# A Machine-Learning Approach to Generalisation of GIS Data

## Forghani, A.,<sup>2,3\*</sup> Kazemi, S.<sup>1</sup> and Bruce D.,<sup>2</sup>

<sup>1</sup>Murray-Darling Basin Authority, GPO Box 1801 Canberra ACT 2601Australia <sup>2</sup>Science, Technology, Engineering and Mathematics (STEM), University of South Australia (UniSA), Adelaide, Australia, E-mail: alan.forghani@unisa.edu.au <sup>3</sup>Faculty of Science & Technology, University of Canberra (UC), Canberra, Australia *\*Corresponding Author* 

## Abstract

The automation of map generalisation in this study involves an expert system approach that consists of four main components including knowledge acquisition, an inference engine, knowledge representation and a user interface. The acquired knowledge was then utilised to build a knowledge-based solution: a 'Generalisation Expert System' (GES) developed in Java, Python and C programming environments for the delivery of generalised geographical features. Its capabilities are demonstrated in a case study through generalising several line and polyline databases over the study area in Canberra, Australia. The cartographic and GIS software communities will benefit from this study through access to a set of tools, guidelines and protocols that incorporate a standardised cartographic generalisation methodology. The results of the trials utilising GES were analysed: a series of generalisation routines were performed to assess the quality of simplification results for different spatial layers. Cartometric measures such as the total length and number of line or polyline segments were used as indices of generalisation to quantify generalisation performance for the target small scale. For example, there are 101,228 segments in 1:250,000 scale and 9,491 segments in 1:500,000 scale contours over the study area. This requires a reduction in the complexity and the density of elevation data. Changes in the representation of contour features at 1:250,000 and 1:500,000 scales as a result of generalisation were quantified. Outputs from map derivation have been analysed applying the Radical Law, this determines the retained number of objects for a given scale change and the number of objects of the original source map. Testing demonstrated that the implemented algorithms in GES are able to extract characteristic vertices on the original entity lines and polylines (e.g. for roads) while excluding non-characteristic lines and polylines to reduce irrelevant computation. This study has demonstrated reasonable improvements in Vertex Reduction, Classification and Merge, Enhanced Douglas-Peucker and Douglas-Peucker-Peschier algorithms. The test results show that GES generalises line features accurately while still maintaining their geometric relations. Existing generalisation software requires advanced technical skills from users; GES however, has a basic and user friendly Graphical User Interface (GUI) which is an advantage to users with basic technical skills and understanding of spatial data management. Changes to geographic parameters should be updated in multi-scale maps and spatial databases in near real-time. GES can be developed as a potential tool for generalising largescale maps into smaller scales, and creating maps of different themes across a variety of scales. Test results have also demonstrated that the methodology developed improves the efficiency of line and polyline generalisation. This study aims to contribute to generalisation system design and the production of a clear framework for users. Experiments presented in this book can be applied to real world problems such as the generalisation of road networks and area features using GES. Future research should be directed towards developing web mapping platforms with generalisation functionality at varying scales.

## 1. Introduction

Spatial intelligence decision-making relies on accurate, economical and viable digital spatial information products which underpin 'egovernment' initiatives and location-based services. Automated generalisation systems reduce the cost of maintaining multiple data models and digital maps at different scales. To facilitate automated generalisation, a detailed generalisation framework for deriving multiscale spatial data has been developed by the author (Kazemi and Forghani, 2016). This involved an assessment of existing generalisation systems, undertaking an international survey of cartographic generalisation practices and developing a knowledge-based expert system on feature generalisation. In order to overcome inadequate handling of scale inconsistency in modern day spatial updating efforts, a Generalisation Expert System (GES) has been built



Java-Python-C, enabling automated generalisation of thematic data at a range of scales. This is achieved by a new approach to spatial data revision that assimilates a scale-independent data model and interactive cartographic generalisation processes.

The approach takes advantage of NMA, SMA and a number of software vendor's inputs in relation to the knowledge acquisition process of cartographic practices, via the cartographic generalisation survey reported by Kazemi and Forghani (2015). The findings from the survey are used to help define the requirements for a knowledge-based spatial database in which the cartographic knowledge and heuristic rules are formulated as a series of rules. These feed into a GES to deliver coherent capabilities and automate the generalisation of features as much as possible for 'derivative mapping'. This paper reports on the development of a GES. A brief description of key steps in building a GES and its components are presented. Its capabilities will be demonstrated in a case study involving the simplification of 1:250,000 national topographic data to 1:500,000 scale over Canberra, Australia. The GES has a simple Graphical User Interface (GUI) that can assist users without requiring a high level of technical skill and knowledge of spatial data management.

#### 2. Methodology

In this study a rule-based generalisation expert system was developed. It is interfaced with GIS software. Fundamentals of the expert system incorporate cartographers' knowledge and intuition to generalise lines and polylines. Rules are symmetrical so they can be processed in either a forward or a backward approach. A forward chaining (data-driven) approach attempts to reach a conclusion through bottom-up reasoning; initially starting from the proof and details, to the top-level conclusions that are based on facts. The operator begins by assuming a conclusion to be true and then applies the rules to prove it (Giarrantano and Riley, 1989). In contrast, a backward chaining (goaldriven) approach starts processing a given problem from the hypotheses (top-down inference) down to the lower level facts that support the hypotheses. It begins with a conclusion and proves the conclusion via providing the veracity of each premise in a left to right, or top to bottom order. As explained by Kazemi and Forghani (2015), expert systems are composed of four key parts: (a) knowledge representation; (b) inference engine; (c) knowledge representation; and (d) user interface (see Figures 1a, 1b and 1c). These are implemented specifically for semi-automated road network generalisation. The research methodology consists of three major components:

- Cartographic knowledge acquisition, which was completed through a survey, resulting in production of a series of cartographic rules to build a GES;
- An assessment of existing generalisation systems as a test bed; and
- Collection and processing of various spatial datasets.

GES uses operations, algorithms and knowledgebased rules that are elaborated on in the following sections of this chapter. Previous research on road generalisation was reviewed in Peschier, (1997), Kazemi et al., (2004a and 2004b), Kazemi et al., (2016) and Forghani and Kazemi (2016). Key components and the methodology of this research are shown in Figures 1(a)-(c) and include the following:

- A detailed review and evaluation of current generalisation methods and solutions.
- Developing a conceptual methodology to generalise 1:250,000 national topographic data into 1:500,000 and 1:1,000,000 scales.
- Cartographic knowledge acquisition by undertaking an International Cartographic Generalisation Survey and analysis of the survey results to build a rule-based expert system.
- Rule collection and encoding in Java-Python-C programming languages to construct the GES.
- Demonstration of GES application for spatial data mining and generalisation for three different spatial features.
- Assessment of the overall generalisation performance using referenced maps and experienced cartographer's feedback.

## 2.1 System Architecture and Key Features

A graphical overview of GES architecture is shown in Figures 2 and 3 and its GUI, key features, tools, and windows are presented in Figure 4. The system is built in the Java-Python-C programming environments with input from the cartographic knowledge acquisition process, based on the International Cartographic Generalisation Survey (Kazemi and Lim, 2007 and Kazemi and Forghani, 2015). The system consists of four main components: graphical interface, setting, algorithms, and outputs of spatial attribute data.







*International Journal of Geoinformatics,* Volume 17, No. 2, April 2021 Online ISSN 2673-0014/ © Geoinformatics International Knowledge--Based Generalisation of Geospatial Data 41-59



Figure 1c: Architecture of GES

Each component is explained in the following sections. The main knowledge-based contextual information is as follows; these characteristics assume that roads are already identified and mapped:

Geometry or Topology of Features Such as Roads: Topology maintains spatial relationships among geographic features. During generalisation processes, GIS tools handle topology through a set of validation rules that define how features are to share a geographic space (e.g. polygons should not overlap and lines should not cross) whilst preserving shared geometry, connectivity and other related topological relationships (Lee, 2004). The shape of a road is an important factor for a contextual classification. Functional requirements, terrain and engineering limitations affect the geometrical and

physical form of a road. For example, slope, width, and curvature of a road require an upper boundary. Roads can intersect but rivers join. In the GES, the existing road map was added to the knowledgebased data. Referenced road maps contain clues that guide the spatial analyst to recognise analogous roads according to structural characteristics and other parameters. ArcGIS<sup>TM</sup> is used to construct the topology of features in order to obtain input to the GES. For example, 'undershoot' errors can be removed by ArcGIS<sup>TM</sup>. This maintains connectivity of road segments in the road coverage.



Figure 2: Roadmap showing the rationale and steps for organisation of the chapter and testing GES



Figure 3: Graphical user interface of GES

![](_page_4_Figure_6.jpeg)

45

Land Use: Land use classification typically relies on visual interpretation. However, this model is not sufficient for roads that do not exist on maps or extend over multiple regions because, in general, the density of roads is related to the type of land use. For example, in a high density urban area, the road network is likely to be complex and dense. Different types of land cover can be associated with different types of road network topology. That is, many crossroads are connected to many road segments in an urban area. This contextual information enables the analyst to extract roads.

Drainage Pattern: The drainage pattern is also incorporated, as it has an effect on the appearance of road structures. Roads normally follow contour lines in valleys but are less curved than channels and rivers. This is essential for road construction in order to minimise the number of bridges on roadriver crossings. Co-linearity and connectivity are also considered, and drainage networks are used to avoid confusion between water and bridges as a road segment may be bounded by water.

*Elevation:* Topography has an effect on the appearance of road structures. A digital elevation model can be used to indicate plausible road tracks in an image. Even in a mountainous area, a road between regions (e.g. towns and countryside having almost the same altitude) follows a similar altitude. In an area with a high slope, a line is unlikely to be a road unless it is approximately parallel to the contours, although there are of course exceptions. For example, line elements (e.g. fire lanes in a forest) are at right angles to the relief. Therefore, an elevation layer is employed in the GES.

Induction is here taken to be a kind of reasoning from premises. Rules of inference are systems for philosophical making deductions. Within discussions 'rule of inference' and 'deduction' are often used loosely and interchangeably. In this paper the deduction was used as a step where one removes the most recent hypothesis (still available) and prefixes it to the previous step. This is shown by unindenting one level. In mathematical logic, the deduction theorem is a metatheorem of propositional and first-order logic that is a formalisation of the common proof technique in which an implication  $A \rightarrow B$  is proved by assuming A and then deriving B from this assumption conjoined with known results. The deduction theorem explains why proofs of conditional sentences in mathematics are logically correct. Though it has seemed "obvious" to mathematicians literally for centuries that proving B from A

conjoined with a set of theorems is tantamount to proving the implication  $A \rightarrow B$  based on those theorems alone. The deduction theorem is sometimes taken as a primitive rule of inference. In this study, rule deduction deals with the knowledge acquisition stage that is known as the inference mechanism.

In this phase the information generated and collected from the prior phases is aggregated into a rule-based view to maintain consistency and reliability for the utilisation of a cartographers' experience in feature generalisation. Once the knowledge is collected in an appropriate form, the next step is to convert the collected knowledge into a programming set in order to build a rule-based expert system and to develop procedural code to execute instructions against a database. A rule is a combination of knowledge that represents an antecedent or condition and its immediate consequence or conclusion. Examples of generic rules used in the construction of the GES are shown in Figure 4. Rule deduction is concerned with knowledge representation and map production, which are driven by cartographic knowledge using the spatial database. Many GIS applications have been inductive rather than deductive (e.g. Forghani, 1998 and Forghani, 2000). Inductive structures deal with finding general rules and inferring from examples. On the other hand, deductive structures are those in which a known general relationship is employed for particular observations. This allows the users to query the identification of all areas in which a known relationship or desired set of premises are satisfied.

The forward chaining process is considered in order to evaluate all rules for a given feature segment. The computation time linearly increases as the number of vertices and lines increases. The forward chaining (bottom-up) search is time consuming since it leads to a larger number of choices. This is particularly problematic if the database is large, i.e. the number of features and rules is large. For example, if there are 1121 line segments in an elevation layer, then a significant number of vertices have to be processed. This process takes approximately 8 hours using Compaq Presario v2000 Notebook (Windows XP, Intel Pentium M 710 - 1.4 GHz, 256 MB RAM, and 40 GB hard drive). In addition, the more rules, the more computation time is required. In practice, it is tedious and time-consuming to use such a database since it requires a powerful computer with huge RAM and hard drive space. In this study, a small subset of the data over Australia's capital city, Canberra (covers approximately 23,630 km<sup>2</sup>) is tested (Figure 3). Generalisation operations and

Agorithms of the GES (Kazemi, et al., 2013; Kazemi and Forghani 2015 and 2016) are shown in Figure 4. Some require further implementation/ development (Table 1). The three main modules in the GES include: shapes.py, a class of shape objects supported by the GES for Points, Areas (Polygons) and Data Structure. It also includes a set of data structures representing entities such as maps, tables, points and generalisation algorithms.

TC 1	1	1 0	COLO	1			1	1 .	1
Igh	10	I · Statuc o	T I VHN	General	1001101	onerations	anda	laorit	hma
1 au	IC.	L. Dialus U		gunua	isation	obciations	s anu a	πεσπι	IIIII5
				D					

Generalisation operations and algorithms	Comment
Selection	This operation is operational in GES
Classification	This operation is operational in GES
Typification	This operation is immature in GES
Symbolisation	This operation is immature in GES
Vertex Reduction	This algorithm is operational in GES
Merge	This algorithm is operational in GES
Douglas-Peucker- Kreveld (1973) and Peschier (1997)	This algorithm is operational in GES

![](_page_6_Figure_4.jpeg)

- # Threshold of Line Tolerance for DP algorithm. Default Value is 0.0 # If Line Tolerance <= 0.0, GES will not run DP algorithm even if Apply DP = Yes
- # default Value is 0.0Line \_Tolerance = 0.005

Rule1

![](_page_6_Figure_8.jpeg)

# Threshold of Point Tolerance for vertex reduction.
# If Line\_Tolerance <= 0.0, GES will do nothing
# default Value is 0.0Point \_Tolerance = 0.15</pre>

Rule2

Figure 4: Examples of GES rules

Table 2: GES executable programs

Executable programs	Comment				
shpparser.c/shpparser.exe	parse input shape data (binary) and then output as plain text file (ASCII)				
dbfparse.c/dbfparse.exe	parse input dbf data (binary) and then output as plain text file (ASCII)				
shpwriter/shpwriter.exe	write out final results into shape file (binary)				
dbfwriter/dbfwriter.exe	write out final results into dbf file (binary).				

## 2.1.1 Communication among components

GES consists of three main components that are implemented in various programming languages. Communications between these components are carried out in varying ways:

- Direct calling using built-in functions:
  - o Java methods call C executable programs
  - Python algorithm methods call C executable programs
- Indirect calling uses Unix/Linux Shell Scripts as a bridge:
  - Java GUI calls Python algorithms using 'gui\_run\_algorithm.sh'

• The returned data is stored in temporary plain text files.

GES uses the Shapefile C Library [open source software] (http://download.osgeo.org/shapelib or http://shapelib.maptools.org/dl) read and write shape and dbf files (binary). There are four executable programs in the GES that were developed from the Shapefile Library (Table 2).

#### 2.1.2 Java client for generalisation definition

A Java interface through a GUI is provided to set up the parameters for the generalisation processes. Figure 4 shows visualisation of input and output for the interactive completion in case user/cartographer

Intervention is required. Spatial data generalisation requires a large number of user/cartographerdefined parameters. In the GES (Kazemi, 2011) the majority of those parameters are established by the mapping specification, defining what is to be achieved, coupled with a set of constraints that introduce limits on allowable change (Neuffer *et al.*, 2004). Main features/functions include:

- Choice of user-defined knowledge rules for defining thresholds and algorithms.
- Choice of displaying different parts of GIS data, e.g. lines, vertices, and labels.
- Choice of display colours for different sets of GIS data.
- Displaying both input and output GIS maps in the same window for comparison.
- Highlight selected shape files with prompt text.
- Zoom and/or move the GIS map.

## 2.1.3 C client for GIS data interoperability

This component utilises the open source Shapefile C Library (Version 1.2), developed by Warmerdam (1999), to parse input shape files data and generate output shape files. The Shapefile C Library provides the functions for developing simple C programs for reading, writing and updating (to a limited extent) ESRI shapefiles, and their associated attribute files (.dbf). Four C-based standalone mini-programs have been developed (see Figure 1b).

2.1.4 GES directories and coding components The following provides a summary of GES descriptions for the directories:

# **Main Directory**

#### • GES Sub-Directory

- bin: Executable C program, Unix Shell Scripts
- classes: Compiled Java byte code classes for GUI
- config: Knowledge rules and GesLog.log
- data: Example input shp/dbf files
- doc: Documentation for GES. Currently includes only documents for python components, e.g. algorithm.html, datastructure.html and shapes.html
- lib: The Shapefile C library
- python: Python code, including three modules and driver class
- Src: Java source code for GUI
- Tmp: Temporary directory for storing temporary output files, e.g. tempdata\_shp.out and tmpdata\_dbf.out
- start\_ges.sh a shell script for running the GES Java GUI

To run the GES the user requires a Unix-like environment. The Cygwin is free software that provides a Unix-like environment on a Windows platform. The user needs to download and execute setup.exe from the Cygwin website: http://cygwin.com/setup.exe

#### To start the GES program:

- Run *Cygwin*. This requires command prompts such as the syntax below
- *pwd* to check current location then change the directory to where *ges* folder is
- *cd ges* and run ./*start\_ges.sh*

## 2.1.5 Input parameters

The inputs include shape and dbf files (binary) and knowledge rules (e.g. 'ges\_rule1.dat' in the 'config' directory). Refer to example below:

## EXAMPLE

*ges\_rule1.dat (in the 'config' directory)* Specifies all rules (algorithms and thresholds) to be applied to the GES program. Lines starting with # are the comment lines and will be ignored by the program.

# This the knowledge rule file for GES
# Flag for whether or not to print out processing information to the GesLog.log file (Yes or No)
# Default value is Yes.
Write_Log_File = Yes
# Flag for whether or not to mark key points before processing (Yes or No)
# Default value is No.
Mark_Key_Points = Yes
# Flag for whether or not to merge polylines which share the same key properties (Yes or No)
# Default value is No.
Merge_Polylines = Yes
# Flag for whether or not to allow overlapped line segments before simplification (Yes or No)
# Default value is Yes.

Line Overlap = Yes # Threshold of Point Tolerance for vertex reduction. # If Line Tolerance  $\leq = 0.0$ , GES will do nothing # Default value is 0.0 Point Tolerance = 0.0015# Flag for whether or not to apply Douglas-Peucker (DP) algorithms (Yes or No) # Default value is No. Apply DP = Yes# Threshold of Line Tolerance for DP algorithm. Default value is 0.0 # If Line Tolerance  $\leq = 0.0$ , GES will not run DP algorithm even if Apply DP = Yes # Default value is 0.0 Line Tolerance = 0.005# Threshold of Minimum Line Segment Length. # Any line segments whose length is less than this threshold will be removed. # If Min Line Segment Length  $\leq = 0.0$ , GES will do nothing *# Default value is 0.0* Min Line Segment Length = 0.0

## 2.2 Generalisation Outputs

The outputs consist of temporary ASCII (plain text) shape and dbf files, binary shape and dbf files, and a log file to store runtime processing information ('GesLog.log' in the 'config' directory). See example below:

EXAMPLE					
GesLog.log (in the 'config' directory)					
Records all running information for each execution of the GES. The contents of this log file can be used for testing and comparison of results under different input conditions. For example, how long does it take to run a process (CPU time), how many shapes and points (vertices) have been reduced after execution (input shapes against output shapes)					
EXAMPLE					
1. Inputs: Elevation					
Shape Type = Arc					
Minimum Bounds(148.763, -35.92, 0, 0) Width = 0.635					
Maximum Bounds(149.398, -35.126, 0, 0) Height = 0.794					
Number Of Shapes = $1121$					
Number Of Points = $101228$					
2. Parameter Settings:					
Mark Key Points = Yes					
Merge Polylines = Yes					
Allow Line Overlap = Yes					
Point Tolerance $= 0.025$					
Apply DP Algorithm = Yes					
Line Tolerance $= 0.025$					
Minimum Line Segment Length $= 0.0$					
3. Outputs:					
Number of shapes $= 1121$					
Number Of Points = 9491					
4. Total CPU Time (h:m:s):					
8:8:51.267					

GES is able to generalise both lines and polygons. These differ from each other in terms of algorithmic implementation in the GES. For lines, GES will execute the following tasks:

• Find all crossing vertices for lines (e.g. roads), those lines/roads have higher road classes (>3).

Mark these crossing vertices as Key Points which must be preserved.

- Merge lines/roads that have the same road name and can be connected together. This will reduce the number of shapes but the total number of vertices will be the same.
- Apply Douglas-Peucker algorithm to those merged lines (e.g. roads).

## .2.1 Input dataset for evaluation

Parts of road, elevation and vegetation layers from the national topographic database of Australia (1:250,000 national topographic data) for Canberra, Australia, were selected as lines and polylines to be generalised and simplified. A case study of real time roads, vegetation and elevation datasets applying different rules and input parameters over this area is shown in Figures 4-6. The following operations are required to generalise a road network. The outputs of generalisation workflow applying these operations are shown in Figures 4-6.

#### 2.2.2 Contour generalisation

There is an ample of literature regarding generalisation of contour lines on a map which joining points (closed curve) of equal height or depth above or below a level, usually mean sea level (e.g. Fei, 1998, Zhang et al., 2009, Ai, 2019 and Yan, 2019). The height or depth difference between successive contour lines is the contour interval as becomes greater and greater when the map scale becomes less and less. Generalisation of raster data such as GA digital elevation models (DEMs) which are arrays of regularly spaced elevation values referenced horizontally into a Universal Transverse Mercator (UTM) (cartographically) is another form of generalising elevation data in essence. It is rational to think that contours support visual communication, while serving as a representation of a surface.

Contour simplification was carried out in order to generate cartographically acceptable graphic shapes for 1:500,000 scale topographic maps.

![](_page_9_Figure_8.jpeg)

Figure 4a: Road network generalisation of a 1:500,000 map from the original 1:250,000 map Geoscience Australia © 2001

![](_page_9_Picture_11.jpeg)

![](_page_10_Figure_0.jpeg)

Figure 4b: Examples of improvements in node junctions, bends and curvatures

![](_page_10_Figure_2.jpeg)

Figure 5: Contour generalisation of a 1:500,000 map from the original 1:250,000 map with 100m interval, Geoscience Australia © 2001

*International Journal of Geoinformatics,* Volume 17, No. 2, April 2021 Online ISSN 2673-0014/ © Geoinformatics International Knowledge--Based Generalisation of Geospatial Data 41-59

51

![](_page_11_Figure_1.jpeg)

Figure 6: Native vegetation generalisation of a 1:500,000 map from the original 1:250,000 map. The area includes vegetation polygons from the range of native vegetation classifications described as part of the 1:250,000 national topographic data specifications, Geoscience Australia © 2001

The Douglas-Peucker (DP) simplification algorithm was used; firstly, contours were simplified with a 0.024m tolerance, and then line-crossings with the error band contours were checked and the line segments were marked. A visual output of the simplification results with original contours is depicted on Figure 5. Contours were simplified within the defined vertical positional accuracy, but maximum horizontal positional accuracy was defined by the simplification tolerance used in the algorithm. The simplification tolerance plays a function in determining the precincts of the simplified contours for flat areas. The simplified contours were obtained within defined spatial accuracy (GA, 1999 and 2001).

#### 2.2.3 Generalisation of native vegetation

Users of native vegetation data require different levels of detail that vary from the species of an individual stand, to a class or community of vegetation. These demand different levels of abstraction, i.e. at different scales to satisfy the requirements of various applications. A review of literature necessitates finding an answer to questions such as how to aggregate, what should be aggregated, when to aggregate and how much abstraction is required. This can be achieved by using rule-based generalisation of bio-geographical principles and the spatial distribution of vegetation, which enables automation of the process of multiscale representation of vegetation.

The 1:250,000 native vegetation of Canberra is taken as a case study of polygon generalisation (Figure 6). To produce simplified vegetation patches, the seven land cover categories were identified, starting from the most detailed information at the largest scale possible. In addition to the above mentioned features, several examples of applying knowledge-based rules (rule 1 & 2) over various features are presented by Kazemi (2011).

### 3. Results

The results were analysed (Figure 10 and Tables 3-5). A series of generalisation trials were conducted to gauge the sensitivity of simplification results to the different spatial layers. Generalisation performance comparisons are summarised in Tables 3-5. According to McMaster (2001)'s cartometric measures, the total length and number of line/polyline segments is used as an index of generalisation generalisation to quantify performance for the target small scale (Figure 10). For example, there were 101,228 segments and 9,491 segments in the 1:250,000 scale and the 1:500,000 scale contour data respectively over the study area. This means reduction in the complexity and the density of elevation data. Changes in the

representation of contour features at 1:250,000 and 1:500,000 scales as a result of generalisation are quantified. The outputs from map derivation have been analysed applying the *Radical Law* (Topfer and Pillewizer, 1966) both employing the GES. The *Radical Law* (Topfer and Pillewizer, 1966 and Kazemi and Forghani, 2016) determines the retained number of objects for a given scale change and the number of objects of the source map. In addition, the outputs compare favorably with GA source datasets such as 1:500,000 and 1:1,000,000.

The generalisation over roads and vegetation performed better than contours (elevation) data. The outputs of roads, elevation and vegetation generalisation are shown in Figures 5-8 respectively. The results have been visually assessed through cartographic assessment. Feedback from both the qualitative (expert assessment) and quantitative (cartometric measures) evaluations using measures of goodness-of-fit (Visvalingam and 1990) was incorporated Whyatt, into the development of an enhanced version of GES (v1). A case study of line and polyline generalisation; the original input GIS maps; and the model-based graphical generalised map were demonstrated earlier (Figures 4 - 6).

![](_page_12_Figure_7.jpeg)

![](_page_12_Figure_8.jpeg)

Table 3: Roads network generalisation input and output parameters for deriving 1:500,000 maps from the source data of the 1:250,000 national topographic data. Some of the key input and setting parameters include Shape Type = Arc, Minimum Bounds (148.763, -35.92, 0, 0), Width = 0.635, Maximum Bounds (149.398, -35.126, 0, 0, Height = 0.794, Number of Shapes = 592, Number Of Points = 51998, Mark Key Points = Yes, Merge Polylines = No, Allow Line Overlap = Yes, Point Tolerance = 0.024, Line Tolerance = 0.024, and Apply DP Algorithm = Yes.

	]	Inputs		Outputs			
Number of	Number	Point	Line	Number of	Number	Total CPU Time (h:m:s)	
Shapes	of Points	Tolerance	Tolerance	shapes	of Points		
456	3457	0.015	0.015	240	693	0:0:1.584	

Table 4: Elevation generalisation input and output parameters for deriving 1:500,000 maps from the source data of the 1:250,000 national topographic data. Some of the key input and setting parameters include Shape Type = Arc, Minimum Bounds (148.763, -35.92, 0, 0), Width = 0.635, Maximum Bounds (149.398, -35.126, 0, 0, Height = 0.794, Number of Shapes = 592, Number Of Points = 51998, Mark Key Points = Yes, Merge Polylines = No, Allow Line Overlap = Yes, Point Tolerance = 0.024, Line Tolerance = 0.024, and Apply DP Algorithm = Yes.

		Inputs		Outputs		
Number of	Number	Point	Line	Number of	Number	Total CPU Time (h:m:s)
Shapes	of Points	Tolerance	Tolerance	shapes	of Points	
592	51998	0.024	0.024	592	3541	2:2:38.7

Table 5: Native vegetation generalisation input and output parameters for deriving 1:500,000 maps from the source data of the 1:250,000 national topographic data. Some of the key input and setting parameters include Shape Type = Arc, Minimum Bounds (148.763, -35.92, 0, 0), Width = 0.635, Maximum Bounds (149.398, -35.126, 0, 0, Height = 0.794, Number of Shapes = 7, Number Of Points = 6085, Mark Key Points = Yes, Merge Polylines = Yes, Allow Line Overlap = Yes, Point Tolerance = 0.035, Line Tolerance = 0.035, and Apply DP Algorithm = Yes.

	]	Inputs		Outputs		
Number of	Number	Point	Line	Number of	Number	Total CPU Time (h:m:s)
Shapes	of Points	Tolerance	Tolerance	shapes	of Points	
7	6085	0.035	0.035	7	338	0:0:6.916

Selection is considered a pre-processing stage where the content of the map is determined. Figures 6-9 show the selection process for cartographic features. The features and their corresponding attributes required for the composition of the map are selected and retrieved from the relevant spatial database layer. Scale and map particularities are in use throughout the selection. In the expert system environment the cartographer introduces the category, the scale, and the boundaries of the new map or chart and the system identifies the layers that can be used (original selection). The selection of the features to be portrayed on the map is realised in the GES. The selected features are transferred to and organised in the expert system environment, and those to be considered for portrayal are chosen in accordance with their thematic characteristics (thematic selection). Quality assessment was carried out firstly by a direct visual comparison between the output (Figures 8-10) and the existing maps (1:500,000 and 1:100,000). Notwithstanding the fact that comparison with paper maps is a subjective

matter as outlined in other studies (e.g. Chaudhry and Mackaness, 2008), it nevertheless provides an indicative success measure for the algorithm. The results of the GES were compared with existing digital datasets and the results were in high agreements when compared with existing generalised mapping products (Geoscience the digital 1:500,000, Australia's 1:100,000 topographic data and 1:1,000,000 Global Map data). It was noted that the algorithm had maintained the overall topological relationships successfully.

It is worthwhile to evaluate generalisation of roads through the use of positional accuracy measures, which is an assessment of the closeness of feature location (e.g. road segments) in the dataset in relation to their true positions on the Earth's surface. The positional accuracy generally includes a horizontal accuracy uncertainty, a vertical accuracy uncertainty, and an explanation of how the accuracy uncertainties were determined (ANZLIC, 2001). The horizontal positional accuracy assessment of roads is conducted once all geometric

ransformations have taken place. Measures such as root mean square error or standard deviation can be used to represent the variation of vertical accuracy of roads. Spatial accuracy describes the positional (coordinate) accuracies of spatial data as a result of generalisation process (multiple-scales). the Generalisation quality assessment is a key subject in modern cartography and its automation process influences the 'fitness for use' of spatial data. Thus quality assessment is considered indispensable for generalisation of roads in this study. This section reports on a process used for assessing the positional accuracy of generalised roads. Among several sources of errors identified by researchers (Burrough, 1986, Goodchild and Hunter, 1997; and Hope et al., 2006), topological and generalisation errors are relevant to this study. The spatial errors of linear features composed of scale-dependent errors as a result of generalisation, and sampling of a line of high geometric accuracy represents the amount of deviations between the interpolated line and the original position of the linear feature. Spatial relationships across road features in a multiple feature type database could offer additional information to support the positioning of road segments. This requires assessing topological and positional relationships to a particular road segment. It is possible to investigate the residuals of displaced road intersections using GPS survey points (Hope et al., 2006). The geometric accuracy of linear features such as roads consists of two key parameters, including positional point accuracy for well-defined points on the road network (e.g. intersections), and shape conformity for a road segment compared to another segment indicating to what extent the curvature of the two segments are similar.

Accuracy of simplified and generalised outputs is determined by comparing the positions of ten well-defined ground control points (GCPs) (here defined as road intersections). The point 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. The source map '1:250,000 digital topographic data' has a basic horizontal accuracy of approximately ±120 metres (GA, 2009). Cartographic generalisation of line and area features introduces errors into the derived map. Generalised output maps have been checked against other data sources including Landsat ETM+ satellite imagery, published 1:250,000 digital topographic data and GPS field information. The GPS survey datasets were collected by Geoscience Australia (GA) in February 2007. Post-processing for the 70 GPS survey points was performed by Ultimate Positioning Pty Ltd. GA made available a copy of these data sets for this

55

study. It would be useful to include the photographic location of the survey points but unfortunately no picture was taken during the GPS survey. Ten GPS survey points were used for this purpose (Figure 8) as these points were within the study area. The accuracy of the collected survey data was estimated to be at the sub-metre level.

About 50 GCPs were identified from the 1:25,000 scale topographic data. These GCPs and GPS survey data were overlaid onto the generalised output maps. The tracks and GPS points did not match the road as it was shown on the generalised output maps because of the displacement process, as a result of the simplification operation when originally applied to the 1:250,000 digital topographic data and also during small scale map derivation. For example, to make a map more readable when having multiple adjacent features, one feature may be preserved in its true position and the others may be displaced. Usually hydrographic and transportation features (e.g. railways, roads, tracks) are preserved in their correct positions compared to a real features such as buildings and native vegetation boundaries. Positional accuracies achieved for generalised 1:500,000 and 1:1,000,000 maps are  $\sim 120$ m and  $\sim 470$ m respectively (see Kazemi, 2011). The overall agreement among GPS points and road intersections provides an accurate map of a simplified roads database. Measured errors are therefore considered to be within defined mapping standards. Vertical accuracy assessment was beyond the scope of this study, but would be a worthwhile to test in future studies. In Figure 10 the results achieved for 1:500,000 and 1:1,000,000 scales are shown for most roads. Displacements of up to 510m among 10 road intersections were surveyed. In the case of Old Boboyan Road (number 10), there appears to be a displacement of 710m. In relation to selected roads intersections, Figure 9 shows the displacement values of the coordinate differences from the original datasets (before generalisation) and derived road map (after generalisation), ranging from 2m to 710m for a generalised road map at 1:500,000 (black line) and 1:1,000,000 (blue line) scales. The residuals of displaced road intersections increase as a function of the scale of roads and the geometric characteristics of road segments. For example, the Old Boboyan Road reveals the largest displacement distance caused by vertex reduction. The simplification changes road topology (shape, lengths and angles). A set of check points was used in relation to the derived 1:1,000,000 from 1:500,000 data using the simplification operation. Figure 13 indicates the displacement values of the coordinate differences on the derived road map (after generalisation).

![](_page_15_Figure_1.jpeg)

Figure 8: The GPS survey tracks and GCP points superimposed to a simplified 1:250,000 and 1:500,000 scale road database

#### 4. Concluding Remarks

Current map production systems in general provide complex tools and procedural cartographic protocols that still require cartographer interaction. This includes the selection of information, symbolisation of features, maintaining topological relationships among features, and visualisation of graphical conflicts as a result of scale variation, and so on. An efficient generalisation technique aims to improve the graphical quality and legibility of maps. For many cartographic generalisation procedures there are no algorithmic solutions; instead they often use a combination of manual and semi-automatic approaches. Expert systems use collections of 'rules of thumb' that are mainly heuristic criteria, methods, or principles for deciding which among several alternative courses of action might be taken to solve a particular problem. The strain of maintaining upto-date maps at a range of scales for spatial intelligence decision-making is ever increasing. An automated generalisation system meets these requirements for reduced cost and shorter loading time for web mapping applications. Dynamic database generalisation is required for many applications, such as real time transportation

navigation, mobile mapping, emergency management and cartographic map production. A new, semi-automated, spatial data mining and generalisation system was developed for polygon and line related datasets. The tests established that algorithms implemented in the GES are able to extract characteristic vertices of the original entity lines and polylines (e.g. for roads and native vegetation), while excluding non-characteristic vertices to reduce complexity and improve the efficiency of line/polyline generalisation.

This study has demonstrated improvements in reduction. vertex classification and merge, Douglas Enhanced and Peucker, (1973)(Visvalingam and Whyatt, 1991) and Douglas-Peucker-Peschier (in Kreveld and Peschier, 1998) algorithms. The test results confirm that the GES generalises line features well and maintains their geometric relations. The results also compared favorably with existing paper maps (e.g. 1:1,000,000). Existing generalisation software requires advanced technical skills from users, however, the GES has a basic GUI that is an advantage for users with limited technical skills and understanding of spatial data management.

![](_page_15_Picture_8.jpeg)

![](_page_16_Figure_1.jpeg)

(a) The GPS survey points superimposed to a simplified 1:500,000 scale road database

![](_page_16_Figure_3.jpeg)

(b) Assessment of the average shape changes caused by the simplification operation

![](_page_16_Figure_5.jpeg)

![](_page_16_Figure_6.jpeg)

![](_page_16_Figure_7.jpeg)

Changing geographic parameters should be updated in multi-scale maps and spatial databases in "near real time". The conventional methods focus on a single map at a time and therefore result in inconsistency between databases. The GES is a trial tool for generalising large-scale maps into smaller scales, and creating maps of different themes across various scales.

#### Acknowledgements

The authors would like to acknowledge that this research was not possible without sponsorship and assistance of several individuals and organisations. We thank also the anonymous reviewers for their excellent feedback. The paper is extracted and modified from previous publications as referenced in this document.

#### References

- Ai, T., 2019, A Generalization of Contour Line Based on the Extraction and Analysis of Drainage System. Spatial Data Analysis, Statistics and Uncertainty Modelling - Working Groups from International Society for Photogrammetry and Remote Sensing (ISPRS) Commission IV, WG IV/3. Accessed online 8 March 2019: https://www.isprs.org/proceedings-/XXXV/congress/comm4/papers/549.pdf
- ANZLIC, 2001, ANZLIC Metadata Guidelines: Core Metadata Elements for Geographic Data in Australia, Version 2. ANZLIC Spatial Council, Canberra, Australia. Accessed online 25 June 2007: http://anzlic.org.au/asdi/metaelem.htm
- Burrough, P. A., 1986, *Principles of Geographical Information Systems for Resources Assessment*. Clarendon Press, Oxford, UK.
- Budrevicius1, J. D., Papšien, L. and Beconyt, W., 2017, Automatic Generalization of Cartographic Data for Multi-scale Maps Representations. Environmental Engineering 10th International Conference. Vilnius Gediminas Technical University, Lithuania, 27–28 April 2017, 1-6.
- Chaudhry, O. and Mackanessa, W. A., 2008, Automatic Identification of Urban Settlement Boundaries for Multiple Representation Databases. *Computers, Environment and Urban Systems*, Vol. 32, No. 2, 95-109.
- Douglas, D. H. and Peucker, T. K., 1973, Algorithms for the Reduction of the Number of Points Required to Represent a Digitised Line or its Caricature. *Canadian Cartographer*, Vol. 10, No. 2, 112-122.
- Forghani A., 2000, Decision Trees for Mapping of Roads from Aerial Photography Employing a GIS-Guided Technique. *Proceedings of the 10<sup>th</sup>*

Australasian Remote Sensing and Photogrammetry Conference, Adelaide, Australia 21-25 August 2000, 1-12.

- Fei, L. F., 1998, Experiments on the Group Generalization of Contour Lines for Topographic Maps, *Geo-spatial Information Science*, Vol. 1(1), 85-95.
- Forghani, A., 1998, A Knowledge-Based Approach to Mapping Roads from Aerial Imagery Using a GIS Database. PhD Dissertation, Surveying and Spatial Information Science, the University of Tasmania, November 1998, Hobart, Australia, 1-300.
- Forghani, A., 2000, Semi-Automatic Detection and Enhancement of Linear Features to Update GIS Files. *Proceedings of the XIXth ISPRS Congress, July 2000, Amsterdam, Netherlands,* CD-ROM procs, 8.
- 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
- Geoscience Australia (GA), 2001, Helping the Global Community to Resolve Global Problems through Spatial Information. *Proceedings of the AURISA 2001 - the 29th Annual Conference of AURISA, 19-23 November 2001, Melbourne, Australia,* CD-ROM procs, 12.
- Giarrantano, J. and Riley, G., 1989, *Expert Systems: Principles and Programming. PWS-KENT Publishing Company. Boston.*
- Goodchild, M. F. and Hunter, G. J., 1997, A Simple Positional Accuracy Measure for Linear Features. *International Journal of Geographical Information Science*, Vol. 11, No. 3, 299-306.
- Hope, S., Kealy, A. and Hunter, G. J., 2006, Improving Positional Accuracy and Preserving Topology Through Spatial Data Fusion. Proceedings of the 7<sup>th</sup> International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, 5-7 July 2006, Lisbon, Portugal, 675-684.
- Kazemi, S., Lim, S. and Ge, L., 2004a, Is Automated Generalization There Yet? Proceedings of the 12th Australasian Remote Sensing and Photogrammetry Association Conference, Fremantle, Western Australia, 18-22 October 2004. CD-ROM Procs., 1-11.
- Kazemi, S., Lim, S. and Rizos, C., 2004b, A Review of Map and Spatial Database Generalization for Developing a Generalization Framework. Proceedings of the XXth Congress of the International Society for Photogrammetry and Remote Sensing, Istanbul, Turkey, 12-23

July, in DVD Commission 4. CD-ROM procs, 1221-1226.

- Kazemi, S., and Lim, S., 2007, Deriving Multi-Scale GEODATA from TOPO-250K Road Network Data. *Journal of Spatial Science*, Vol. 52(1), 165-176.
- Kazemi, S., 2011, Cartographic Knowledge-Based Generalisation of Spatial Features. PhD Thesis in Surveying & Spatial Information Systems, UNSW, Sydney, Australia, 1-218: http://handle.unsw.edu.au/1959.4/50854.
- Kazemi, S. and Forghani, A., Cho, G. and McQueen, K., 2013, Integrated Cartographic and Algorithmic Approach for Road Network Database Generalisation. *International Journal* of Geoinformatics. Vol. 9, No. 4, 63-76.
- Kazemi, S. and Forghani, A., 2015, Cartographic Knowledge Acquisition: from Theory to Practice. *International Journal of Geoinformatics*, Vol. 11, No. 2, 57-72.
- Kazemi, S. and Forghani, A., 2016, Knowledge-Base Generalisation of Road Networks. *International Journal of Geoinformatics*, Vol. 12, No. 1, 1-13.
- Kreveld, M. J. and Peschier, J., 1998, On the Automated Generalisation of Road Network Maps. Proceedings of the 3<sup>rd</sup> International Conference on GeoComputation, 17-19 September 1998, University of Bristol, United Kingdom. Accessed online 12 July 2006: http://www.geocomputation.org/1998/21/gc\_21. htm
- Lee, D., 2004, Geographic and Cartographic Contexts in Generalisation. *Proceedings of the ICA Workshop on Generalisation and Multiple Representation*, 20-21 August 2004, Leicester, UK. Accessed online 1 March 2006: http://ica.ign.fr/Leicester/paper/Lee-v2-ICAWorkshop.pdf
- McMaster, R. B., 2001, Measurement in Generalization. *Proceedings of the 20th International Cartography Conference, August* 6-10, Beijing, China, 2082-2089.
- Neuffer, D., Hopewell, T. and Woodsford, P., 2004, Integration of Agent-Based Generalisation with Mainstream Technologies and Other System Components. Proceedings of the ICA Workshop on Generalisation and Multiple representation, 20-21 August 2004, Leicester, UK. CD-ROM Procs, 1-11.
- Peschier, J., 1997, Computer Aided Generalisation of Road Network Maps. M.Sc Thesis. Department of Computer Science, Utrech University, mon.nl/thesis.htm; Accessed 25 March 2004.

- Topfer, F. and Pillewizer, W., 1996, The Principles of Selection, A Mean of Cartographic Generalization. *Cartographic Journal*, Vol. 3(1), 10-16.
- Visvalingam, M. and Wyatt, J. D., 1990, The Douglas-Peucker algorithm for Line Simplification Reevaluation through Visualization. *Computer Graphics Forum*, Vol. 9(3), 213-228.
- Warmerdam, F., 1999, The Shapefile C Library Purdue University, USA. Accessed Online 2006: www.engineering.purdue.edu
- Yan, H., 2019, Description Approaches and Automated Generalization Algorithms for Groups of Map Objects. *First Edition 2019*, Kindle Edition, ISBN-13: 978-9811336775.
- Zhang, X., Zhu, X. and Yang, Y., 2009, Employing united Delaunay Triangulation in Contour Lines Generalization. Proceedings of the IEEE - 17th International Conference on Geoinformatics, airfax, VA, USA, 1-9.

![](_page_18_Picture_18.jpeg)