on generalisation operators in order to generate parameter files when initiating batch processing and interactive generalisation. For example, road segments which are shorter than a certain length, which cause conflict in the final map and which are not significant for presentation on the map, may be eliminated. For all these features, it would be ideal to show a screen-shot of the pull down menu listing options/parameters that is (are) applied to each of four subsets. However, the DynaGen™ software is no longer available for this study due to an expired grant licence at the time of rewrite of this paper (September, 2015).

3.2 Line Smoothing

Smoothing and simplification operations result in a reduction of plotting time, a reduction of storage space, faster vector-to-raster conversion, and faster vector processing. However, there is a difference between simplification and smoothing. In other words, line simplification deals with the representation of the curvature of the line by fewer points, while the smoothing operation deals with representation of the line with fewer sharp angles to improve aesthetics. The line averaging (smoothing) algorithms shift the position of points making up a line, in order to remove small perturbations and capture only the most significant trends of the line. Hence, the smoothing is applied to improve the appearance of digitised lines or, more simply, for cosmetic modification of lines as described by McMaster and Shea, (1992) and Hunter (1998). Three well known smoothing algorithms (Brophy, Simple Average and Weighted Average) were applied in this study (Figure 2). These are embedded in DynaGen™. Each algorithm uses a variety of parameters. The Brophy algorithm uses *look ahead factor*, *densification* and *smooth factor*. The Simple Average algorithm uses *look ahead factor* and *densification*. The Weighted Average algorithm shares the principles of two earlier algorithm's parameters, and brings *Weight* into the equation (Intergraph, 2004). Applied parameters include:

a) *Densification,* which temporarily adds vertices along the geometry in order to move the real vertices further, the temporary vertices being deleted once the smoothing process is completed,
b) *Look ahead* which determines how many points ahead of and behind the point that is to be smoothed are used in the smoothing of that point,
c) *Smooth* which determines the distance a point is to be shifted over the radius of the Brophy circle, and
d) *Weight* which determines the weight given to the points during the averaging process.

The smoothing process has been run on a number of road samples from 1:250,000 national topographic dataset (Figure 2). The effects of the selected smoothing algorithm on that particular segment of the road are highlighted within the red circles.

The road shape is changed during the vertex reduction. The road shape is important when it comes to generalisation due to its effect on coincidence (e.g. roads passing over the top of buildings), features, and accuracy. Producing a readable map and, at the same time, preserving accuracy is the core task of the science of cartography. It is a trade-off that must be balanced in order to achieve a desirable number of vertices while not causing deterioration of the quality of the remaining linear features. However, to overcome this problem generalisation with an appropriate level of tolerance, suitable settings and appropriate tools is required. This was achieved by several attempts to find an appropriate tolerance setting. It has to be noted that reduction of the vertices will result in low data volume. Smoothing operators do not change the topology of the feature; all existing topological relationships are maintained without creating new topological relationships. This ensures beginning and ending points of lines are not moved or removed.

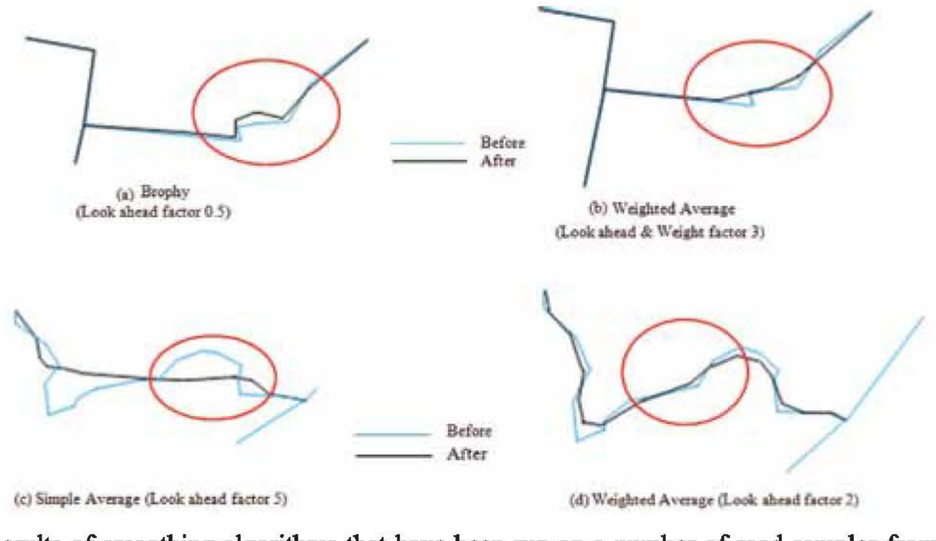

Figure 2: Results of smoothing algorithms that have been run on a number of road samples from 1:250,000 national topographic road databases over the study area. Data courtesy of Geoscience Australia © 2004

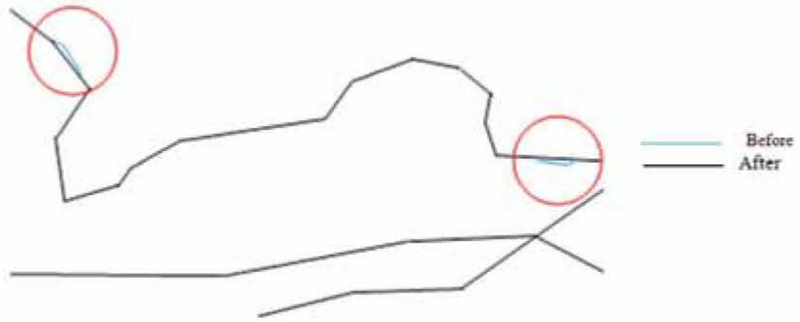

Figure 3: The results of Douglas-Peucker simplification algorithm using tolerance of 3m. Data courtesy of Geoscience Australia © 2004

3.3 Line Simplification

DynaGen^{TM} offers six simplification algorithms which can be used to simplify roads data. The results of the Douglas and Peucker (1973) simplification algorithm are presented below (Figure 3), and the line simplification algorithms and their parameters are tabulated (Table 1). The minor bends along the road segments are deleted to improve the cartographic representation of the roads.Line simplification demonstrated that the Douglas and Peucker (1973) algorithm produces the most desirable generalised results of all the applied simplified algorithms (Alves, 2010). It operates on entire lines. This is because topological and geometrical relations of adjacent objects are appropriately handled when the generalised elements are close to each other, and where there is a possibility of either overlap or intersect. Also, the Point Relaxation simplification algorithm performed better in terms of aesthetic quality of simplified lines than the other remaining algorithms of Lang, Nth Point, Reumann-Witkam, and VectGen. For example, Reumann-Witkam simplification algorithm involves three main characteristics: applying two parallel lines to identify search region; calculating initial slope of the search region; and line processing sequentially until one of the edges of the search corridor intersects a line. These algorithms have been widely discussed in the literature and are not described here (e.g. Visvalingam and Williamson, 1995 and Muller, 1995).

3.4 Line Typification

This operation is considered a cartographic generalisation task that makes maps more readable by reducing density and details of lines, points and polygons (Kazemi et al., 2004a) while retaining the character of the original features. In the context of linear feature generalisation it gives a representative pattern of the more significant features in the road network while removing line features based on their proximity to each other. Two popular line typification algorithms, namely tree levelling and conflict resolution with a number of parameters, were tested in this study. The tree-levelling algorithm uses three parameters: minimum terminal branch length, maximum branch retention level, and identify critical lines. This handled conflict resolution reasonably well as demonstrated in Figure 2. Similarly, conflict resolution utilises three parameters: *minimum spacing tolerance, proximity limit,* and *identifies critical lines*. Description of these typification parameters follows. *Minimum terminal branch length* specifies the minimum terminal line length. For example, roads shorter than the minimum line length and having no branches are removed. *Maximum branch retention level* retain the maximum number of levels in the tree. If the value is set to 0, all input roads are removed. (c) *Identify critical lines* enables selection of a set of lines as critical lines that are always a part of the representative pattern of lines to retain on the map. *Minimum spacing tolerance* determines the minimum line spacing tolerance. In this regard roads shorter than the minimum spacing tolerance are selected for removal. The algorithm basically reduces the density and simplifies the distribution and pattern of the network. The result should be a connected and less congested network that represents a similar pattern at the smaller scale.

4. Assessing Generalisation Performance

The line simplification algorithms change the geometry of a line by eliminating a number of vertices applying tolerance values. This is a factor used to determine the influence of the algorithm on cartographic line simplification processes. Several quantitatively and qualitatively measures have been developed to assess the shape of the simplified lines.

Table 1: Line simplification algorithms and their parameters (Intergraph, 2004)

Algorithm	Parameters	Comments
Douglas-Peucker	Z-retention flag and tolerance	Z-retention flag of 1-10m and tolerances of 0.1 to 5 is used.
Lang	Z-retention flag and tolerance	Z-retention flag of 1-10m and tolerances of 0.1 to 5 is used.
Nth Point	Z-retention flag and N	Z-retention flag of 1-10m and point spacing of 1 to 20 is used.
Point Relaxation	Z-retention flag and relaxation radius	Z-retention flag of 1-10m and relax circle radius 0.2-5m is used. But the optimum value is found 3.93m.
Reumann-Witkam (Reumann and Witkam, 1974)	Z-retention flag and tolerance	Z-retention flag of 1-10m and Corridor Tolerance of 0.2-5m is used.
VectGen	Z-retention flag and tolerance	Z-retention flag of 1-10m and tolerances of 0.2-5m is used.

Researchers (e.g. McMaster, 1986, Buttenfield, 2002 and Skopeliti and Tsoulos, 2001) have used a number of mathematical measures (i.e. length, number of line reductions/density, angularity, curvilinearity and density) for evaluating line simplification. McMaster's (1986) cartometric measure of total length and number of roads (density) based on feature types is considered here as an index of generalisation to quantify generalisation performance for the target small scale (see Figure 4). Changes in the representation of road features at 1:250,000, 1:500,000 and 1:1,000,000 scales as a result of generalisation were quantified. Roads outputs from map derivation have been analysed applying the *Radical Law* (Pfer and Pillewizer, 1966) both employing ArcGIS™ and DynaGen™ systems. The *Radical Law* (Equation 1) determines the retained number of objects for a given scale change and the number of objects of the source map. The Equation 1 computes the number of roads in the map after the line simplification operation:

$$n_T \approx n_S \sqrt{S_S / S_T}$$

Equation 1

Where n_S is the number of objects in source dataset; S_S is source scale; S_T is scale after transformation; and n_T is number of objects in the dataset after transformation. The source scale database of 1:250,000 national topographic roads have 1,202 segments and the derived roads have 856 segments after the line simplification and smoothing operations. When applying the *Radical Law* formula, there are 765 lines lines in 1:500,000 scale and 584 lines in 1:1,000,000 scale roads respectively (Figure 5). This means reduction in the complexity and the density of roads. These conform to the equation proposed for line simplification by Kazemi and Lim (2007) and Kazemi *et al.,* (2009).

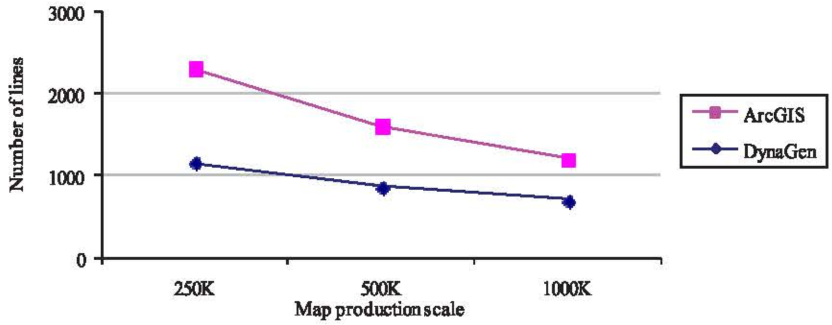

Figure 4: Assessment of generalisation results through ArcGIS™ and DynaGen™ systems; the source scale database is 1:250,000 (250K) national topographic data and outputs of generalised maps are road features at 1:500,000 (500K) and 1:1,000,000 (1000K) scales. The variation between ArcGIS™ and DynaGen™ generalisation results could be related to the implementation of simplification algorithms

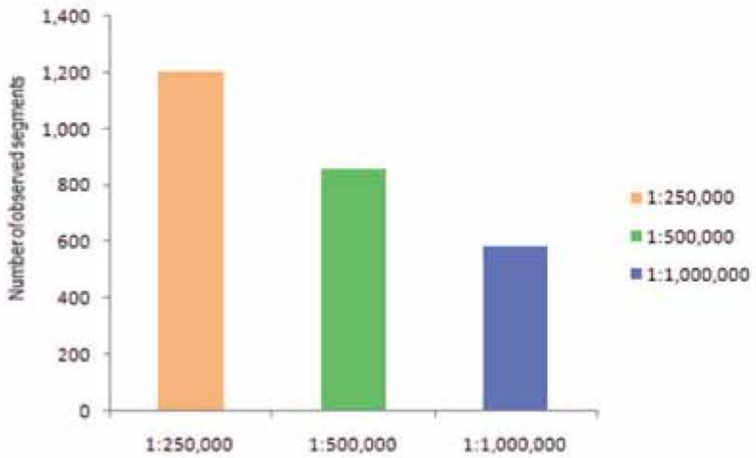

Figure 5: Number of observed road segments (lines) for 1:250,000 (the source scale database, TOPO250K), 1:500,000 (outputs of generalised road segments using TOPO250K), and 1:1,000,000 scales (outputs of generalised road segments using 1:500,000).

Table 2: A summary of the operations and algorithms available on the employed commercial generalisation platforms (adapted from Mackaness et al., 2007) and their effectiveness (based on the qualitative assessment and experts knowledge). The + and ++ mean low and high respectively

Operations / Algorithms	**ArcGIS™ Generaliser**	**DynaGen™**
Aggregation	Merging +	Line and area aggregation ++
Collapse	Area to Point Line to Point Point to Point Area to Line Area to Edge 2 Lines to Line ++	Area to Point Line to Point Point to Point Area to Line Area to Edge 2 Lines to Line +
Displacement	Editing mode +	YES +
Selection	Selection by Location Selection by Attribute ++	Selection by Location (queries) Selection by Attribute (using Intergraph Geomedia□) +
Simplification	Line and Area +	Line and Area ++
Smoothing	Line and Area +	Line and Area using Brophy & Averaging algorithms ++
Typification	NA	Yes +

These results have also been analysed qualitatively through visual assessment/inspection of the results of generalisation that was also carried out by superimposing the versions of roads at scales of 1:250,000, 1:500,000 and 1:1,000,000 successively with ETM+ imagery (composite bands 1, 2, 3) and respective raster maps (Table 2). Digitization of road segments based on high resolution-freely available imagery could provide basic parts of the road network which then, could be used as a reference for comparison with the generalization results. In this way, accuracy assessment would be a matter of comparing vector entities; and generalization results would be compared to more accurate than ETM+ images, road segments. However, it was not the focus of this study as per explained previously (see Section 2). The derived maps have reasonable agreements with the existing small-scale road maps such as the Global Map at 1:1,000,000 scale. The outputs from map generalisation have been analysed applying the *Radical Law* and changes in the representation of road features at 1:250,000, 1:500,000 and 1:1,000,000 scales as a result of generalisation were quantified (see Figure 5 and Figure 4). In addition, a quantitative evaluation of the results of generalisation is attempted applying positional accuracy measure. Accuracy of simplified and generalised outputs is determined by comparing the worst case and best case scenarios of positional displacement errors. The locations are assessed on the generalised maps (1:500,000 and 1:1,000,000) and corresponding positions from the published 1:250,000 national topographic maps and Landsat 7 ETM+ imagery. The source imagery has a basic horizontal accuracy (RMSE x,y) of approximately ± 15m (Smith *et al.,* 2003) and the '1:250,000 digital topographic data' has a basic horizontal accuracy of approximately ± 120 metres (GA, 2009). Positional accuracies for generalised roads (using 18 points from the Landsat imagery and the 1:250,000 data) are of the order of 140 - 180m RMSE values respectively; the measured displacements are approximately within the standards.

5. Conclusions and Recommendations

This paper introduced a knowledge-based approach to the generalisation of a road network database. This was conducted through application of the principles of generalisation using Intergraph's DynaGen™ software. The derived maps were satisfactory when compared with the existing small scale road maps such as the *Global Map* at scale of 1:1,000,000. The evaluation suggests that such generalisation methodology will be useful for building a practical generalisation framework and workflow. It has been experienced that the generalisation operations/algorithms, and their parameters embedded in the DynaGen™ system deliver coherent capabilities to automate the generalisation process for practical production applications. Dynamic generalisation has potential advantages. The following observations should be taken into consideration to maximise its advantages:

- To simplify roads, the Douglas and Peucker (1973) simplification algorithm produces satisfactory outputs. Topological and geometrical relations of adjacent objects are well managed when the generalised elements are close to each other and may overlap or intersect.
- For the feature elimination function, it is suggested that the database should be made more 'intelligent' by incorporating object-oriented technology to enable the software to define feature class and attribute conditions of features and enforcing this into map specifications by removing features that do not meet product requirements.
- To increase the efficiency of this approach, cartographic knowledge can be encoded into an expert system as was mentioned here by means of an example. An implementation of an expert system has been completed.
- To enhance current generalisation practice in national mapping organisations it is important to communicate generalisation problems and requirements, and to evaluate existing generalisation systems (e.g. comparing algorithms and approaches) and measure the fitness of a generalisation approach applying existing software. This recommendation is also supported by Stoter et al., (2004).

It is worth noting that the generalisation operations/algorithms and their parameters embedded in the DynaGen™ system deliver capabilities comparable to other existing systems such as ArcGIS™. The results from the work conducted here show the dynamic generalisation approach has potential for the generalisation of road networks. Building an expert system is recommended in order to integrate generalisation algorithms with cartographers' experience that will assist cartographers in choosing the appropriate techniques for map and database generalisation tasks such as feature displacement.

Acknowledgements

The lead author would like to acknowledge that this research was not possible without sponsorship and assistance of several individuals and organisations. In particular, the University of New South Wales (UNSW) granted the lead author with the Faculty of Engineering's Research PhD Scholarship and Supplementary Engineering Award (SEA). The authors are grateful to several reviewers including Professor Chris Rizos, Head of the School of Surveying and Geospatial Engineering of UNSW, Professor Gary Hunter of University of Melbourne, Professor Graeme Wright of Curtin University of Technology, Mr Graham Baker of Geoscience Australia (retired) and Dan Lee of ESRI Redlands. They participated in various discussions on the topic of generalisation and provided constructive suggestions toward the lead author's doctoral research. Special thanks are also extended to the University of California Berkeley for hosting Dr Sharon Kazemi as Visiting Scholar and Dr Alan Forghani as Visiting Fulbright Scholar from 2006 to 2007. We thank also the anonymous reviewers for their excellent feedback. The paper is extracted and modified from previous publications.

References

Alves, M., 2010, Polygonal Line Simplifying Methods Applied to GIS. *Proceedings of the FIG Congress 2010. Facing the Challenges – Building the Capacity,* 11-16 April 2010, Sydney, Australia, CD-ROM Procs, 9.

Arnold, L. M. and Wright, G. L., 2005, Analysing Product-Specific Behaviour to Support Process Dependent Updates in a Dynamic Spatial Updating Model. *Transactions in GIS,* Vol. 9, No. 3, 397-419.

Bart, G., Glaser, E., Sibler, P. and Weibel, R., 2013, Analysis of Swiss German Syntactic Variants using Spatial Statistics. In: Álvarez Pérez, X.A., Carrilho, E. and Magro, C. (eds.): *Current Approaches to Limits and Areas in Dialectology.* Newcastle upon Tyne: Cambridge Scholars Publishing, 143-169.

Benz, S. and Weibel, R., 2014, Road Network Selection for Medium Scales using an Extended Stroke-Mesh Combination Algorithm. *Cartography and Geographic Information Science*, 41(4): 323-339.

Bereuter, P. and R. Weibel., 2010, Generalisation of Point Data for Mobile Devices: A Problem-Oriented Approach. *Proceedings of the 13th Workshop of the ICA Commission on Generalisation and Multiple Representation,* 12-13 September 2010, Zurich, Switzerland, CD-ROM procs, 8.

Bereuter, P. and Weibel, R. 2013. Real-Time Generalization of Point Data in Mobile and Web Mapping using Quadtrees. *Cartography and*

Geographic Information Science. 40(4): 271-281.

Brophy, D.M., 1973. An automated methodology for linear generalisation in thematic cartography. *Proceeding of the American Congress of Surveying and Mapping Conference, 11 March 1973, Washington D.C., USA, pp. 300-314.*

Buttenfield, B., 2002, Transmitting Vector Geospatial Data Across the Internet. *Proceedings of the 2nd International Conference on Geographic Information Science,* 25-28 September 2002, Boulder, Colorado, USA, 51-64.

Chybicka, I., Iwaniak, A. and Wieslaw, O., 2004, Generalisation of the Topographic Database to the Vector Map Level 2 - the Components of the Polish National Geographic Information System. *Proceedings of the ICA Workshop on Generalisation and Multiple Representation,* 20-21 August 2004, Leicester, UK, CD-ROM procs, 14.

Douglas, D. H. and Peucker, T. K., 1973, Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or its Caricature. *The Canadian Cartographer,* Vol. 10, No. 2, 112-122.

Geoscience Australia (GA), 1999, Topographic Data and Map Specifications (CD-ROM). Technical Specifications, Version 3.4, National Mapping Division, Geoscience Australia. Accessed online 5 December 2004: http://www.ga.gov.au/mapspecs/250k100k/section3

Hardy, P., M.O. Briat, C. Eicher, and T. Kressmann, 2004. Database-driven cartography from a digital landscape model, with multiple representations and human overrides. *Proceedings of the ICA Workshop on Generalisation and Multiple Representation, 20-21 August 2004, Leicester, UK, CD-ROM procs, 12pp.*

Hunter, G. J., 1998. Generalisation Techniques for Spatial Data', 451-620 Lecture Notes, The University of Melbourne. McMaster, R.B. and Shea, S. 1992, Generalisation in Digital Cartography, The Association of American Geographers, Washington D.C., 3, 71-91.

Intergraph, 2004, DynaGen software documentation (CD-ROM). P.O. Box 240000, Huntsville, AL 35813, USA.

Iwaniak, A. and Paluszynski, W., 2001, Generalisation of Topographic Maps of Urban Areas. *Proceedings of the 20th International Cartographic* Conferences, 6-10 August 2001, Beijing, China, CD-ROM Procs, 12.

Accessed online 19 July 2003: Accessed online 5 July 2004: http://icaci.org/documents/ICC_proceedings/ICC2001/icc2001/author.htm

Joao, E. M., 1998, The Algorithmic Approach in Causes and Consequences of Map Generalisation, Taylor and Francis Ltd, London, ISBN: 9780849339325.

Karsznia, I., 2008, An Attempt of Road Network and Settlement Generalization in the General Geographic Database using DynaGEN Environment. *Geodezja i Kartografia.* Vol. 57, No. 2, 81-95.

Kazemi, S., Forghani, A., Cho, G. and McQueen, K., 2013, Integrated Cartographic and Algorithmic Approach for Road Network Database Generalization· *International Journal of Geoinformatics.* Vol. 9, No. 4, December, 2013, 73-86.

Kazemi, S., and S. Lim, 2007. Deriving multiscale GEODATA from TOPO-250K road network data. *Journal of Spatial Science, Vol. 52, No. 1, pp. 171-182.*

Kazemi, S., Lim, S. and Paik, H., 2009, Generalisation Expert System (GES): A Knowledge-Based Approach for Generalisation of Line and Polyline Spatial Datasets. *Proceedings of the Surveying & Spatial Sciences Institute Biennial International Conference.* 28 September - 2 October 2009, Adelaide, Australia, 717-729.

Kazemi, S., Lim, S. and Rizos, C., 2009b, Interactive and Automated Segmentation and Generalisation of Raster Data. *International Journal of Geoinformatics.* Vol. 5, No. 3, 65-73.

Kazemi, S., and Lim, S., 2007a, Deriving Multiscale GEODATA from TOPO-250K Road Network Data. *Journal of Spatial Science.* Vol. 52, No. 1, 171-182.

Kazemi, S., Lim, S. and Ge, L., 2005b, An Interactive Segmentation and Generalisation Framework for Mapping Features from Landsat 7 imagery. *Proceeding of the First International Symposium on Cloud-prone & Rainy Areas Remote Sensing (1st CARRS),* 4-8 October 2005, Hong Kong, 1285- 1296.

Kazemi, S., Lim, S. and Ge, L., 2005a, Integration of Cartographic Knowledge with Generalisation Algorithms. *Proceedings of the IEEE Geoscience & Remote Sensing Symposium.* 25-29 July 2005, Seoul, South Korea, 3502–3505.

Kazemi, S., Lim, S. and Rizos, C., 2004a, A Review of Map and Spatial Database Generalisation for Developing a Generalisation Framework. *Proceedings of the XIX Congress of 2004 the International Society for Photogrammetry and*

Remote Sensing, 12-23 July 2004, Istanbul, Turkey, 1221-1226.

Lecordix, F., Regnauld, N., Meyer, M. and Fechir, A., 2005, MAGNET Consortium. *Proceedings of the 8th ICA Workshop on Generalisation and Multiple Representation.* 7-8, July, 2005, La Coruna, Spain Zurich, CD-ROM procs, 8.

Lüscher, P. and Weibel, R., 2013, Exploiting Empirical Knowledge for Automatic Delineation of City Centres from Large-Scale Topographic Databases. *Computers, Environment and Urban Systems*. 37(1): 18-34.

Mackaness, W. A, Ruas, R. and Sarjakoski, L. T., 2007, A Synopsis of Generalisation Operations. In Generalisation of Geographic Information: Cartographic Modelling and Applications. Edited by W. A. Mackaness (University of Edinburgh, Scotland, U.K), A. Ruas (IGN Laboratoire COGIT, St. Mande, France), L. T. Sarjakoski (Finnish Geodetic Institute, Masala, Finland), April 2007, Elsevier Science, ISBN 13: 978-0-08-045374-3.

McMaster, R. B., 1986, A Statistical Analysis of Mathematical Measures for Linear Simplification: an Empirical Study. *The American Cartographer*. Vol. 13, No. 2, 103-116.

McMaster, R. B. and Shea, K. S., 1992, Vector-Based Generalisation in Digital Cartography. Association of American Geographers, 1710 16th St NW, Washington, D.C, USA, 71-98.

McMaster, R. B. and Shea, S., 1992, Generalisation in Digital Cartography. The Association of American Geographers, Washington D.C., 3, 71-91.

Meng, L., 1997, An in-Depth Investigation on the Long Lasting Problem of Cartographic Generalisation. Technical Report, May 1997, Swedish Armed Force, 1-75.

Muller, J. C., 1995, GIS and Generalisation: Methodology and Practice. Editor(s): J-P Lagrange; R Weibel; J. Muller, University of Ruhr, GISDATA Series. 20 April 1995, Bochum, Germany, ISBN: 9780748403196.

Pfer, T. and Pillewizer, W., 1966. The Principles of Selection. *The Cartographic Journal*. Vol. 3, No. 1, 10-16.

Ruas, A., 1998, A Method for Building Displacement in Automated Map Generalisation. *International Journal of Geographical Information Science*, Vol. 12, No. 8, 789-803.

Skopeliti, A. and Tsoulos, L., 2001, A knowledge Based Approach for the Generalisation of Liner Features. *Proceedings of the 20th International Cartography Conference*, 6-12 August 2001, Beijing, China, CD-ROM Procs, 10.

Smith, C.J.H., L. Wang, and N. Higgs, 2002. The refinement of Landsat 7 passes of the Australian continent with accurate ground control points. *Proceedings of the 11th ARSPC Conference, 2-6 September 2002, Brisbane, Australia, CD-ROM procs, 9pp.*

Stoter, J. E., Karrak, M. J. and Knippers, R. R., 2004, Generalisation of Framework Data: A Research Agenda. *Proceedings of the ICA Workshop on Generalisation and Multiple Representation, 20-21 August 2004, Leicester, UK.* Accessed online 4 April 2005: http://ica.ign.fr/bibliography/display_text.php?id_text=50

Venkateswaran, R., Weibel, R. and Purves, R. S. 2014, Exploring and Visualising Differences in Geographic and Linguistic Web Coverage. *Transactions in GIS*, 18(6): 852-876.

Weiss, R. and Weibel, R., 2015, Road Network Selection for Small-Scale Maps using an Improved Centrality-Based Algorithm. *Journal of Spatial Information Science*, 9: 71-99.

Technitis, G., Othman, W., Safi, K. and Weibel, R., 2015, From A to B, Randomly: A Point-To-Point Random Trajectory Generator for Animal Movement. *International Journal of Geographical Information Science.* DOI: 10.1080/13658816.2014.999682.

Visvalingam, M. and Williamson, P. J., 1995, Simplification and Generalisation of Large Scale Data for Roads - A Comparison of Two Filtering Algorithms. *Cartography and Geographic Information Systems,* Vol. 22, No. 4, 264-275.

Watson, P. and Smith, V., 2004, Interoperability of Agent-Based Generalisation with Open, Geospatial Clients. *Proceedings of the ICA Workshop on Generalisation and Multiple Representation*. 20-21 August 2004, Leicester, UK, CD-ROM Procs, 5.

Yang, W. and Gold, C. M., 2004, A System Approach to Map Generalisation. *Proceedings of the ICA the Workshop on Generalisation and Multiple Representation*. 20-21 August 2004, Leicester, UK, CD-ROM Procs, 10.

Zeiler, M., 1999, Modelling Our World - The ESRI guide to geodatabase design. ESRI Press, Redlands, California, USA, ISBN 1-879102-06-1.