

**MODELLING GEOGRAPHIC PHENOMENA AT  
MULTIPLE LEVELS OF DETAIL: A MODEL  
GENERALISATION APPROACH  
BASED ON AGGREGATION**

**OMAIR Z CHAUDHRY**

DOCTOR OF PHILOSOPHY  
THE UNIVERSITY OF EDINBURGH

2007

## **Declaration of Originality**

This thesis has been composed by me and contains work that is entirely my own except where stated otherwise. This work contains no material that has been accepted for the award of any other degree or diploma in any university or tertiary institution.

Omar Z Chaudhry

26<sup>th</sup> October 2007

## **Abstract**

Considerable interest remains in capturing once geographical information at the fine scale, and from this, automatically deriving information at various levels of detail and scale via the process of map generalisation. This research aims to develop a methodology for transformation of geographic phenomena at a high level of detail directly into geographic phenomena at higher levels of abstraction. Intuitive and meaningful interpretation of geographical phenomena requires their representation at multiple levels of detail. This is due to the scale dependent nature of their properties. Prior to the cartographic portrayal of that information, model generalisation is required in order to derive higher order phenomena typically associated with the smaller scales. This research presents a model generalisation approach able to support the derivation of phenomena typically present at 1:250,000 scale mapping, directly from a large scale topographic database (1:1250/1:2500/1:10,000). Such a transformation involves creation of higher order or composite objects, such as settlement, forest, hills and ranges, from lower order or component objects, such as buildings, trees, streets, and vegetation, in the source database. In order to perform this transformation it is important to model the meaning and relationships among source database objects rather than to consider the object in terms of their geometric primitives (points, lines and polygons). This research focuses on two types of relationships: taxonomic and partonomic. These relationships provide different but complimentary strategies for transformation of source database objects into required target database objects. The proposed methodology highlights the importance of partonomic relations for transformation of spatial databases over large changes in levels of detail. The proposed approach involves identification of these relationships and then utilising these relationships to create higher order objects. The utility of the results obtained, via the implementation of the proposed methodology, is demonstrated using spatial analysis techniques and the creation of 'links' between objects at different representations needed for multiple representation databases. The output database can then act as input to cartographic generalisation in order to create maps (digital or paper). The results are evaluated using manually generalised datasets.

## **Acknowledgements**

This thesis would not be a reality without the support of numerous people and institutions. I think the words *thank you* are not enough to express how grateful I am for their needed support but this is all I have.

Firstly, I would like to thank William Mackaness from the bottom of my heart for introducing me to this exciting field of map generalisation, for his guidance throughout, his constant input and much needed support, as well as for his friendship. He is a true inspiration and a great person to work with.

I am grateful to Nicholas Hulton and Nicolas Regnaud for their supervision, comments and suggestions throughout the research and also for reviewing the thesis. Thanks to Patrick Revell for his suggestions and quick replies to all my questions and to Ross Purves for his comments and suggestions on our hills and range boundary detection methodology.

I would also like to thank, my office mates, Karin Viergever, Nick Cutler, Iain Cameron and Sebastien Nobert for providing a conducive and friendly working environment. Also thanks to all my friends in Edinburgh for making the last three years full of pleasant memories.

I am grateful to Ordnance Survey for funding this research, provision of data, and for providing the opportunity to work with the generalisation group at Ordnance Survey for three months. Special thanks to the University of Edinburgh for funding this research and for providing all the required resources and facilities.

I am grateful to my parents, my sisters for all their love, support and their special prayers for my success. Also to my wife/life, Mehreen, for always being there and for being an enduring source of happiness. This research would not have been possible without her support and love.

Above all, I am thankful to The Almighty for always showing me the right path and helping me in every difficulty.

## Contents

Declaration of Originality .....	i
Abstract .....	ii
Acknowledgements .....	iii
Contents .....	iv
List of Figures .....	viii
List of Tables.....	xvi
<b>CHAPTER 1: Introduction.....</b>	<b>1</b>
1.1 Scale and Geographic Phenomena .....	1
1.2 The need for ‘Automatic’ Generalisation .....	2
1.3 Problem definition.....	5
1.4 Aim of Research.....	6
1.5 Key Objectives .....	8
1.6 Principle Rationale .....	8
1.7 Structure of Thesis .....	9
<b>CHAPTER 2: Relevant Concepts of Spatial Data .....</b>	<b>11</b>
2.1 Spatial Data Modelling .....	11
2.2 Field and Object Structure Approach.....	12
2.3 GeoSpatial Ontologies .....	14
2.4 Classification of Objects .....	16
2.5 Relationships .....	17
2.5.1 Topological Relationships.....	17
2.5.2 Taxonomic Relationships.....	19
2.5.3 Partonomic Relationships .....	20
2.5.4 Taxonomy and Partonomy .....	24
2.6 Generalisation of Spatial Databases.....	26
<b>CHAPTER 3: Generalisation in Digital Domain: Focus on Database Transformation .....</b>	<b>27</b>
3.1 Objectives of Generalisation .....	27
3.2 Model generalisation and Cartographic Generalisation.....	28
3.3 Effects of Model and Cartographic Generalisation.....	30
3.4 Multiple Representation Database .....	31

3.4.1	Creation of MRDBs .....	34
3.5	Critical Factors of Model Generalisation.....	37
3.5.1	Data Modelling.....	37
3.5.2	Objects and Relations.....	37
3.5.3	Conditions and Constraints .....	38
3.5.4	Operations .....	38
3.6	Strategies of Model Generalisation (aggregation methods).....	41
3.6.1	Geometry Driven Generalisation .....	42
3.6.2	Structural Generalisation.....	42
3.6.3	Class Driven Generalisation.....	43
3.6.4	Functionality (or Partonomy) Driven Generalisation .....	44
3.7	Classes of the target Data Model .....	46
3.8	Populating the Target classes .....	47
3.8.1	Pre-processing: Classification of Source Database.....	48
3.8.2	Database Enrichment .....	49
3.8.3	Selection and Aggregation .....	50
3.9	Summary .....	50
<b>CHAPTER 4: Specification and Description of the Source Datasets.....</b>		<b>52</b>
4.1	Datasets .....	52
4.1.1	Primary Dataset: OS MasterMap .....	52
4.1.2	Additional Datasets .....	56
4.2	Pre-processing: Classification.....	58
4.3	Summary .....	61
<b>CHAPTER 5: Identification of ‘Container’ Boundaries.....</b>		<b>62</b>
5.1	Modelling Boundaries .....	62
5.2	Methodology for Settlement Container Boundary.....	63
5.2.1	A Distance based Clustering Approach .....	65
5.2.2	Calculation of Citiness .....	67
5.2.3	Creation of the Boundary of the Settlement Container.....	69
5.2.4	Implementation and Results.....	71
5.3	Methodology for Forest Container Boundaries.....	73
5.3.1	Area Patch Methodology.....	75

5.3.2	Extension to Area Patch Methodology .....	77
5.3.3	Implementation and Results.....	79
5.4	Methodology for Hill and Range Container Boundaries .....	81
5.4.1	Calculating the prominence of a Summit.....	83
5.4.2	Modelling morphological variance .....	87
5.4.3	Calculating the extent of each summit.....	89
5.4.4	Classification into Hills and Ranges .....	91
5.4.5	Implementation and Results.....	93
5.5	Summary .....	95
<b>CHAPTER 6: Database Generalisation via Aggregation based on Partonomies 97</b>		
6.1	Making Partonomies Explicit.....	97
6.1.1	Modelling a palimpsest of partonomies .....	99
6.2	Database Transformation .....	100
6.2.1	Selection and Aggregation of Road and Rail class objects.....	101
6.2.2	Aggregation via Partonomy .....	102
6.2.3	Selection and Aggregation via specific rules.....	104
6.3	Implementation .....	107
6.4	Summary .....	109
<b>CHAPTER 7: Results and Discussions ..... 110</b>		
7.1	Case Studies .....	110
7.2	Evaluation .....	117
7.2.1	Visual Comparison.....	117
7.2.2	Using text points from cartographic dataset .....	125
7.2.3	Quantitative Evaluation.....	130
7.3	Utility of Results .....	133
7.3.1	Spatial analysis.....	134
7.3.2	Identification and Enrichment of new Composite Classes .....	137
7.3.3	Creation of links for MRDBs.....	138
7.4	Generalisation errors due to inconsistencies in the source database.....	139
7.5	Degree of Fuzziness in Results .....	140
7.6	Further Discussion .....	141
7.7	Summary .....	143

<b>CHAPTER 8: Conclusions and Future Work.....</b>	<b>145</b>
8.1 Summary of Thesis .....	145
8.2 Major Achievements .....	146
8.3 Future work .....	149
<b>References .....</b>	<b>153</b>
<b>Preface.....</b>	<b>166</b>
Appendix I.....	168
Appendix II .....	179
Appendix III.....	197
Appendix IV.....	224
Appendix V .....	242
<b>Curriculum Vitae .....</b>	<b>265</b>



## List of Figures

Figure 1.1: Different phenomena represented at different levels of detail (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved). Houses visible at 1:25,000 scale, blocks visible at 1:50,000 and settlements at 1:250,000 .....	3
Figure 1.2: Incremental vs Direct generalisation of spatial databases .....	7
Figure 1.3: What is perceived is governed by scale (and theme). (a) Non similar objects at 1:25,000 have been aggregated to create a object at 1:250,000 (b) (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved).....	7
Figure 2.1: Two principle approaches for structuring spatial data (a) field approach (b) object structured approach (Molenaar 1998).....	13
Figure 2.2: Topological relationships between polygon objects with no disconnected parts .....	19
Figure 2.3: A possible classification hierarchy for transportation. Different levels of hierarchy reflect different levels of detail (modified after Peng (1997)).....	20
Figure 2.4: Composite object consisting of two component object(van Smaalen, 2003) .....	22
Figure 2.5: An example of an aggregation hierarchy of a city in terms of its parts (modified after Peng (1997)).....	24
Figure 3.1: The first abstraction of reality creates the primary digital landscape model (DLM), from which a digital cartographic model (DCM) can be produced – either directly from the DLM or via the process of model generalisation and the creation of secondary landscape models (after Grünreich (1985)).....	29
Figure 3.2: Effects of cartographic generalisation. a) Metaphorical representation of the database b) Cartographic representation of the database results in conflict between symbols c) Displacement is applied to remove the conflict (modified after (Sarjakoski, 2007)).....	31
Figure 3.3: The relationship between MRDB , model and cartographic generalisation .....	32

Figure 3.4: Representation levels for a building object in a MRDB (modified from Kilpelainen, 1997).....	33
Figure 3.5: Matching settlement objects from Strategi cartographic dataset overlaid on building objects in OS MasterMap Topography layer (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved). A few building objects fall outside the settlement boundary due to cartographic operations such as displacement performed on the Strategi dataset .....	36
Figure 3.6: Example of selection operator. (a) A graph theoretic representation of road network at 1:50,000 (b) has been generalised to network at a lower level of detail (1:100,000) (c) whilst avoiding becoming disconnected (scanned from Mackaness and Beard (1993)).....	40
Figure 3.7: Aggregation of individual objects into a composite object. The geometry of the composite object is created by merging the faces (polygons) of individual objects.....	41
Figure 3.8: Example of geometry driven generalisation. Elimination of preliminary scrub and farm shown in a) that are below the area threshold creates holes (b). To avoid this, the scrub object is aggregated with the tree object and the small farm objects have been aggregated to create a single farm object (c) .....	42
Figure 3.9: (a) Example of a Classification Hierarchy or taxonomy. (b) Class driven generalisation based on the taxonomy .....	43
Figure 3.10: Aggregation hierarchy and aggregation of objects based on their partonomic relationships. Here objects belonging to different taxonomies have been aggregated to create objects (block, settlement) that are functionally (or partonomically) related instead of taxonomically .....	45
Figure 3.11: Overall methodology .....	49
Figure 4.1: Extract of OS MasterMap Topography layer (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved) .....	54
Figure 4.2: Polygon features form a complete geometric partition i.e. there is no overlap between features.....	56

Figure 4.3: Extract of ITN data segments are classified based on their descriptive term (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved).....	57
Figure 4.4: Link between ITN roads and road objects in the source database. The descriptive term of ITN roads is used for classification of corresponding road objects in the source database. The objects in the two datasets are linked via cross reference table that contains object identifier (Toids) of both datasets .....	60
Figure 4.5: Classification of road objects in the source database shown in Figure 4.4 based on the descriptive term values of corresponding ITN roads and classification rules .....	61
Figure 5.1: The major stages of methodology used for determination of settlement container boundaries .....	65
Figure 5.2: Typical objects (buildings) selected from the source database. Three low dense areas are ringed (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved) .....	66
Figure 5.3: The output boundaries generated by distance based clustering algorithm using a 100m threshold. The low dense areas ringed in Figure 5.2 have become part of the output boundaries .....	67
Figure 5.4: Interpolated and normalised surface based on citiness values (using equation 1) for buildings in Figure 5.2. The three prominent settlements are ‘Livingston’, ‘Mid Calder’ and ‘East Calder’ in Scotland.....	68
Figure 5.5: Process of expansion, aggregation and elimination for three examples (a). (b) expansion of objects propotional to their citiness value. (c) Aggregation of overlapping objects. (d) Elimination of small boundary objects and holes .....	70
Figure 5.6: Resultant settlement container boundaries generated by the algorithm superimposed on the input buildings (Figure 5.2) .....	71
Figure 5.7: Resultant boundary generated by the algorithm with input buildings selected from the source database. The region is in the Peak District in England .....	72

Figure 5.8: Resultant boundary generated by the algorithm with input buildings selected from the source database. The region is around the town of Peebles in Scotland .....	72
Figure 5.9: Resultant boundary generated by the algorithm with input buildings selected from the source database. The input dataset consists of approximately 130,000 buildings. The region is Edinburgh city and its surroundings in Scotland.....	73
Figure 5.10: Methodology for creation of forest container boundaries .....	74
Figure 5.11: Main stages of area patch methodology .....	75
Figure 5.12: Application of Müller and Wang (1992) area patch methodology on tree objects (a). Result of expansion and contraction (b). Resultant forest boundary object (c) .....	76
Figure 5.13: Forest container boundary (dark grey) overlaid over original tree objects (light grey) selected from source database. The region is south of the town of ‘East Calder’ in Scotland .....	79
Figure 5.14: Output forest container boundaries (dark grey) overlaid on source database tree objects (light grey). The region is Livingston and its surrounding in Scotland .....	80
Figure 5.15: Output forest container boundaries (dark grey) overlaid on source database tree objects (light grey). The region is Peebles and its surroundings in Scotland.....	81
Figure 5.16: The overall method by which hills and ranges are identified.....	83
Figure 5.17: (a) Source DTM (b) Contours at 5m interval from source DTM. Contours have ‘spikes’ and are broken at certain places .....	85
Figure 5.18: (a) Smooth or generalised DTM (b) Resultant contours at 5m interval including height labels .....	85
Figure 5.19: Contours created from a smooth DTM (Figure 5.18a). Summit points along with their elevation values are shown within each of the highest contours. Key contours of summit A and B are highlighted in bold. Note that all the summit points that are inside the key contour ‘a’ of summit A are of lower elevation than summit A (232m).....	86

Figure 5.20: Approximate profile of transect from Figure 5.19 showing the Prominence of summit A and B .....	86
Figure 5.21: Multiscale morphometric classification of DTM in Figure 5.19. The kernel used ranges from 3*3 to 51*51 .....	88
Figure 5.22: Morphologically variable polygons of morphometric features as depicted in Figure 5.21 .....	89
Figure 5.23: Determining the extent of a summit A.(a): Key contour A, morphologically variable polygons (b) next higher level contour is selected (c) next higher level contour is selected (d) the resultant extent of summit A.....	90
Figure 5.24: (a) Summits and their extents identified in Figure 5.23a (b) the summits and extents of Figure 5.19 .....	91
Figure 5.25: Summit A is a range since its extent contains summit B which on the other hand is a hill. Similarly summit C is a hill because it does not contain any summit in its extent.....	92
Figure 5.26: Hill and range boundaries above threshold of 35m. The summit B (prominence 19m) in Figure 5.19 has been aggregated into its Peak A (prominence 156m). .....	92
Figure 5.27: Input DTM for Edinburgh and surrounding in Scotland .....	93
Figure 5.28: Resultant hills and range container boundaries for Figure 5.27 above 35m prominence.....	94
Figure 5.29: Input DTM for region surrounding Peebles in Scotland .....	94
Figure 5.30: Output container boundaries of hills and ranges for Figure 5.29 above 35m prominence .....	95
Figure 6.1: Container boundaries (forest and settlement) using the approach proposed in Chapter 5 overlaid on source database objects. The container boundaries are labelled with their object identifier. A portion of overlapping area has been highlighted in red and is shown in Figure 6.2 .....	98
Figure 6.2: Partonomic relationships for a few source database objects in the region highlighted in Figure 6.1. The rows highlighted in the table are those objects having multiple partonomies .....	99

Figure 6.3: Using multiple partonomic information to support multi criteria spatial analysis. For instance finding all objects that are part of leafy-suburban and hilly region.....	100
Figure 6.4: Aggregation of road segments based on adjacency and being of the same class.....	102
Figure 6.5: Overlapping composite objects due to m:n partonomic relationships for the region highlighted in Figure 6.2. The forest (light green) and settlement object (light grey) overlap because of aggregation based on multiple partonomies without any constraints .....	103
Figure 6.6: Aggregation stages for the creation of (non overlapping) composite objects from component objects in the source database .....	105
Figure 6.7: (a) Figure 6.1 (b) Resultant objects created by the proposed aggregation approach. The resultant objects are thematically and geometrically partitioned .....	107
Figure 7.1: Source Database objects classified according to the pre-processing stage of the proposed methodology (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved). The region is Livingston in Scotland.....	111
Figure 7.2: Settlement and forest container boundaries for the input region shown in Figure 7.1 .....	111
Figure 7.3: Objects of the resultant database classified according to the target data model .....	112
Figure 7.4: Source Database objects classified according to the pre-processing stage of the proposed methodology (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved). The region is Peebles in Scotland.....	113
Figure 7.5: Settlement, forest, hills and range container boundaries for the input region shown in Figure 7.4 .....	113
Figure 7.6: Objects of the resultant database classified according to the target data model .....	114

Figure 7.7: Source Database objects classified according to the pre-processing stage of the proposed methodology (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved). The region is the city of Edinburgh and its surroundings in Scotland .....	115
Figure 7.8: Settlement, forest, hills and range container boundaries for the input region shown in Figure 7.7 .....	116
Figure 7.9: Objects of the resultant database classified according to the target data model .....	116
Figure 7.10: Collection of lake patches. Taken from Bertin (1967) scanned from Müller & Wang (1992). The region is Les Dombes, located North East from Lyon (France) .....	118
Figure 7.11: Various generalised solutions (at same scale) for data shown in Figure 7.10, from Bertin (1967) (scanned from Müller & Wang (1992)).....	118
Figure 7.12: Settlement, forest, road and railways objects selected from the resultant database (see Figure 7.3 for comparison). A few differences with Figure 7.13 are ringed .....	120
Figure 7.13: Strategi dataset illustrating settlement, forest, rail and roads in the region shown in Figure 7.1 (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved) .....	120
Figure 7.14: Settlement, forest and roads objects selected from results shown in Figure 7.6 for comparison against Figure 7.15 .....	122
Figure 7.15: Strategi dataset illustrating settlement, forest, rail and roads in the region shown in Figure 7.4 (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved) .....	123
Figure 7.16: Settlement, forest, road and railways objects selected from resultant database shown in Figure 7.9 for comparison.....	124
Figure 7.17: Strategi dataset illustrating settlement, forest, rail and roads in the region shown in Figure 7.7 (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved) .....	125

Figure 7.18: Differences between road objects in the resultant database (in grey polygons) and in Strategi (blank polylines). Junctions are not present as a separate object in Strategi, dual carriage ways have been collapsed into single road polylines in Strategi in contrast to objects in the results .....	132
Figure 7.19: Example of spatial analysis on the enriched source database. (a) Resultant database. (b) All buildings in East Calder selected in source database (in yellow). (c) All objects that are part of East Calder with degree of partonomy above or equal to minimum degree of partonomy threshold (DT) used during aggregation stage (section 6.2.3).....	135
Figure 7.20: Example of spatial analysis. Finding the shortest road path (highlighted in red) between 'Livingston' and 'East Calder'.....	137
Figure 7.21: Creation of instance of new composite classes that are combinations of existing composite classes. The figure illustrates the creation of hill_forest (composite class) object via the aggregation of objects that are part of a hill as well as a forest.....	138
Figure 7.22: (a) Railway objects (black polygons) in the source database are disconnected (b) The resultant rail objects are thus disconnected in the target database.....	140
Figure 7.23: What does 'London' mean?- A phenomenological view much closer to our concept of London .....	143



## List of Tables

Table 3.1: Class of the target data model .....	47
Table 4.1: Important attributes of the Topography Layer of OS MasterMap used in this research.....	55
Table 4.2: Classes of the source data model .....	59
Table 7.1: Settlement points distances from the closest settlement objects in Figure 7.3 .....	126
Table 7.2: Text points distance from the closest settlement, forest, hill and range objects in Figure 7.6.....	126
Table 7.3: Text points distance from the closest settlement, hill and range objects in Figure 7.9 .....	128
Table 7.4: Total number of objects in the source database, number of objects per class in the resultant database and in the Strategi for the case study shown in Figure 7.1 .....	130
Table 7.5: Total number of objects in the source database, number of objects per class in the resultant database and in the Strategi for the case study shown in Figure 7.4.....	130
Table 7.6: Total number of objects in the source database, number of objects per class in the resultant database and in the Strategi for the case study shown in Figure 7.7.....	131
Table 7.7: Comparison of cumulative areas for settlement and forest class objects in the results and strategi for the three case studies .....	133

## **CHAPTER 1: Introduction**

### **1.1 Scale and Geographic Phenomena**

‘All geographic processes are imbued with issues of scale’ (Taylor, 2004 p.214). Scale has a huge effect on the determination of what phenomena can be viewed, what patterns can be portrayed and what processes can be inferred. The scale of observation is critical to the discernment of pattern and the identification of various types of relationships implicit among the representation of a set of geographic phenomena (Ruas & Mackaness, 1997). Scale is not only important to geographers but also to disciplines such as ecology, meteorology, climatology, geology, economics, sociology and political science. It is indeed quite difficult to identify a completely “scale-less” discipline (Sheppard & McMaster, 2004). There is a strong relationship between phenomena, process and scale of observation. As scale changes, different processes and phenomena become apparent and different patterns emerge (Mackaness, 2007). Typically we are interested in viewing both the precise detail of the phenomena represented as well as the broad linkages across regional and global spaces.

Viewing and analysing geographic space at various levels of detail is common practice in the geosciences (Leitner, 2004). The activity helps to discern the operational scales of geographic phenomena, the extent and permanence of patterns, which in turn sheds light on the underlying processes and their interactions (Monmonier, 1991). There is no single scale at which to view the world. Different scales enable representation of different information which is useful for different applications. Spatial data portrayed at multiple scales in map form has existed for thousands of years (Turnbull, 1989). Cartographers have long understood the link between scale and task. Monmonier (1984) argues that it is a travesty not to supply mapping at multiple scales. It is not the case that any one scale contains more or less information, more that they contain different, albeit related information (Mackaness, 2007). For instance for pedestrian navigation within a city we require spatial data at large scale (showing detail) since it contains information at the street level; for

navigation between or across cities a coarser view is required for the purposes of planning and to gain a better sense of overall distance and direction. Thus there is a need to represent spatial data at different levels of details to discern fundamentally different processes and patterns.

## **1.2 The need for 'Automatic' Generalisation**

To fulfil the requirement for spatial data at different scales, National Mapping Agencies (NMAs) maintain and provide maps at a variety of scales or levels of detail (Figure 1.1). As an example, if a user wished to view individual houses then a 1:25,000 scale topographic map might be appropriate. A more generalised 'block view' is available at 1:50 000 scale. In order to see an entire urban settlement, it would be necessary to view this information at smaller scales (say 1:250,000 scale) (Figure 1.1). Traditionally it has been the job of the cartographer to decide how the information is best filtered in order to be visualised at a smaller scale. The cartographer was responsible for selecting and symbolising the data critical for the intended task. This process of filtering data from large scale to smaller scale is called 'Map Generalisation'. The goal of map generalisation is to give emphasis to salient objects and their properties whilst omitting less important qualities with respect to the scale and the purpose of a map (McMaster & Shea, 1992).

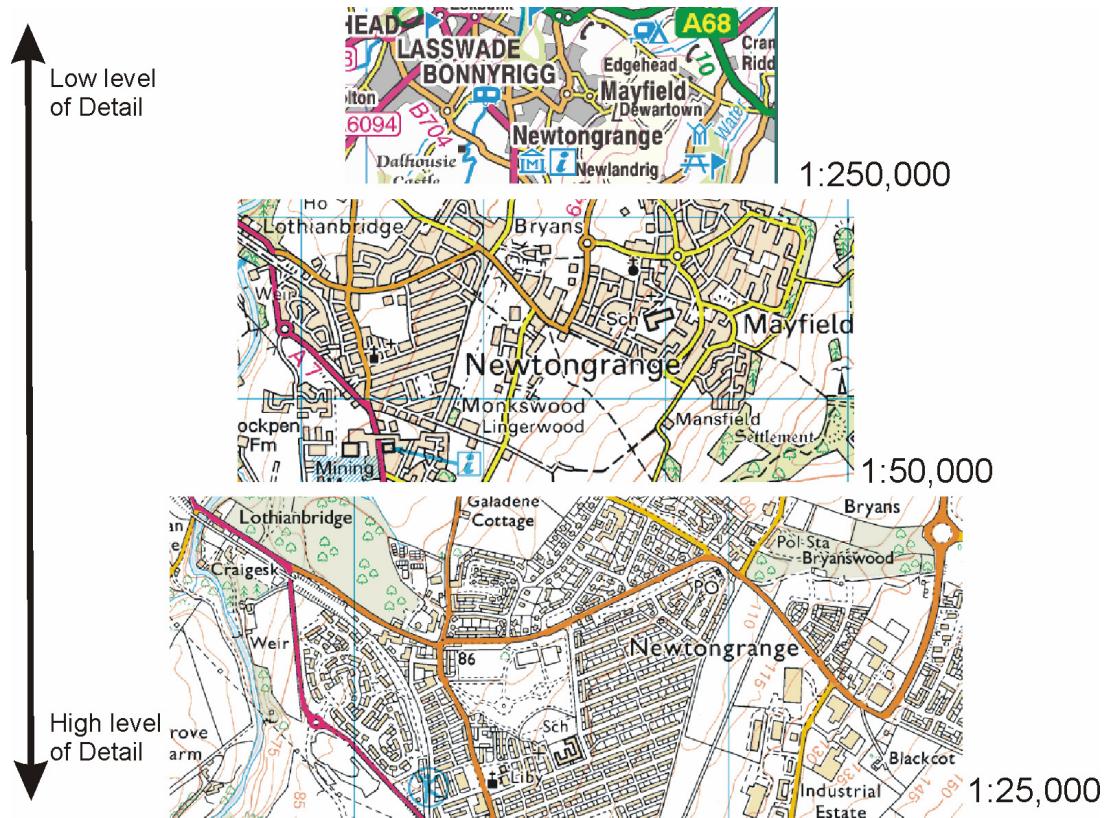


Figure 1.1: Different phenomena represented at different levels of detail (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved). Houses visible at 1:25,000 scale, blocks visible at 1:50,000 and settlements at 1:250,000

Manual map generalisation is a slow, labour intensive and expensive process. The vision is that computers could support computer aided cartography which would overcome these shortfalls and provide more reliable, easy to implement solutions – resulting in significant cost savings. Information technology has not devalued the power of the map, but it has driven a series of paradigm shifts in how we store, represent and interact with geographical information (GI) and has offered new ways of visualising and exploring GI. Historically the paper map reflected our current knowledge of the geography of the world. Now it is the database that has become the knowledge store, with the map as the metaphorical window by which geographic information is dynamically explored (Mackness, 2007). In these interactive environments, the art and science of cartography is being extended to support the integration of distributed data collected at varying levels of detail, whilst conforming to issues of data quality and interoperability. With respect to map generalisation, the challenge is in developing a set of algorithms and methodologies that mirror the

service traditionally provided by the human cartographer, yet take advantage of the paradigm shift afforded by information science in how we interact with, and explore geographic information (Mackness & Chaudhry, 2007b).

Within this paradigm shift, the requirement to view the world at different scales (or multiple levels of detail<sup>1</sup>) has remained, as has the requirement to produce high quality cartographic products. The transition from paper to digital mapping initially involved digitising paper maps at different scales and storing them (and maintaining them) in separate databases (very much reflecting traditional paper map production techniques within National Mapping Agencies). However there is huge redundancy in this model – changes in the real world had to be reflected in changes in each of the independent databases. But a line of thinking has emerged which has challenged the wisdom of this approach, asking whether it is possible to store the phenomenon once (at a very high level of detail), and then to apply a range of algorithms in order to control the selection and representation of the phenomenon in a form appropriate to the intended scale. There are significant benefits to this line of thinking: such as cost effectiveness, database consistency, ease of updating process and integration with third part data sets. These benefits are reflected in developments into Multiple

---

<sup>1</sup> In terms of spatial databases term ‘level of detail’ is more appropriate

Representation Database i.e. databases in which different representations are linked (discussed in Chapter 3) (Mustière & van Smaalen, 2007; Sarjakoski, 2007).

These benefits are premised on the existence of a set of algorithms that can, with minimum intervention from the user, control the selection and representation of geographic phenomenon according to a specified scale and theme. The science (or automated art) of ‘map generalisation’ is all about designing such algorithms; algorithms that manipulate and symbolise the geometric primitives stored in the database. Map generalisation can also be viewed as a service; in the anticipation of users unfamiliar with cartographic concepts, and with poor evaluation skills (Mackaness & Chaudhry, 2007b). In other words, a level of automation is required that anticipates users who are unaware of generalisation procedures and do not have the necessary cartographic training to perform manual generalisation.

### **1.3 Problem definition**

Automating the process of map generalisation has been a research field for more than three decades. But still there are very few complete commercial solutions. The question is why a task which is performed reasonably ‘easily’ by humans has become such a difficult problem in the digital domain (Sheppard & McMaster, 2004). There are several reasons for this such as the challenge of formalising rules, understanding of the link between scale and phenomena, assessment of results, determination of appropriate parameter values and constraints that control the automatic decision making process. There has been a failure to recognise generalisation as a modelling process. Rather it is seen as some ‘drawing’ process at the end of analysis (Mackaness, 2007). Most focus in generalisation research has been towards the development of algorithms that improve the graphic or visual quality of the output map. But with the increased use of geographic information systems (GIS) and developments in spatial databases, generalisation is now a part of the framework of geographic information processing. The graphic output (paper or digital map) now acts as a window by which we can search and explore the underlying database. “Generalisation must be based on process rather than graphical appearance” (Müller, 1989 p.203). “By dropping this (asthetic) constraint it should simplify matters

significantly and get to the core of the problem more easily” (van Smaalen, 2003 p.2). The first part of automatic generalisation is the transformation of geographic phenomena in the source database into the required phenomena. The appropriate graphic display of the resultant phenomena then becomes a secondary objective. This research takes a modelling perspective of map generalisation and focuses on the database side of the problem. The next sections present the aim and objectives of this research and the justification for doing this research.

#### **1.4 Aim of Research**

The aim of this research is the automatic derivation of spatial objects typically present at a notional 1:250,000 scale database directly from a large scale spatial database (in this case, OS Master Map) via the automated process of (model) generalisation.

Transformation of a spatial database over large changes in scale is not straight forward and is more than just a process of *subselecting* the data. The transformation of the database involves creation of higher order concepts such as cities, forest regions, and mountain ranges from source concepts objects in the source database (such as buildings, trees and groups of hills). In this transformation process it is important to model phenomena in a meaningful way (Ormsby & Mackaness, 1999), rather than to consider the object in terms of its geometric primitives (points, lines and polygons) (Nyerges, 1991). It requires the modelling of relationships, both thematic and spatial, of objects in the database (Ruas & Mackaness, 1997). These relationships illustrate the role of objects in the database, their association with each other and also their link with the required concepts. Modelling of these relationships is thus essential for transformation of the database.

Minsky (1975) made the observation: that you cannot tell you are on an island by looking at the pebbles on a beach. At one scale you see the pebble, at another scale you see a continuous coastline, yet there is an interdependence between and among geographic phenomenon that leads us to believe that objects at small scale can be automatically derived from data stored at fine scale. This is an assumption that pervades the science of map generalisation. We can either derive databases at lower

levels of detail using a ‘cascade’ approach (Figure 1.2a) or we can ‘jump’ straight from fine scale to the scale required (Figure 1.2b). This research takes a direct approach for the transformation of a source database into a target database. Mostly the research in generalisation has followed a stepwise approach (Figure 1.2a). This is because of the graphic requirement to an output with close resemblance to the input. But over large changes in the level of detail or scale there is fundamental changes in content and the incremental approach is not appropriate. In making these ‘jumps’ (Figure 1.2b) we need to cross what Müller (1991) referred to as ‘conceptual cusps’. These cusps exist in the generalisation continuum – points at which representations fundamentally change. An obvious example would be the point at which a collection of separate objects typically associated with ‘town’ (such as dense buildings and network structures), are aggregated, and wholesale replaced with a single feature (Figure 1.3) or the point at which we no longer perceive the ‘pebble’ but ‘see’ the beach (Mackness, 2006).

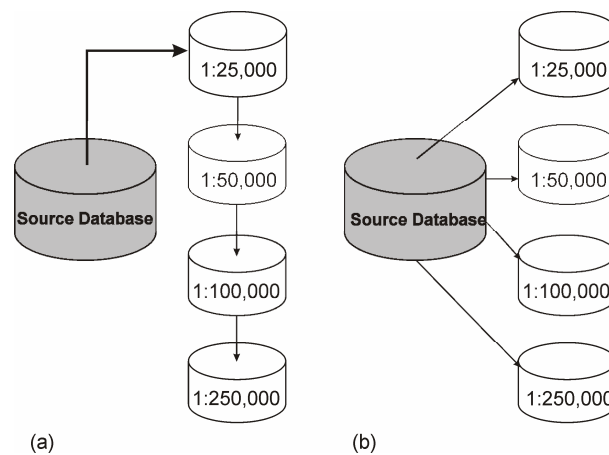


Figure 1.2: Incremental vs Direct generalisation of spatial databases

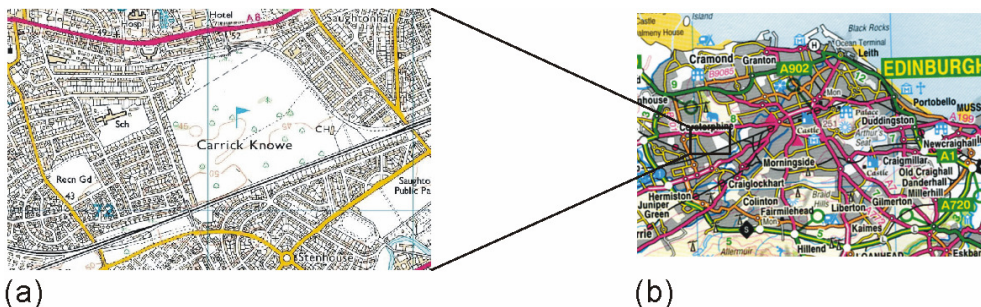


Figure 1.3: What is perceived is governed by scale (and theme). (a) Non similar objects at 1:25,000 have been aggregated to create a object at 1:250,000 (b) (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved)



## 1.5 Key Objectives

The aim of this research will be achieved via the following key objectives:

- To demonstrate the feasibility of *direct* transformation (Figure 1.2b) of source database objects into required objects at lower level of detail (1:250,000);
- To illustrate the importance of relationships especially *partonomic* relationships in terms of database transformation and as a way of linking geographic phenomena at different levels of detail;
- To demonstrate the utility of the transformed database in terms of spatial analysis and as a prerequisite to cartographic generalisation;
- To implement and evaluate the model as a basis for demonstrating proof of concept, addressing pragmatic issues in the handling of large volumes of data;

The methodology utilises the Ordnance Survey (OS) topographic dataset (OS MasterMap Topography Layer and ITN layer) along with OS digital terrain model dataset (Land-Form PROFILE Plus). OS's cartographic product Strategi (1:250,000) was used for evaluation and determination of various thresholds. The implementation was done using a relational database, open source code and programming in Java and SQL.

## 1.6 Principle Rationale

The research will help in assessing the viability of systems that are capable of deriving small scale databases directly from large scale data (thus obviating the need for multiple datasets); it will advance techniques in the modelling of geographic phenomenon, in particular our understanding of how information is 'transformed' across scales; increase our understanding of the aggregation process among groups of classes of objects; and result in a better understanding of the links between the form of query and the appropriate level of detail required to support multi scale spatial analysis. The research will demonstrate the close linkage that exists between query

and scale, arguing for a deeper understanding of ideas of aggregation and generalisation that extends beyond the visual view, and includes ideas of database abstraction. The research will demonstrate the utility of the transformed database in terms of spatial analysis routines that could not be applied in the source database without this transformation and will also highlight the importance of the resultant database in terms of linking source and target objects and also as input to cartographic generalisation. The novelty of this research lies in:

- **A Phenomenological Perspective:** Exploring how geographic phenomena merge or separate to create higher order, more generalized forms;
- **Aggregation:** Development of aggregation techniques based on a combination of spatial and thematic relationships;
- **Large Scale Changes:** Developing solutions to perform database transformation (model generalisation) over a large scale change in a single step.

## 1.7 Structure of Thesis

The thesis is divided into eight chapters. An introductory part (Chapter 1 to 3) which is followed by the main stages of the methodology research section (Chapter 4 to 7). The last chapter (8) presents conclusions of the research and presents an outlook on further research. In summary:

**Chapter 2** introduces different spatial concepts, spatial databases, data modelling, classification, relationships, that are important for understanding of the proposed methodology.

**Chapter 3** builds on the introduction of Chapter 1 and the concepts of Chapter 2 and defines the objectives of generalisation. These objectives define the two categories, model and cartographic generalisation. Model generalisation is discussed in more detail since this is the core approach to database transformation. The chapter presents critical factors of model generalisation, different strategies and highlights the importance of a partonomic or functional approach. The chapter ends with the

introduction of different stages of the proposed methodology for database transformation.

**Chapter 4** discusses the input datasets used. Their particular properties are presented. The chapter then presents the classification of the source dataset (the first stage of the methodology).

**Chapter 5** presents algorithms developed for the detection of ‘boundaries’ of objects in the output classes. These boundaries are required for determination of (partonomic) relationships for source objects. It presents three techniques; one for each (composite) class of the target data model. Each approach is explained with the aid of case studies.

**Chapter 6:** This chapter presents methods for determining the partonomic relationships using the boundaries detected in Chapter 5. The chapter then discusses the issue of multiple partonomies and how these relationships along with taxonomic relationships are used for database transformation. Aggregation rules are presented for this transformation.

**Chapter 7** presents results for three different regions of interest selected from the source database. Different techniques are presented for evaluation of these results. The chapter also discusses the importance of the results in terms of spatial analysis and links for MRDB.

**Chapter 8** concludes the thesis by giving a summary of the thesis and presenting the major achievements. The chapter also discusses the possible avenues of future research.

Throughout this three year research period, outputs have been presented at international conferences and workshops. These are summarised in the author’s curriculum vitae attached at the end of the thesis. We have submitted five papers for publication (all five accepted). Copies of these papers are included in the Appendices section (Appendix I-V).

## **CHAPTER 2: Relevant Concepts of Spatial Data**

This chapter defines the foundation of this research by describing the relevant aspects of spatial data in order to fully convey the process of generalisation. The aim is not to carry out a review of general concepts of spatial data, but to discuss those concepts that are relevant to the understanding of the proposed methodology.

### **2.1 Spatial Data Modelling**

This thesis concerns itself with automated map generalisation and to properly understand how things can be approached it is important to fully appreciate the building blocks that determine how real-world entities are represented in the computer environment. Thus the way in which generalisation can be facilitated is itself influenced by the way in which the information describing the location and nature of spatial entities is represented by computers.

In the cartographic literature there is no clear distinction between a dataset and database. In this thesis the term dataset refers to spatial data captured by a surveyor or by photogrammetric techniques at a certain scale or level of detail. The term 'database' or 'spatial database' here refers to a spatial dataset together with a database management system. A spatial database like any ordinary database includes standard data types such as text, number, date but also includes spatial data types which are used to model the geometry of spatial data items. In this thesis the term 'spatial entity' refers to something that exists in reality and an 'object' is the representation of that entity in a spatial database. Spatial database can also model spatial and thematic relationships between objects. Whilst a spatial database is the core content of any digital generalisation process, the underlying modelling process is an essential step that creates a meaningful database for any given application (Peng, 1997).

A spatial database contains data that is a representation of a certain perception of geographic space; in other words, a model of the real world. There are many descriptive models for geographic phenomena and processes with different levels of

complexity (Molenaar, 1996a). These models express the way the world is perceived and is governed by the application of interest (Burrough & Frank, 1996). The process of modelling involves creating a representation schema for the real world phenomena that later can be implemented in a computer environment and be used for building a database. This process is called 'data modelling' or 'spatial data modelling' (Molenaar, 1998; Peng, 1997; Peuquet, 1984; Worboys *et al.*, 1990). The process of modelling is the most essential step that creates a meaningful database useful to a given application (Nyerges, 1991). This is because it identifies which entities are useful for a given application, what are their relationship, how they will be structured and implemented (Molenaar, 1996b).

The data modelling process that defines how the data are going to be structured such that it is interpretable by the computer hardware is called 'physical data modelling'. At a higher level is the 'logical data model'. This deals with how the data are structured in the database. But before data can be mapped onto logical and physical data models, relevant spatial entities, how they are structured (section 2.2), how these are categorised (section 2.3 and 2.4), their properties and their mutual relationships (section 2.5) need to be identified. This level of data modelling is called 'conceptual data modelling' (Molenaar, 1996a, 1998; Peng, 1997). Of these three, the conceptual data model plays a central role in the process of generalisation (Peng, 1997) since it provides a design or framework to interpret the database. Understanding of the database in terms of meaning and relationships of its object is essential for its transformation (Nyerges, 1991). The relevant aspects of conceptual data modelling are discussed in more detail in subsequent sections.

## **2.2 Field and Object Structure Approach**

Spatial objects in a spatial database are a representation of the real world. They contain both thematic and geometric (spatial) information that are normally represented through their thematic and geometric attribute values. There are two main structuring techniques for linking the thematic and geometric values i.e the field approach and the object-structured approach (Bian, 2007; Burrough, 1996; Goodchild, 1992; Molenaar, 1998; Worboys, 1995).

A field approach represents the earth's surface as a continuum. Typical examples of geographic phenomena that are best modelled as a field are air pressure, elevation or temperature. In the field model each attribute is assumed to vary continuously and smoothly over space. The values of these attributes are considered to be position dependent (Figure 2.1a). The representation of such a field in a spatial database requires that the continuum is discretised in the form of points or finite cells often in the form of a regular grid. For instance a digital terrain model (DTM) is usually represented as a grid of cells (raster data structure). Each cell in the DTM represents the terrain elevation and the size of the cell represents the resolution of the grid.

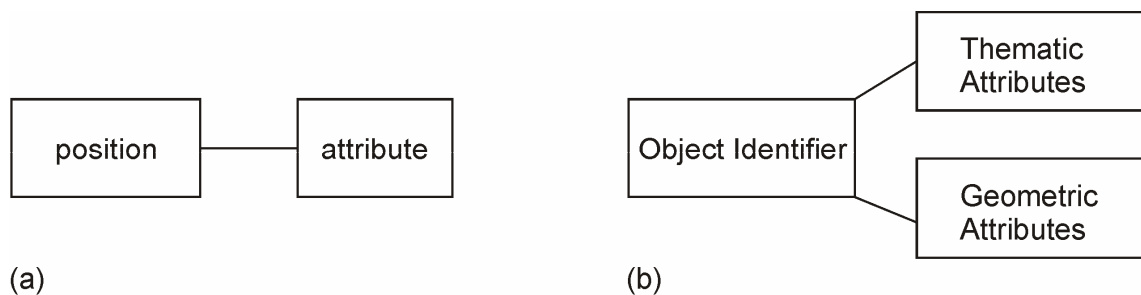


Figure 2.1: Two principle approaches for structuring spatial data (a) field approach (b) object structured approach (Molenaar 1998)

An object approach reflects the perception of the real world in terms of discrete entities that have crisply defined extents. An object approach is typically used to model heterogeneous objects (where attributes of the object apply uniformly to the whole object) and their boundaries have been accurately surveyed. For instance buildings, roads, land parcels. Spatial entities such as cities, towns, forests, hills and ranges are examples of spatial regions with no exact boundaries but can be perceived and modelled as spatial objects depending on how they are defined (Mark & Smith, 2004) and modelled (discussed in more detail in Chapter 5).

Typically a spatial entity in the object approach is modelled by crisply delineated 'points', 'lines' and 'area' (or polygon) objects. A point represents a pair of coordinate values, lines link series of exactly known coordinates (points) and area objects (polygons) are defined by boundary lines. In an object structured approach the link between thematic data and the geometric data is usually made through an object identifier (Figure 2.1b) (Molenaar, 1998). This object identifier is unique for

each object. The data structure commonly used for object structured approach is the vector data structure. In this research the primary and target database are modelled using an object approach (in vector format) though an additional dataset (DTM) used was structured in raster format.

### 2.3 GeoSpatial Ontologies

Geo-Ontology is recognised as an emerging research initiative by the University Consortium for Geographic Information Science (Agarwal, 2005; Mark *et al.*, 2004; UGIS, 1998). The term ontology has been widely used in information systems and in philosophy in a number of ways. Guarino and Giaretta (1995) discuss the issues regarding the use of this term and provide a formal definition of ‘Ontology’ (with an upper-case O) as “that branch of philosophy which deals with nature and organisation of reality” (Guarino & Giaretta, 1995). In this sense the ontology of geographic phenomena would deal with the totality of geospatial concepts, categories, their properties, relationships and geo-spatial processes and with their interrelations at different resolutions and representations (Mark *et al.*, 2004; Smith & Mark, 2003). The term ontology (with a lower-case o) in information and artificial intelligence refers to “specification of conceptualisation”(Gruber, 1993; Gruber, 1995). The ontological process is defined as the ‘definition or identification of entities that can act as referents for capture of the real world’ (Agarwal, 2005; Frank, 1997; Mark *et al.*, 1999; Milton & Kazmierczak, 2003). There are a number of ways of conceptualising the real world hence there can be multiple ontologies (Bittner & Geoffrey, 2001; Fonseca, 2001). This process of conceptualisation involves a system of concepts and categories which divide the corresponding universe into objects, processes and relationships between objects and processes (Smith & Mark, 2001).

Everything that exists in the spatio-temporal world is an entity and an ontology is captured by depiction of the entities which exist within a given portion of the world at a given level of generality. It includes the types of entities and relations that exist between entities (Grenon & Smith, 2004). The ontologies of geographic phenomena can belong to a specific domain or application or can be generic i.e. upper level ontologies (Agarwal, 2005; Fonseca, 2001; Grenon & Smith, 2004; Kuhn, 2001). For instance a cartographer might use a different definition for a city as compared to the definition used by an economist. Similarly for other geographic phenomena such as a lake, mountain, hill, or forest different disciplines might use different ways of defining each of these concepts. The goal is to express the different views of the real world via ontologies in a manner that provides a formal way of sharing knowledge across domains. A geospatial ontology within a specific domain will define the relevant concepts, categories, spatial objects, their boundaries (fiat or bona fide), and spatial and non spatial relationships (Smith & Mark, 2001; Smith & Varzi, 2000). Each of these components of spatial objects are described in more detail in subsequent sections.

Because there are different levels of detail in the real world that need to be represented, so there are different levels of ontologies (scale dependency) (Reitsma & Bittner, 2003; Smith & Mark, 1998). The level of detail of an ontology is related to the level of detail of the geographic information. Geospatial ontologies is a vast and developing area of research. Here the intension is to describe the relevant concepts of this field that are important to the understanding of subsequent discussions and the proposed methodology. The purpose here is to describe the concepts of ontologies in terms of generalisation in order to provide a conceptual framework in which to 'position' this work – namely the creation of objects of higher level ontologies from the spatial objects at more detailed levels.



## 2.4 Classification of Objects

A data model defines the *type* of objects in a database. Each object in the database has its own set of thematic and geometric attributes. But this does not mean that all attributes are different as compared to other objects. Objects may have common behaviour or attributes (thematic or geometric). Also objects can have similar attribute values. Thus objects can be associated with each other in terms of common attribute values or common behaviour. This helps in dealing with groups of objects that can be represented and generalised in the same way. The process of categorisation of objects is called *classification*. A class is a set of objects that share a common attribute structure (object class) or attribute values (data class) (Molenaar, 1998). For instance all objects in the database that have the ‘descriptive’ attribute value ‘Building’ can be classified as a members of the ‘Building’ class. Similarly we can have tree, road, railway and land parcel classes. The objects of a class are called its instances. The relationship between an object and its class is an ‘is-a’ type. If each object in a given spatial database is an instance of some class and each object is an instance of one class only then can classes form a ‘thematic partition’ of the dataset (Molenaar, 1998).

A common way of storing the new class information of an object in a database is to store the name of the class as an attribute value (data class). Another way is to create a new class with its own attribute structure and then transfer the appropriate objects into it. Objects of such classes will have the same attribute structure (object classes). Each class has a certain criterion that needs to be met for an object to be categorised as an instance of that class (Carnap, 1956). This criterion is termed the *intension* of the class and helps to differentiate objects belonging to different classes (Molenaar, 1998). For instance, the descriptive attribute equal to ‘Building’ is the intension of the above example of a building class. A detailed data model will have a detailed set of classes and as we move towards more abstract data models specialised classes are replaced by general classes. This idea is explained further in next section.

## 2.5 Relationships

Objects in a spatial dataset are related just as real world entities have relationships with each other. Modelling these relationships is required in order to understand the role of each object in the database. Thus new objects can be created by the combination of group of objects that have a similar role. Just as a spatial object has both thematic and geometric properties similarly they have both spatial and thematic relationships. Different type of relationships for spatial objects for map generalisation have been summarised by Steiniger and Weibel (2007). Relevant to this research are topological (spatial) relationship and two types of thematic relationships (taxonomic and partonomic). These relationships provide means of converting the source database objects into required objects.

### 2.5.1 Topological Relationships

The word *topology* is derived from the Greek for place: *topos*. Topology is the mathematical study of geometrical properties of objects that are preserved when the object is distorted. In topology two objects are considered to be the same if either can be distorted to form the other without being cut or torn. For instance in topology a circle is equivalent to an ellipse into which it can be deformed by stretching. Similarly a sphere is equivalent to an ellipsoid. However a sphere and a torus have different topologies. The field of topology can be further divided into three sub fields: point set topology or general topology, algebraic topology and geometric topology (Bredon, 1993; Moise, 1977; Rotman, 1988; Willard, 1970). The most basic and fundamental is point- set topology. It studies properties of spaces and maps such as connectedness, compactness and continuity. It defines the basic notions such as open, closed sets, interior, boundary, exterior, closure, incidence, inclusion, neighbourhood, closeness, compactness, continuous function and provides theorems to prove them.

Topology is important in cartography because it enables us to understand different types of spatial relationships, called topological relationships, between spatial objects. Egenhofer and Herring (1990) proposed a method for deriving topological relationships based on point set theory of algebraic topology (Egenhofer & Herring,

1990). For spatial objects that are structured in vector format their topological structure can be analysed using graph theoretic elements i.e node (point), segment (lines) and faces (polygons) (Gersting, 1992; Liu, 1983; Molenaar, 1998). In the point-set model each object is modelled as a point set with three elements; a boundary, an interior and an exterior. The topological relationship for each object is determined by an intersection of each of these three elements against elements of another object. This results in a nine intersection matrix (Egenhofer & Franzosa, 1991; Egenhofer & Herring, 1990) that describes the possible topological relationship that can exist between spatial objects. In this research we deal principally with polygon objects. The topological relationships for two simple polygons (with no disconnected geometries and holes) are summarised in Figure 2.2 (Egenhofer & Franzosa, 1991). The topological relationships have been further extended to model more complex spatial objects (Cohn & Gotts, 1996; Egenhofer *et al.*, 1994a; Tryfona & Egenhoer, 1997).

Topology for a set of objects in a given data set can either be explicitly stored or derived at the point at which they are needed. A few data structures have extended the primary vector data structure in order to store the topology as part of the data structure. Examples include Arc/Node format or formal data structure (FDS) (Molenaar, 1998) or extended formal data structure (EFDS) (Peng, 1997). The methodology presented in this work did not require that topology be modelled as part of the dataset. Instead topological relationships were determined using the topological operations in the spatial database (Oracle Spatial 10g) as and when they were required (discussed more in Chapter 6).

Topological relations between spatial objects impose certain constraints that are critical for any generalisation process (Harrie & Weibel, 2007; Molenaar, 1998; Paiva, 1998; Peng, 1997). For instance an object should not overlap with another object as a result of a generalisation process. If this results from some generalisation operation (for instance due to the exaggeration of some objects) then the conflict needs to be resolved. Topological relations are also useful in the creation of objects at higher levels of abstraction via composition or aggregation of source database objects. These relationships make it possible to formulate consistency rules at

several levels of abstraction and during update operations (Molenaar, 1998). These relationships will be used in Chapter 6 in order to identify associations between objects at different levels of abstraction. This will enrich the source database required prior to its transformation.

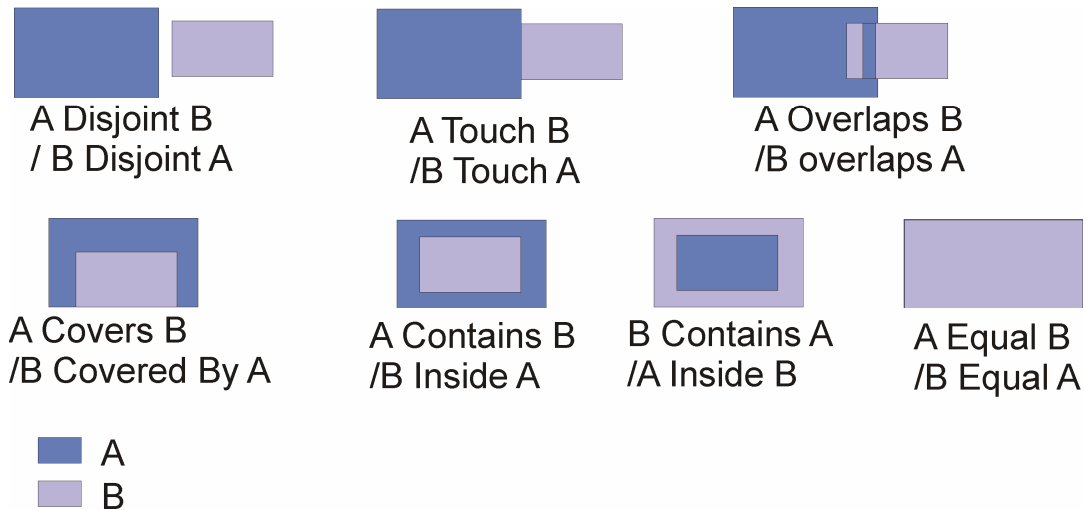


Figure 2.2: Topological relationships between polygon objects with no disconnected parts

### 2.5.2 Taxonomic Relationships

As explained in section 2.4, classification is the process of categorising objects into classes. Some classes can be more detailed than others. The level of classification depends upon the resultant application and the level of detail in the attributes. For instance a vehicle class can further be classified (specialised) into cars, motor bikes, vans, trucks, and bicycles or a building class can be further refined into a house or factory or shop class. Similarly a road class can be further classified into a motorway class or highway class or pedestrian path class (Figure 2.3). Each of these classes can be further sub classified. In these examples vehicle, building and road classes are more generic classes and are called ‘super’ or ‘parent classes’ whereas motor bikes, shop and motorway are more detailed classes and are called ‘sub classes’.

Sub classes and their super classes form a hierarchical structure called as *classification hierarchy* (Molenaar, 1993; Molenaar, 1998; Smith & Smith, 1977) or *taxonomy* (van Smaalen, 2003) (Figure 2.3). The relationship between a sub-class

and super-class is of ‘is a’ type and is called a taxonomic relationship. An advantage of taxonomies or classification hierarchies, in terms of generalisation, is the level of abstraction associated with the different levels of the hierarchy (Figure 2.3). Thus the taxonomic relationships provide ways of transforming a complex model into a less complex one (model generalisation). One can readily envisage different spatial dataset granularities associated with each level in the classification hierarchy. The lower levels in the hierarchy correspond to higher levels of detail and thus more detailed data. Whereas the higher levels correspond to higher levels of abstraction and thus lead to less detailed data (both thematic and geometric). Changing the sub class into classes at higher levels in the same classification hierarchy would mean transforming the model from a higher level of detail to a higher abstraction level. This will lead to a generalisation process that converts instances of sub class classes into instances of super classes. This type of database transformation is discussed in more detail in Chapter 3 (section 3.6.3).

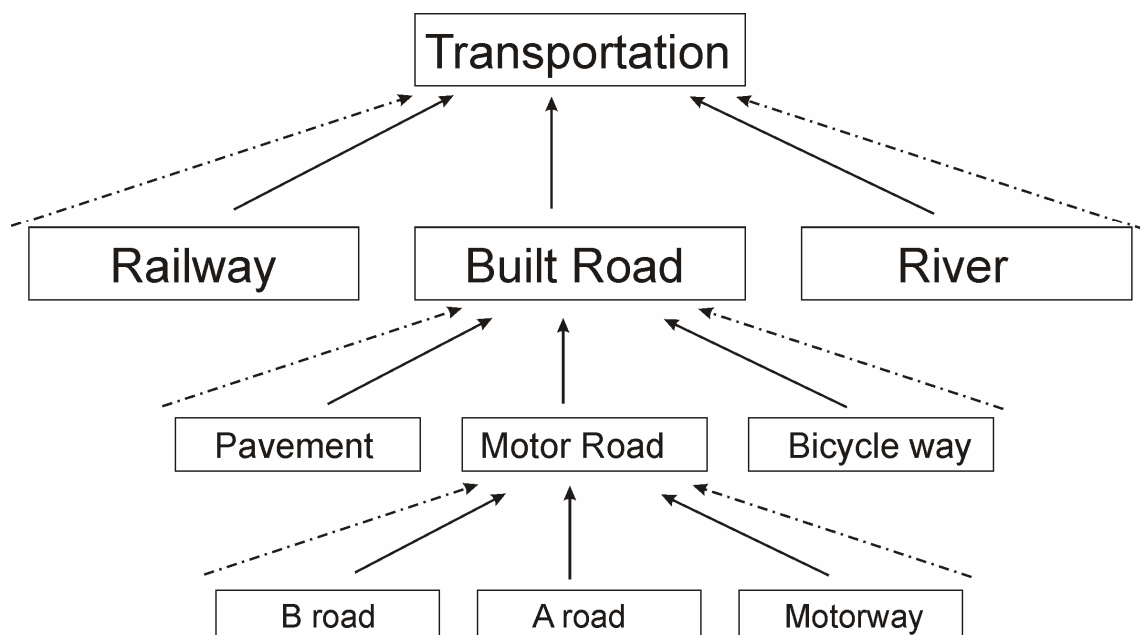


Figure 2.3: A possible classification hierarchy for transportation. Different levels of hierarchy reflect different levels of detail (modified after Peng (1997))

### 2.5.3 Partonomic Relationships

We can classify and relate objects according to their thematic similarity in the sense that they belong to the same class or share a common super class. But objects

belonging to a different classification hierarchy can also be related. For instance an object of class building block is a combination of objects instances of the classes street, garden, pavement and buildings. The relationship between these objects is of the type 'part of' (a street is part of a building block and so is the pavement, building and garden). This type of relationship is called a partonomic or functional relationship (Molenaar, 1998; van Smaalen, 1996). Partonomic relationships have their roots in the theory of *mereology*. Mereology is the theory of relationships of part to whole and the relations of part to part within a whole (Simons, 1987; Varzi, 2003). 'Parts' provide a conceptual skeleton for linking the functionality, appearance and behaviour of the resultant concept with their constituent parts (Varzi, 2003). This provides ways of transforming concepts at higher level of detail into concepts at lower level of detail. It mirrors our functional and conceptual understanding of geographies.

A class, such as building block, that is created by the composition of classes belonging to different classification hierarchies based on partonomic relations and is called a composition class. And classes that are involved in its composition are called component classes (van Smaalen, 2003). Accordingly the instance of a composition class is called a composite object. And an instance from which a composite object is created is called a component object. Both the thematic and geometric descriptions of a composite object are normally derived from the geometric and thematic descriptions of its constituent objects but it can also have additional attributes (Figure 2.4). For instance, a city object can have the additional attribute of 'population density' or 'building density'. When objects are aggregated into a composite object their thematic descriptions become one. A composite object does not necessarily have to be contiguous. It can consist of several disconnected parts. For instance an archipelago is a collection of disconnected islands. A composite object is a relationship between two or more component objects seen as a new object (Smith & Smith, 1977).

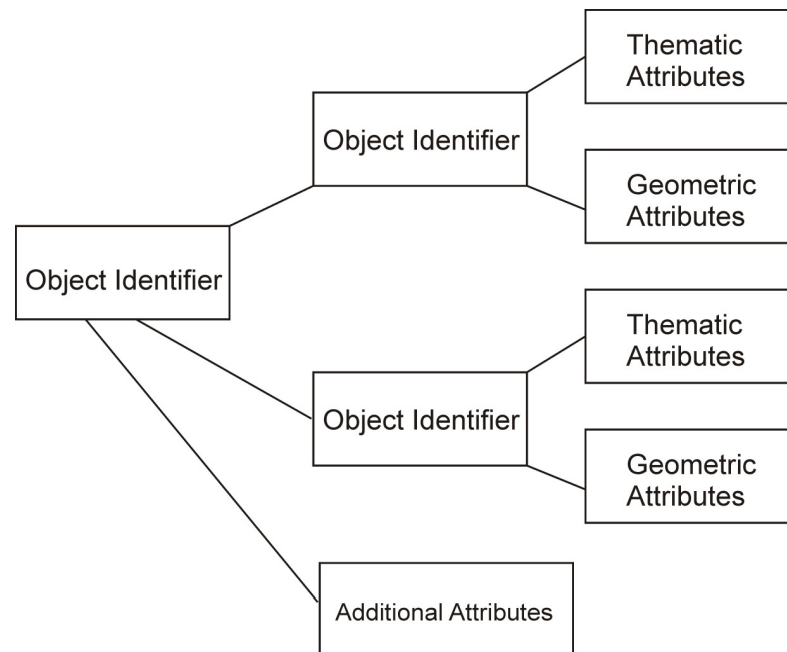


Figure 2.4: Composite object consisting of two component object(van Smaalen, 2003)

Partonomic relationships have been further classified in research in a number of different ways (Gerstl & Pribbenow, 1995; Iris *et al.*, 1988; Winston *et al.*, 1987). These classifications could be based on functional, compositional, structural or behavioural properties. ‘Organ-body’ or ‘engine-car’ are examples of functionally based classification whereas ‘a bunch of grapes’ or ‘a pint of milk’ are examples of compositional classification. It is important to point out that due to the nature of these relationships it is quite difficult to use any single classification schema that is appropriate in all cases (Gerstl & Pribbenow, 1995). However these classifications distinguish *necessary/canonical/good* parts from *optional/facultative/bad* parts (Cruse, 1986; Gerstl & Pribbenow, 1995; Tversky, 1990). Good or necessary parts are those parts which are both perceptually salient and functionally significant. Optional or bad parts are those parts which are occasionally found, or do not spring to mind when considering a typical definition. For instance ‘the seat of the chair’, ‘the blade of the saw’ are necessary parts of a chair or a saw whereas wheels of a chair or the cover of a saw are optional (or ‘bad’) parts. In a spatial context ‘trees of the forest’, ‘buildings of the settlement’ are good parts whereas scrub or lamppost are their optional parts respectively. This property of partonomic relationships can be

used in order to determine the extent of a composite object using their typical or necessary parts and will be explored in depth as key component of this research.

Just as in the case of taxonomic relationships, a super-class can be a sub-class of the next super-class in the classification hierarchy. Similarly a composite class can be a component class of another composite class. For instance, a building block can be a component class of a district which can be a component class of a city or settlement (composite class). This results in the creation of a hierarchy called an *aggregation hierarchy* or *partonomy* (Hughes, 1991; Molenaar, 1998; Peng, 1997; Thompson, 1989; van Smaalen, 1996) (Figure 2.5). Similar to a classification hierarchy this hierarchy also reflects levels of abstraction (or different levels of aggregation).

Composite classes and their instances (composite objects) are present as one moves up the hierarchy and their constituent components are presented below. This implies that the introduction of a composite class in the model will result in a transformation of the model from a lower abstraction to a higher level of abstraction. A generalisation process will be required for the creation of instances (objects) of new composite objects from existing object instances of the existing component classes (see section 3.6.4 for more discussion). The hierarchy between composite objects and component objects is called an *object hierarchy* (Molenaar, 1998). It is important to point out here that a partonomy reflects a part-of relationship between component and composite classes whereas an object hierarchy establishes part-of relationships between the component objects and the composite object.



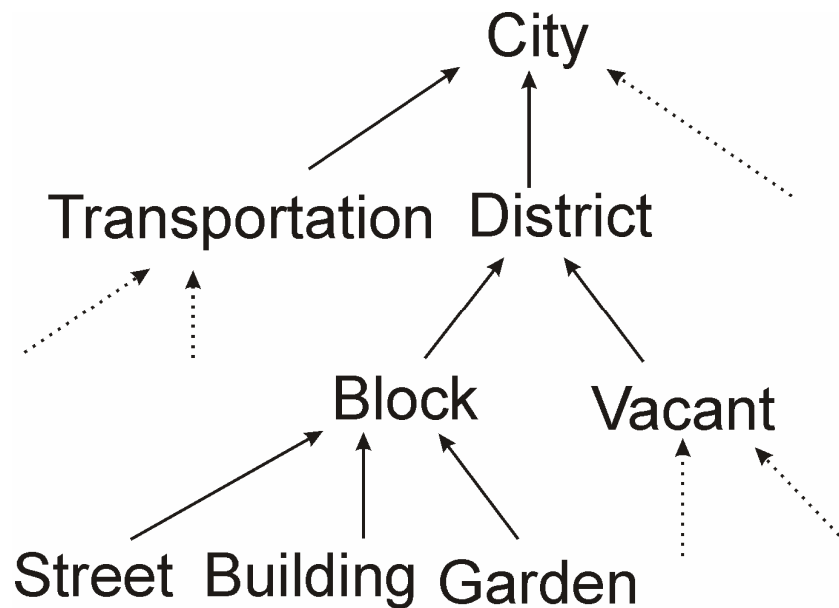


Figure 2.5: An example of an aggregation hierarchy of a city in terms of its parts (modified after Peng (1997))

The application of partonomic relationships extends beyond the classification and generalisation of spatial objects to other disciplines. For instance partonomic relationships are significant for classification in the medical domain since diagnoses, medical procedures and findings commonly relate to anatomical objects and their parts (Bernauer, 1996; Smith & Rosse, 2004). Partonomies are also important for structuring events. Like objects, events belong to categories and like objects they have parts (Barker & Wright, 1954; Zacks & Tversky, 2001). Event partonomies have been studied by observing how people segment activities as it happens. Structured representations of events can relate partonomy to goal relationships. Such representations have been shown to drive narrative comprehension, memory and planning. Computational models provide insight into how these representations might be organized and transformed (Zacks & Tversky, 2001).

#### 2.5.4 Taxonomy and Partonomy

Both taxonomic and partonomic relationships are the result of different but complementary modes of classification (Tversky, 1990). In simple terms a taxonomy is a classification based on similarities between classes whereas partonomy is a classification based on shared functionality between different classes. A partonomy can be created from a single instance of a class whereas a taxonomy is

the output of comparison between different instances (Tversky, 1990). Taxonomic relationships enable us to make inferences based on the knowledge of common properties shared by a class of entities. So in a taxonomy, classes at lower levels inherit attributes from classes at higher taxonomic levels. This contrasts with partonomies, which usually do not permit property inferences. A piston is part of a car but it does not inherit the properties of a car. Similarly a road can be part of a city but it's not a kind of city. But partonomic relationships might reveal a different kind of inference important in human cognition, the inference from appearance to function (Tversky & Hemenway, 1984). Partonomic relationships reveal subcomponents of an object and the relations among them whereas taxonomic classification reveals the properties or attributes shared by different instances.

In a taxonomy the relationship between a super-class and sub-class is usually 1:M (one to many) (Peng, 1997). Similarly the relationship between a class and its instances is also 1:M. In a partonomy the relationship between a composite class and component class can be M:N (many to many). For instance a garden (component class) can be part of a building block (composite class) and can also be part of park (composite class). Similarly the relationship between composite objects and component objects is M:N. For instance a railway station can be part of the railway network and also part of a city – reflecting the function of connecting the populous into the network. Similarly a river can be part of a hydrological system comprising rivers, lakes and streams. And the same river can also be part of a transportation network for shipping comprising rivers, streams and canals (Molenaar, 1998). Thus there are many to many relationship between the different aggregation levels. But for a given application it is valid to define criteria for composite objects such that they are mutually exclusive i.e. component objects belong to one composite object (Molenaar, 1998). This avoids conflicts during the process of database transformation (discussed further in Chapter 6).

Both taxonomic and partonomic relationships provide invaluable classification structures for our conceptual understanding of geographic space. Thus provide framework to carry automated generalisation process – that is to say the extraction

from lower to higher orders of abstraction (Mackness, 2006). This approach, will be used in this research as a basis for database transformation.

## 2.6 Generalisation of Spatial Databases

A spatial data model is the abstraction of certain real world phenomena at a certain level of detail. A spatial data model should always be constructed at a detail such that the modelled phenomena as well as the underlying process are meaningful and relevant for the given application (Müller *et al.*, 1995; Weibel, 1995a). A higher complexity model does not necessarily mean that it will be appropriate for a relevant application. This is because some detail may not be relevant and required information may be hidden by the “noise of detail”. Hence before a database can be constructed we need to determine the relevant aspects of reality required for the resultant application. This involves specifying the types (classes) of objects and their mutual relationships. A spatial database is then an instance of a particular spatial data model containing the objects instances of the classes defined by the new data model.

Objects in a database can be viewed as graphic representations. Because it is concerned with the graphical display it is scale dependent. The legibility of the graphic and the message that it may convey to the users are the main aspects to be considered. The process of transforming spatial data from one level of detail to another level of detail (i.e generalisation) thus involves two broad categories. One that focuses on transformation of the spatial data model (and database) from high levels of detail to a spatial data model (and database) at lower levels of detail. The other category of generalisation focuses on creation of a paper or digital representations of objects in the database. It focuses on visual enhancement of the output data. Within these two categories we can define the set of operations that are carried out during the generalisation process. The next chapter discusses these categories in more detail with the main focus on database (or model) generalisation.

## **CHAPTER 3: Generalisation in Digital Domain: Focus on Database Transformation**

This chapter seeks to discuss the process of generalisation in the digital domain with its main focus on the importance of database generalisation. The chapter starts with discussion of the objectives of generalisation. Based on these objectives and the concepts defined in Chapter 2 the current chapter outlines two broad categories of generalisation. The chapter then focuses on the importance of database (or model) generalisation with respect to multiple representation databases (MRDB) that are increasingly becoming important for National Mapping Agencies (NMAs). The next section discusses important factors affecting the process of database (or model) generalisation. The chapter then presents important model generalisation operations with a special focus on the aggregation operation. The last section gives an overview of the different stages of the proposed methodology based on the discussed concepts. The discussion here focuses on generalisation of spatial dataset that are structured using an object based approach in a 2D environment.

### **3.1 Objectives of Generalisation**

Chapter 1 stated that automated generalisation involves both modelling of geographic phenomena and the graphic display of the phenomena. These two processes are discussed in more detail in this chapter. Using the concepts discussed in Chapter 2, the following main objectives of the generalisation process can be defined (Peng & Molenaar, 1995):

- To derive a new (digital) database with different coarser geometric or thematic levels of detail from an existing database which exists at a higher detailed level for a particular application.
- To enhance graphic representation of the database objects when the output scale cannot accommodate the dataset of interest for visualisation purposes.

We carry out these objectives in order to provide users with fundamentally different views of the world – ranging along a continuum from the very detailed through to the

highly synoptic. The first objective relates to the aspect of changing the complexity of the spatial data model whereas the second one relates to the graphic representation of a database. In generalisation research the terms ‘database’ or ‘conceptual’ or ‘model’ generalisation are used to refer to the process that focuses on the first objective (Bertin, 1983; Kilpeläinen, 1997; Mackaness & Chaudhry, 2007b; Molenaar, 1998; Peng, 1997). Whereas the terms ‘graphic’ or ‘view’ or ‘cartographic’ generalisation are used for the process that focuses on the second objective (João, 1998; Kilpeläinen, 1997; Peng, 1997; Weibel & Dutton, 1999). In this research, the terms model and cartographic generalisation will be used.

### **3.2 Model generalisation and Cartographic Generalisation**

The relationship between cartographic and model generalisation is illustrated in Figure 3.1 (Grünreich, 1985). The database containing the first abstraction is typically called a digital landscape model (DLM –Figure 3.1). The DLM (primary) might be created by the acquisition of original information via surveying, photogrammetry or other means of capture. Typically a notional scale is associated with the DLM database though it is perhaps more apposite to talk of ‘level of detail’ or ‘resolution’ (Müller *et al.*, 1995; Peng, 1997). Even at this basic level the modelling of the real world involves generalisation in the form of selection and abstraction according to the conceptual data model (termed ‘object generalisation’). Databases that are derived from the primary DLM (having lower thematic and geometric resolution as a result of model generalisation) are called secondary DLMs. Both primary and secondary DLMs can be used to create a cartographic representation or a digital cartographic model (DCM) via the process of cartographic generalisation.

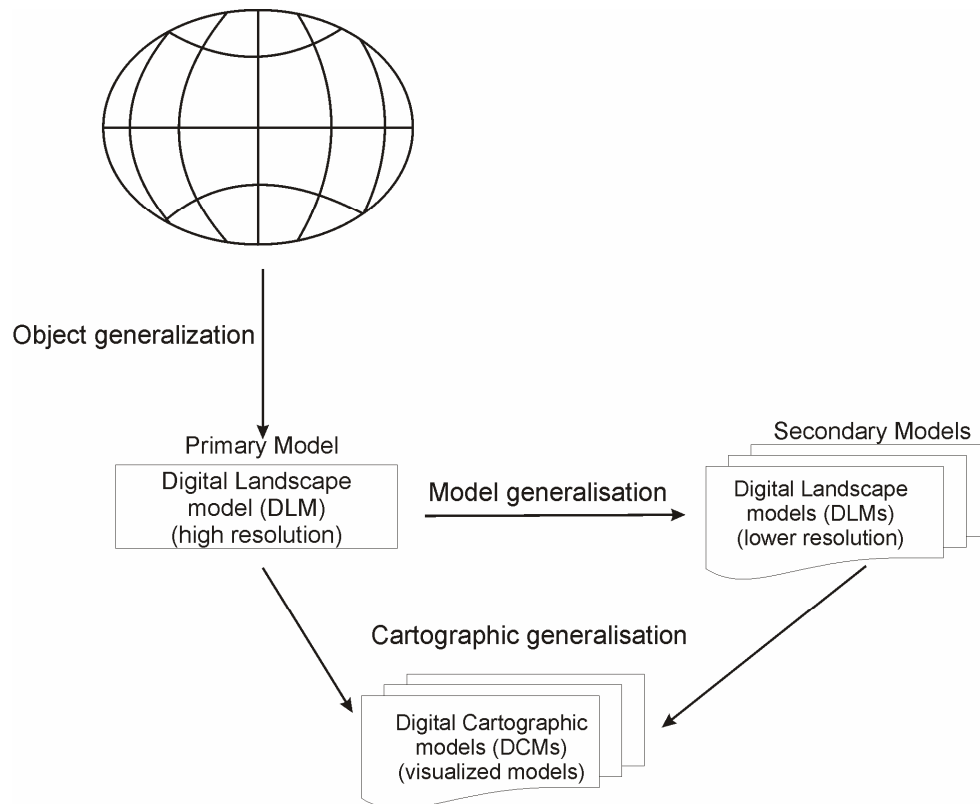


Figure 3.1: The first abstraction of reality creates the primary digital landscape model (DLM), from which a digital cartographic model (DCM) can be produced – either directly from the DLM or via the process of model generalisation and the creation of secondary landscape models (after Grünreich (1985))

The objective of model generalisation techniques is to reclassify and reduce the detail, thereby giving emphasis to entities associated with the broader landscape – thus enabling us to convey the extent of the forests rather than to see the trees, or to see the island chain along the plate margin, rather than the individual islands. The model generalisation process is not concerned with issues of legibility and visualisation. On the other hand cartographic generalisation is a set of techniques concerned with increasing the efficiency with which the map (paper or digital) is interpreted – thus the techniques aim to resolve ambiguity, and to retain those qualities of a representation that best fit with the user’s expectations. The process of cartographic generalisation is subject to the same principles as those that apply in manual generalisation processes (Weibel, 1986). It is argued in research that issues related to database transformation and those related to the graphical limitations of the output medium should be handled separately (Kilpeläinen, 1992; Sarjakoski, 2007). The separation of model and cartographic generalisation helps to manage the

complexity of the task. In display emphasis should focus on optimal graphic communication and on conventional cartographic generalisation (João, 1998) whereas for analysis and modelling attention should be paid to optimal preservation of data characteristics (Weibel, 1986). Müller (1989) pointed out that over small changes in scale objects before and after generalisation are mostly the same and thus geometric or cartographic generalisation can be directly applied on the source dataset in order to create a cartographic output. But over large changes in the level of detail, fundamental changes in the content and here conceptual or model generalisation becomes a critical prerequisite to cartographic generalisation (Weibel, 1995b).

### 3.3 Effects of Model and Cartographic Generalisation

Since model and cartographic generalisation have different objectives, they have different effects on the data, and both affect the accuracy of spatial databases in their own way (Müller, 1991). In model generalisation the reduction of data volume is maximised while at the same time the modification of the source data is minimised (João, 1998). Model generalisation usually involves operations such as selection, simplification, classification and aggregation (discussed in section 3.5.4) (Mackanness & Chaudhry, 2007a, b). On the other hand cartographic generalisation will have more effects in terms of accuracy because it operates by employing ad hoc decisions involving operations such as symbolisation, enhancement, exaggeration, smoothing and displacement (Müller *et al.*, 1995) and is therefore non-statistical in nature (Brassel & Weibel, 1988). Cartographic generalisation *'might cause displacement, distortions and exaggeration of map elements locally, if that is needed to preserve the characteristic look of the map'* (Weibel, 1992 p.314). As is illustrated in Figure 3.2, a building object represented in a database (Figure 3.2a) may be represented by a church symbol, in order to emphasise the object, in the corresponding map (Figure 3.2b). Lake and road objects have also been symbolised and exaggerated for cartographic reasons (Figure 3.2b). This has resulted in an overlap between the church and the road. Symbols need to be displaced in order to avoid this visual conflict (Figure 3.2c). Thus positional accuracy is sacrificed in order to maintain 'clarity' (topological constraint). Of course the positional accuracy in the database remains unchanged and unaffected by this process.

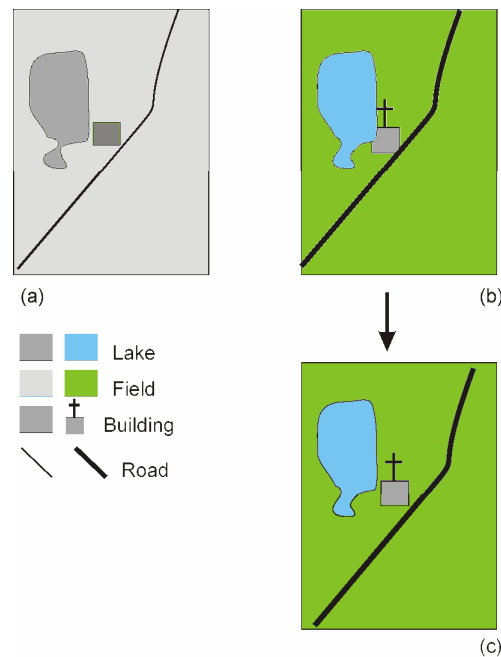


Figure 3.2: Effects of cartographic generalisation. a) Metaphorical representation of the database b) Cartographic representation of the database results in conflict between symbols c) Displacement is applied to remove the conflict (modified after (Sarjakoski, 2007))

Typically model generalisation precedes cartographic generalisation (Weibel, 1995b). Alternatively model generalisation may be required in response to a non-visual query, or as a prerequisite to data analysis. For example the question ‘what modes of travel exist between the cities of Edinburgh and Glasgow?’ requires us first to aggregate together phenomena at the fine scale (in this case dense regions of buildings to create two cities) in order to define the extent and general location of these two entities (Chaudhry & Mackaness, 2007). Only then can we identify, for example, the major roads that connect these two urban centres. The increased use of GIS and the huge number of spatial datasets being gathered from different sources has subsequently increased the importance of model generalisation. The increased importance of MRDBs (discussed in the next section) has also added to the importance of model generalisation.

### 3.4 Multiple Representation Database

In recent decades the central task of NMAs has been to establish spatial databases from which maps can be produced via cartographic generalisation (Sarjakoski,



2007). Mostly the databases are disconnected from each other. Thus the maintenance and updating of these disconnected databases is a major issue for NMAs. At the same time efficient cartographic generalisation methods are still required for map production. During the 1990s and up to present times, the problem of maintaining and updating multiple disconnected databases at different levels of detail has been approached by introducing conceptual models for so-called multiple representation databases (MRDBs) (Kilpeläinen, 1997). The term MRDB refers to a database structure in which several representations of the same geographic entity or phenomenon, such as a building or a road, are stored as different objects, at different levels of detail and are linked in some way (Sarjakoski, 2007). The relationship between MRDB, model and cartographic generalisation is expressed in Figure 3.3.

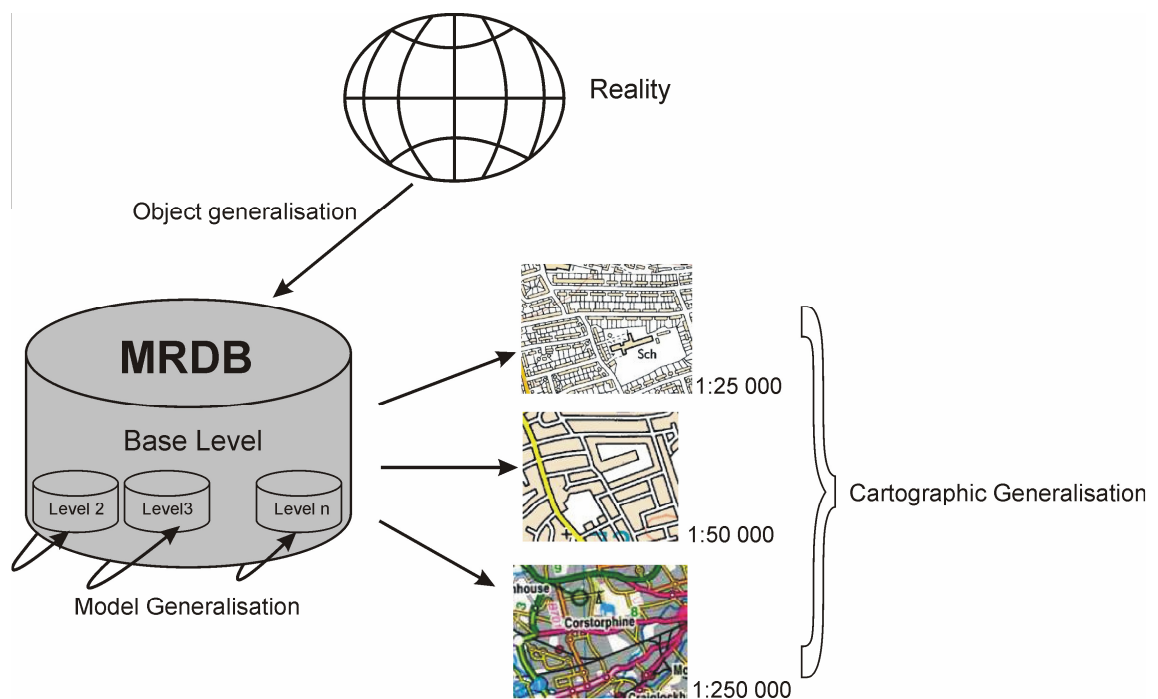


Figure 3.3: The relationship between MRDB , model and cartographic generalisation

A MRDB consists of various representation levels with different degrees of geometric and semantic abstraction providing a set of different views of the same object (Devogele *et al.*, 1996; Kidner & Jones, 1994; Kilpeläinen, 1992; Weibel & Dutton, 1999) (Figure 3.4). A MRDB emphasises the utilisation of geographic databases for various spatial applications, not just those that have been predefined for some specific map scale. Its flexibility lies in its ability to derive different types of

maps from different representation levels of a MRDB using cartographic generalisation techniques that results in the ability to create customised outputs. These databases are also sometimes called multi-scale or multi-resolution databases.

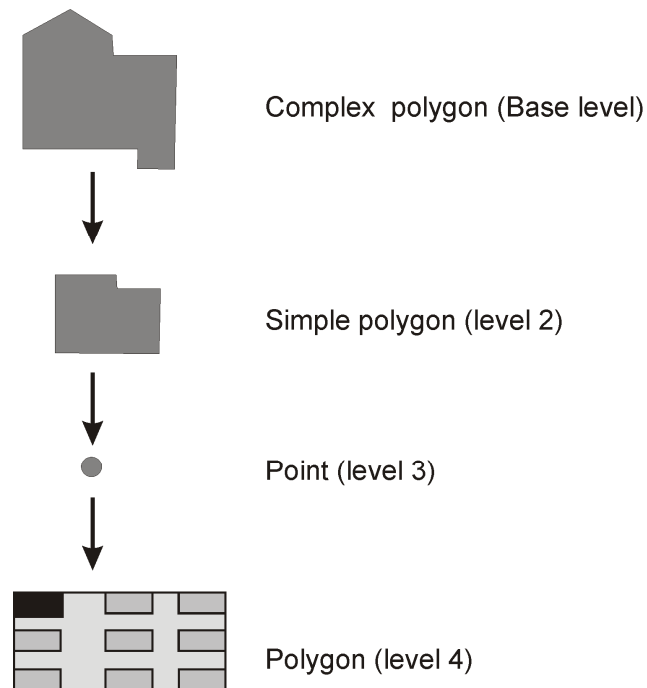


Figure 3.4: Representation levels for a building object in a MRDB (modified from Kilpeläinen, 1997)

Several researchers have explained the benefits of MRDB (Hampe & Sester, 2002; Kilpeläinen, 2001; Sheeren *et al.*, 2004). These can be broadly summarised as:

**Database Maintenance:** One of the biggest advantages of MRDB is the ease of database updating. If objects at different levels of detail are connected, updates done at the base level database can be propagated to smaller levels of detail automatically or at least semi automatically (Badard & Richard, 2001; Egenhofer *et al.*, 1994b; Kilpeläinen, 2001; Kilpeläinen & Sarjakoski, 1995).

**Database Consistency:** Since each object is stored only once, a MRDB avoids data redundancy in the databases. The links between the data also provide a basis for automatic error checking and quality control. If one representation is known to be of better quality then it can be used to control the quality of the latter (Mustière & van

Smaalen, 2007). Integration or 'links' between the data also allows inconsistencies and errors to be detected (Egenhofer *et al.*, 1994b; Sheeren *et al.*, 2004).

**Increased Efficiency:** If the databases at different levels of detail are connected this will increase the speed of retrieving data. This is important for time critical applications such as location based services. Here different levels of abstraction are typically linked by a tree structure; the level of the tree to be displayed is dependent on the current zoom level selected by the user (Jones, 1991; van Oosterom, 1995). This leads to ideas of 'intelligent zoom' where detail increases while zooming in and decreases while zooming out (Frank & Timpf, 1994; Jones, 1991; van Oosterom, 1995).

**Customised Datasets:** It is possible to derive application dependent generalised output from MRDBs. Different objects from different representation levels can be selected to create an output specific to the requirement of the user. In addition to this, these databases can also have multimedia data linked to the geometric representations (Kilpeläinen, 1997). For instance a house object in the base database can be linked with a photograph of that house or the sound representation of nearby traffic. Similarly the MRDB might also store temporal attributes associated with an object such as opening times of specific stores in order to plan and navigate a shopping trip (Linturi & Simula, 2005).

### 3.4.1 Creation of MRDBs

There are two main approaches to the creation of an MRDB (Sarjakoski, 2007) :

- Creating links between objects in existing databases (Harrie, 2001; Sester *et al.*, 1998; Uitermark, 2001)
- Generating smaller scale representations from a single large scale base database via model generalisation (Kilpeläinen, 1997; Kilpeläinen & Sarjakoski, 1995; Martinez & Molenaar, 1995)

The main issue regarding the creation of MRDB from existing databases is the matching of objects at different representations. And this becomes more critical if

the existing databases were created from already generalised maps (via digitization) or cartographic generalisation. This is because of the application of cartographic operations (carried out manually or automatically) such as typification<sup>2</sup> and displacement. After typification it is difficult to determine which cartographic objects represent the same spatial entity (Harrie, 2001). Similarly, due to displacement of features it becomes more difficult to match objects (Figure 3.5). In Figure 3.5 the settlement objects from OS Strategi (a cartographic product) are overlaid on top of OS base database (OS MasterMap Topography Layer). Because of cartographic operations reflected in OS Strategi, (as well as subjectivity in the human interpretive process) and the year of capture the matching is a difficult task. Another major limitation of this approach is that only already existing databases can be linked. Furthermore, quality control in the derivation process is absent since the derivation of the Strategi objects are done subjectively. In other words we lack a consistent, repeatable model by which one can be defined in terms of the other.

---

<sup>2</sup> Typification reduces the number of objects while preserving the distribution pattern (Regnauld and McMaster, 2007).



Figure 3.5: Matching settlement objects from Strategi cartographic dataset overlaid on building objects in OS MasterMap Topography layer (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved). A few building objects fall outside the settlement boundary due to cartographic operations such as displacement performed on the Strategi dataset

The second approach is based on the derivation of smaller representations from a single large scale spatial database (base level Figure 3.3) directly via model generalisation. MRDB and model generalisation are quite closely related (Sarjakoski, 2007). As discussed in previous sections model generalisation does not involve operations that deal with graphic output. Its main aim is the transformation of a database at high levels of detail to one of higher levels of abstraction. This obviates the need for an ‘object matching’ process between abstractions and avoids the need to link existing cartographic datasets. This approach was adopted by Kilpelainen (1997) using an object oriented approach for the multiple representation of building objects at four different levels of detail. The building objects at lower levels of detail were generated from a single large scale database (Figure 3.4). Similarly Martinez and Molenaar (1995) proposed an approach for multiple representation of hydrographic data based on an aggregation hierarchy. An aggregation hierarchy was also used by van Smaalen (2003) for the creation of multiple representations of a topographic dataset. Trying to create a MRDB via

model generalisation brings us back to the initial issue of needing to have an automated model generalisation approach. This research has its central focus on model generalisation and its relevant aspects are now discussed in more detail.

### **3.5 Critical Factors of Model Generalisation**

Liu et al (2003) made the observation that the database transformation is governed by four factors: the data model, the objects and their relations, the conditions and constraints, and a set of operations needed to perform this transformation. Each of these is briefly discussed.

#### **3.5.1 Data Modelling**

A data model defines classes and relationships between different classes and their instances. It determines what classes and which instances of these classes will be contained within a database. It also determines the degree of detail of the target database and the contents of the database. Database transformation takes place via the introduction of a new data model (Peng, 1997) since it defines its own set of classes and consequently a new set of instances (objects of these new classes). Model generalisation is essentially the transformation from one data model to another one. The classification and aggregation hierarchies play a critical role in this transformation (discussed further in section 3.6.3 and 3.6.4).

#### **3.5.2 Objects and Relations**

Objects and their relationships are the core content of any database. When a data model is changed, for the purpose of database transformation, this triggers the process of creating new object as instances of the new classes specified by the new data model. Any change in the objects will also introduce change in the relationships between objects. For example, a group of adjacent spatial objects with different attributes may be aggregated to form a homogeneous object. The individual and the notion of adjacency is 'lost' – overshadowed by the more general characteristic form. The new characteristic form may give emphasis to more gestaltic qualities such as patterns of alignment, or emphasis of orthogonal qualities. Depending upon the level of change, these transformations may include changes in the geometric and thematic

properties of spatial objects (in addition to topological changes). The transformation of a spatial database essentially involves the creation of new objects derived from the objects in the source database. A data model transformation controls the objects and the relation transformation (Liu *et al.*, 2003).

### 3.5.3 Conditions and Constraints

Database transformation is an application dependent process. Each application has a set of conditions and constraints that governs the process of transformation. These conditions and constraints affect properties of the resultant objects. These conditions and constraints define the intensions of the output classes. These may themselves be either statistical or topological. For instance a road network needs to remain topologically connected during generalisation, a surface must be space exhaustive, and settlement objects need to be of a certain size in the resultant database. Before the database transformation is performed, the conditions and constraints affecting the process of transformation must be identified.

### 3.5.4 Operations

In order to carry out database transformation certain generalisation operations are needed. Several researchers have proposed a set of generalisation operations (Li, 2007; McMaster & Shea, 1992; Regnaud & McMaster, 2007). A typical set of model generalisation operations are selection, classification, simplification and aggregation. These operations are critical in database transformation since they determine how objects in the target database will be created from objects in the source database. The operation of classification involves categorisation of objects (as discussed in Chapter 2). Selection and aggregation operations are discussed in more detail in the next sections since they are used extensively in the proposed methodology.

#### **Selection or Elimination**

Selection is the process of selecting a subset of features according to some criteria. The challenge in the selection or elimination (removal of unselected objects) operation lies in deciding which objects will be selected and which objects will be eliminated. The renowned radical law (Töpfer & Pillewizer, 1966) (Equation 3.1) is

one way of linking the level of detail with the number objects in the resultant database or map.

$$n_f = n_a \sqrt{M_a / M_f} \quad \text{Equation 3.1}$$

Where  $n_f$  is the number of objects which can be shown at the resultant scale,

$n_a$  is the number of objects at source scale,

$M_a$  is the scale of the source map,

$M_f$  is the scale of the derived map

But it does not give any information on which objects to retain and which to remove.

The selection or elimination criteria can be based on thematic or spatial attributes.

Often these operators require prior *database enrichment*. The process of making the required knowledge explicit and enriching the source data with this knowledge is known as data enrichment (Ruas & Plazanet, 1997). This process can signify the importance of objects in the database that can be used by the selection operator.

Importance can be measured in different ways. For example in an isolated region a small building might be retained whereas buildings of the same type and size are removed from a built up area at small scale. Many selection algorithms have been proposed. For instance Regnauld (1996) proposed the use of minimum spanning trees for the selection of buildings. Ruas (1998) proposed a density measure approach for reduction in the number of buildings in an urban building block. The selection of linear objects within a network becomes more complex because of the need to retain connectivity in the network. Graph theory has been used to manage problems where the edges or roads are assigned higher weights to maintain the connectivity of the network (Chaudhry & Mackaness, 2005a; Jiang & Claramunt, 2004; Mackaness & Mackechnie, 1999; Mackaness & Beard, 1993; Thomson & Richardson, 1995). Figure 3.6 illustrates the selection of important roads from high level of detail (1:50,000) to lower levels of detail (1:100,000) whilst maintaining the connectivity using graph theoretic techniques (Mackaness & Beard, 1993).



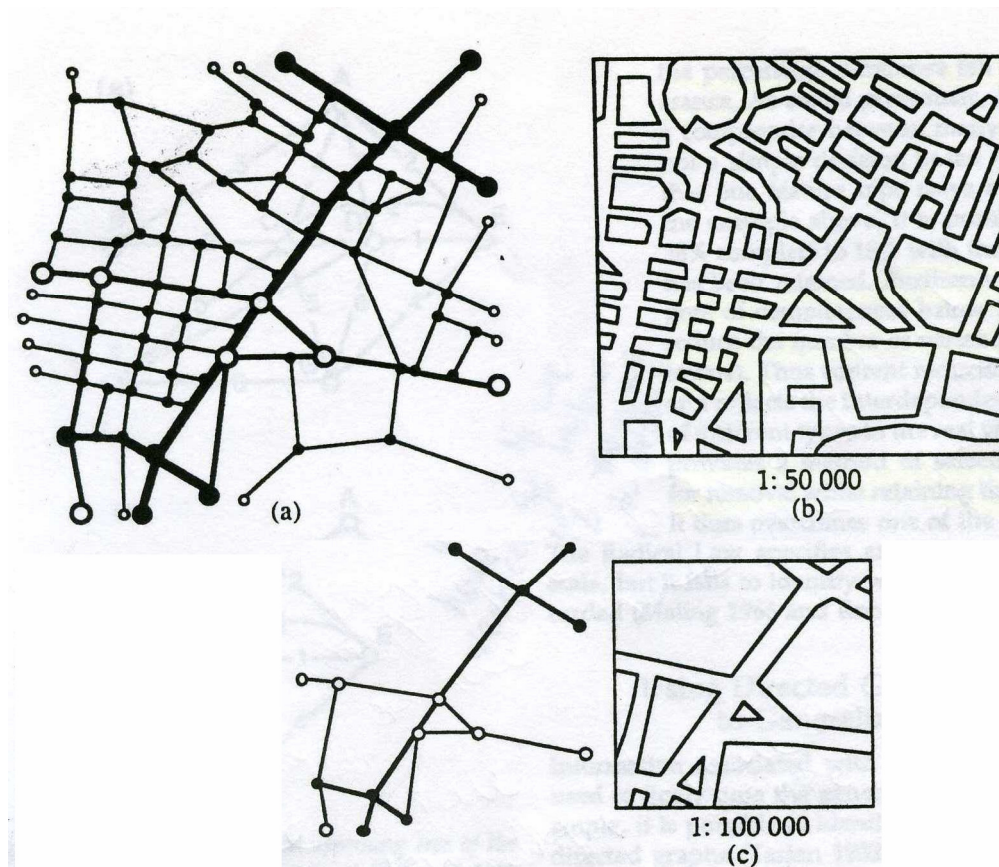


Figure 3.6: Example of selection operator. (a) A graph theoretic representation of road network at 1:50,000 (b) has been generalised to network at a lower level of detail (1:100,000) (c) whilst avoiding becoming disconnected (scanned from Mackaness and Beard (1993))

### Aggregation

Aggregation is one of the most important operations of model generalisation because of its ability to reduce spatial complexity by creating composite objects from their constituent objects in the source database (Mustière & van Smaalen, 2007; van Smaalen, 2003). An aggregation operator involves the process of joining together multiple objects into one object (Frank & Egenhofer, 1988) (Figure 3.7). The geometry of the resultant object is calculated from the geometry of its constituent objects (Molenaar, 1998). Aggregation can also be seen as a reclassification process of all constituent objects that create a composite object (Regnauld & McMaster, 2007). As a result of an aggregation process each source object, that is part of the composite object, becomes an instance of the composite object's class (composite class). Their attributes are modified according to the intension of the new class

(Regnauld & McMaster, 2007). In Figure 3.7 the two houses, land and pavement are all part of a block. Thus a block object can be generated by aggregating the geometries of its component objects. The aggregation operation is based on the ‘part of’ relationships defined in the aggregation and object hierarchies (section 2.5.3) in order to transform the source objects into target database objects.

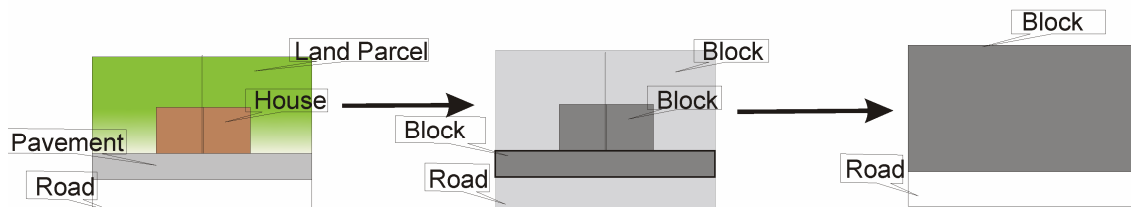


Figure 3.7: Aggregation of individual objects into a composite object. The geometry of the composite object is created by merging the faces (polygons) of individual objects

The aggregation operator is particularly relevant for geometrically or area partitioned datasets. A geometrically partitioned dataset is a dataset in which area objects cover the entire area of interest and do not overlap (space exhaustive) (Molenaar, 1998). In such datasets objects cannot be simply eliminated because elimination would create gaps (holes) in the resultant dataset. In order to avoid these situations objects have to be aggregated with other objects. For the aggregation operator to be applied, we must first identify the objects or groups of objects that will be aggregated. In other words a technique or strategy is required that governs which objects can be aggregated with other objects. A few important strategies for model generalisation processes that triggers aggregation are presented in the next section.

### 3.6 Strategies of Model Generalisation (aggregation methods)

Molenaar (1998) distinguishes four main techniques for the aggregation of objects in model generalisation (section 3.6.1-3.6.4). All these four methods of aggregation are intended for the generalisation of datasets that form thematic and geometric partitions.

### 3.6.1 Geometry Driven Generalisation

In this type of generalisation the resolution (level of detail) of the source database is reduced with respect to the target database resolution. In the case of vector structured datasets this type of generalisation involves defining a geometric threshold such as a minimum area or minimum length or minimum compactness. Objects below the threshold can either be eliminated by aggregating them with adjacent objects or if there is a collection of small objects they can be aggregated to create a larger object which is above that threshold. For example in Figure 3.8 the scrubs and individual farm objects in the source dataset (Figure 3.8a) have areas below the threshold. Simple elimination of these objects would result in gaps (holes) (Figure 3.8b). In order to avoid these situations, the scrub object is aggregated with the large adjacent object (tree) (Figure 3.8c). The cluster of farm objects in Figure 3.8a are mutually adjacent and can be aggregated in order to create a farm object with an area greater than the area threshold (Figure 3.8c).

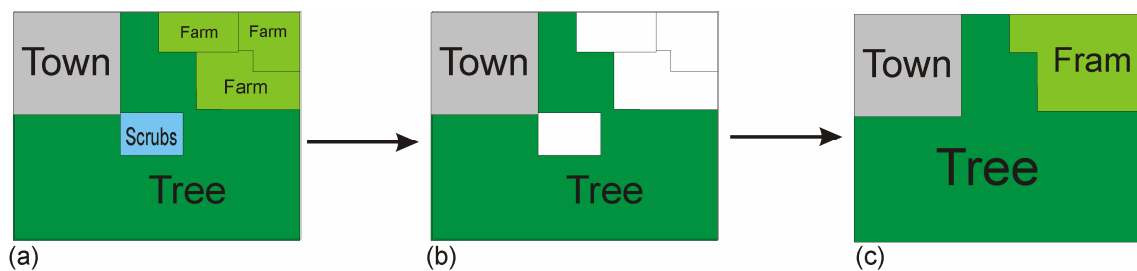


Figure 3.8: Example of geometry driven generalisation. Elimination of preliminary scrub and farm shown in a) that are below the area threshold creates holes (b). To avoid this, the scrub object is aggregated with the tree object and the small farm objects have been aggregated to create a single farm object (c)

### 3.6.2 Structural Generalisation

Structural generalisation is based on the hierarchical relationships in a network structure. Road network generalisation is an example of structure driven generalisation. Where graph theory is usually used for modelling a network (Mackness & Beard, 1993; Thomson & Richardson, 1995) in order to ensure the connectivity of the resultant roads (Figure 3.6). This type of generalisation can also be applied to the aggregation of area or polygon objects. For instance, the aggregation of catchment areas following the elimination of stream elements in

order to retain the output flow at the outlet (Martinez, 1994). In both geometric and structurally driven generalisation techniques, the spatial or geometric properties trigger the generalisation process and the thematic properties are adjusted after aggregation or elimination of the objects. These approaches contrast with the next two approaches where the thematic abstraction triggers the aggregation process in geometric partitioned datasets.

### 3.6.3 Class Driven Generalisation

A spatial database can be transformed via model generalisation based on the similarity between classes of objects in the database. Class driven generalisation makes use of classification hierarchies (taxonomies). As stated in section 2.5.2 classes at different levels in the hierarchy correspond to different levels of detail. More detailed classes are present lower in the hierarchy and general or parent classes are at higher levels in the hierarchy. As illustrated in Figure 3.9, objects belonging to ‘coniferous’ and ‘non coniferous’ tree classes have been aggregated into objects of a generic ‘forest’ class object. The forest object is aggregated with objects of other similar classes (scrubs and grassland) to create objects of a more generic class ‘natural area’. Here the thematic abstraction triggers the spatial abstraction.

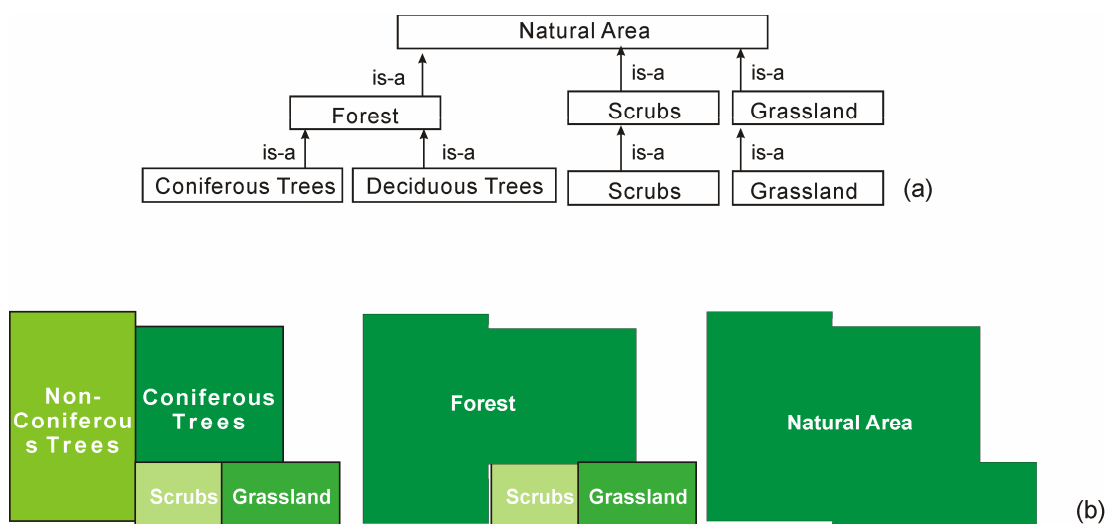


Figure 3.9: (a) Example of a Classification Hierarchy or taxonomy. (b) Class driven generalisation based on the taxonomy

Downs and Mackaness (2002) used a taxonomy for aggregation of different types of rock units in the source database, into more general rock units at lower levels of detail. Similarly Liu (2002) used a class driven generalisation approach for aggregation of land use units in the source dataset by classifying the objects to more general classes and then aggregating the adjacent objects that belong to the same parent class.

Classification or taxonomy driven generalisation is a straightforward approach but can only be followed if such hierarchies exist. A special case of class driven generalisation is 'similarity driven generalisation'. In similarity driven generalisation the strict criteria of aggregating of objects with a common super class is relaxed by using a similarity measure between source and target classes to allow a wider range of choices (Yaolin *et al.*, 2002). These similarity measures can be calculated by comparison of attribute values of objects of different classes. Attributes with ratio scale values can be compared more objectively, in contrast to attributes of a nominal scale (van Smaalen, 2003). Another limitation of this approach is the labour-intensiveness of creating these similarity values between classes (Bregt & Bulens, 1996) and they are therefore rarely available. In both of these approaches it is the thematic similarity that drives the process of generalisation and it depends upon the thematic specification of the dataset (Molenaar, 2004).

#### **3.6.4 Functionality (or Partonomy) Driven Generalisation**

It is not always possible to aggregate source objects into composite objects based on a single taxonomy (Molenaar, 1998). But often we want to combine phenomena from different classifications. Whilst class driven generalisation is valid over small changes in level of detail (van Smaalen, 2003), over large changes in level of detail (as is the case in this research) we need to combine objects that might not have any thematic similarity. As discussed earlier over large scale changes there is fundamental changes in content (Mackaness & Edwards, 2002; Müller, 1989). For instance a building object is valid over a limited range of scale but at higher levels of abstraction we need to switch to class 'block' or 'settlement' in which object such as roads, parks, streets and buildings need to be aggregated (Figure 3.10).

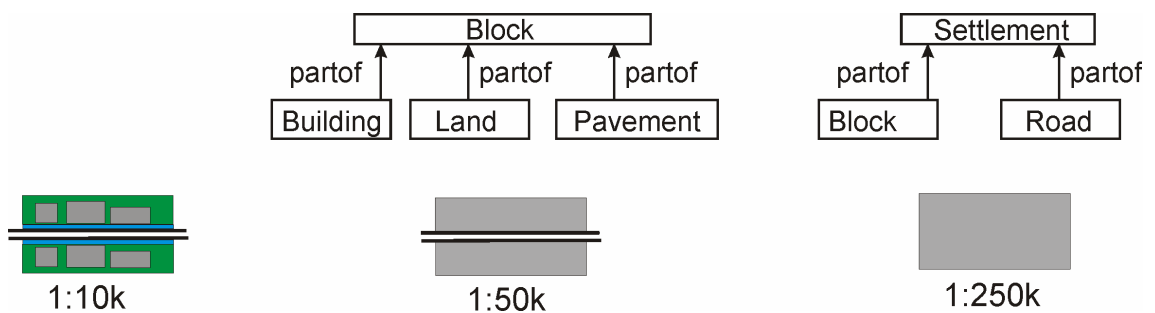


Figure 3.10: Aggregation hierarchy and aggregation of objects based on their partonomic relationships. Here objects belonging to different taxonomies have been aggregated to create objects (block, settlement) that are functionally (or partonomically) related instead of taxonomically

This type of aggregation based on shared function is expressed in terms of partonomic relationships (section 2.5.3). For instance a city might be made up of a density of roads, churches, industrial quarters, stations, and political institutions – it is what defines ‘citiness’ (from a prototypical and functional point of view). All these objects belong to different classification hierarchies, but when in physical proximity and of sufficient density, it is valid to aggregate them and create a new object of class ‘city’. In this manner, a particular set of objects are ‘part of’ a particular instance of a city. Thus if we know the parts of a composite object we can aggregate these parts to create the whole they are part of. In other words, a composite object is actually the merological sum of its constituent parts (Varzi, 2007). This type of model generalisation is termed ‘functionality driven generalisation’ (Molenaar, 1998). In addition to aggregation of objects these relationship also serve as links between source and target objects i.e. links between objects at different representation levels in a MRDB (Mustière & van Smaalen, 2007; van Smaalen, 2003). Robinson (Robinson, 1995) describes an aggregation hierarchy for a building block. Ruas and Lagrange (1995) observed that a hospital is composed of a set of buildings and areas. van Smaalen (2003) proposed an iterative object aggregation approach based on functional or partonomic relationships. This is essentially a stepwise approach for the combination of source classes and resultant

classes. These examples represent the few instances where functional driven generalisation has appeared in research in model generalisation.

Over small changes in scale aggregation can be based on taxonomic classification. But over large changes in the level of detail, as is the case in this research, we require aggregation of objects that are not taxonomically but partonomically related. The next section presents the classes of the target data model identified for this research. Section 3.8 gives an overview of the different stages of the proposed methodology for transformation of source database objects into objects of the required classes.

### **3.7 Classes of the target Data Model**

As stated earlier, a data model defines the classes at the resultant level of detail. The level of detail of these classes depends upon the resultant application. The selection of an appropriate level of detail is comparable to the work of selecting a proper map scale (Peng, 1997) for a given application. The aim of this research is to directly transform objects in the source database (OS MasterMap ) into objects that are present at a notional level of detail found at 1:250,000 scale. In order to perform this database transformation the first step is to identify the classes of the intended data model. Once the required classes have been identified we need an aggregation approach for transforming the source objects into target objects.

The output classes for the target database were identified by examining OS Strategi dataset. The OS Strategi dataset is an OS vector product at 1:250 000 scale (Ordnance Survey, 2007e). From these observations the classes listed in Table 3.1 were identified. In order to create objects of these classes we also need to define the intension (or criteria) of each class. These intensions were determined by studying the existing set of rules (Ordnance Survey, 2005) or by empirical analysis and observations. Table 3.1 illustrates the classes present in the target data model and their intensions.

Table 3.1: Class of the target data model

Output Classes	Intension
Settlement Class	Area of each of its instance must be equal or greater than 0.01 km <sup>2</sup> (Ordnance Survey, 2005).
Forest Class	Area of each object needs to be equal to or greater than 0.25 km <sup>2</sup> (Ordnance Survey, 2005).
Motorway	A road object of type 'Motorway'.
A Road	A road object of type 'A Road'.
B Road	A road object of type 'B Road'.
Minor Road	A road object of type 'Minor Road'.
Junction	A road object that serves as a link between two or more roads of any of above types where at least one should be of a different type.
Rail Class	Object of type 'Rail'
Hill Class	Object with a prominence of at least 35m and not having any other hills in its extent. The intension of this class was determined via empirical observations.
Range Class	Object with a prominence of at least 35m and should have at least one hill in its extent. The intension of this class was determined via empirical observations.

It is important to point out there are a range of classes such as rivers, lakes, coastlines that are also present at low levels of detail (such as at 1:250,000 scale) but in this research we limited it to the transformation of the source dataset objects into objects belonging to the classes listed in Table 3.1. Further research will look into extending the resultant data model to include these classes as well. In order to maintain geometric partition (space exhaustive with no holes) in the resultant database an additional class is needed called 'General Land' class. All source objects that are not classified as part of any of the classes in Table 3.1 are classified as part of this class. This ensures that the resultant database is thematically and geometrically partitioned.

### 3.8 Populating the Target classes

Once the classes of the target data model have been identified their instances need to be created in order to create a database that is an instance of this new data model. Following the concepts and approaches for database transformation defined previously, we present the proposed approach here. The proposed methodology, for database transformation, consists of four main stages (Figure 3.11). The details of



each stage of the methodology are presented in subsequent chapters. The objective here is to give the reader an overview of the proposed methodology using the concepts and discussion presented so far.

### **3.8.1 Pre-processing: Classification of Source Database**

The first stage consists of classification of the source database objects. This is required in order to group objects that are similar from sets of dissimilar objects. The classification process is required in the selection of ‘typical members’ or ‘good parts’ of the resultant composite classes (further discussed in Chapter 5). This process also helps in the selection and aggregation of source objects into composite objects that are thematically the same in the source and target data model such as Motorways, A Road and Rail class objects. The classification stage of the methodology is explained in detail in Chapter 4.

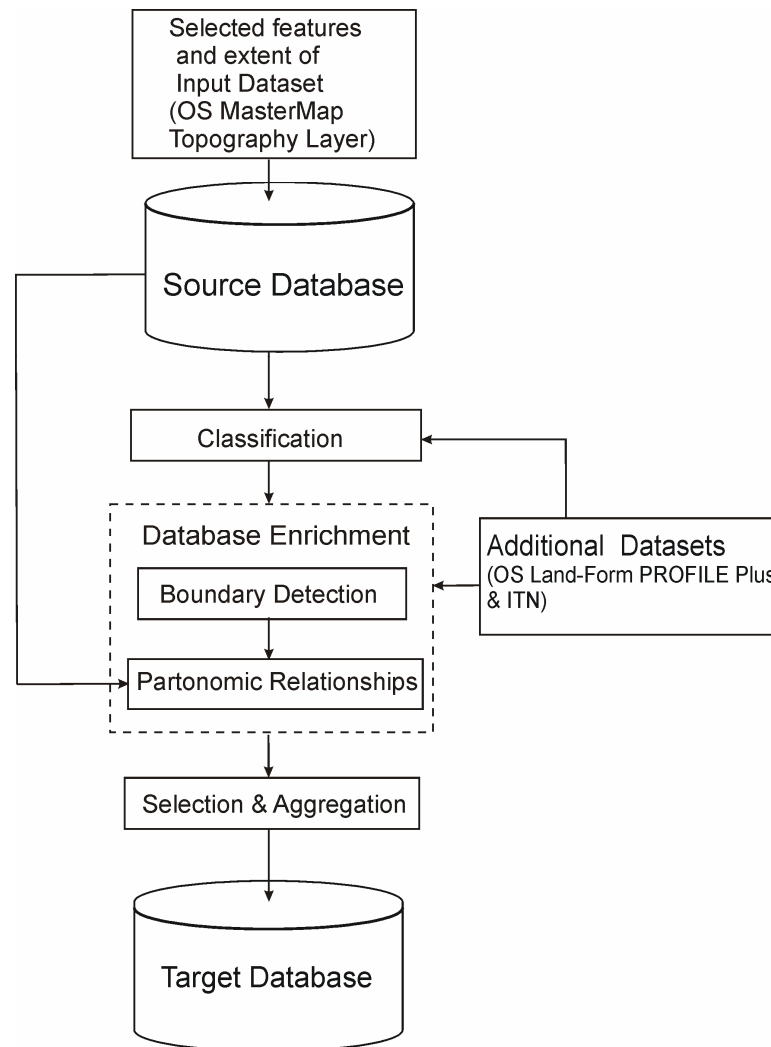


Figure 3.11: Overall methodology

### 3.8.2 Database Enrichment

Several classes in Table 3.1 are composite classes (settlement, forest, hills and ranges). These classes are quite distinct from the classes of the source database objects. The transformation involves determining the component classes of these composite classes. In other words creating the target database from the source database (database transformation) requires the creation of composite objects (instances of *required composite* classes) from the component objects (instances of *source component* classes) in the source database. This requires determination of the partonomic relationships of the source objects in terms of the required composite objects. Such relationships are not explicit in the source database and need to be

made explicit first before transformation can be carried out (Neun *et al.*, 2004). This requires enrichment of the source database.

The database enrichment process presented here in this research was broken down into two sub-stages. The first stage involves determination of the ‘extents’ (or boundaries) of the required composite objects. This process was carried out using specific properties of each of the composite classes independently (Settlement, Forest, Hills and Ranges). The approach developed for each of these classes is explained along with results detailed in Chapter 5.

The resultant boundaries obtained from the approaches in Chapter 5 are called ‘container’ boundaries since these boundaries act as a ‘container’ for the objects in the source database. Using the topological relationships along with area intersection, the partonomic relationships for source objects are determined in terms of the required composite objects. This results in an enriched source database. This stage of the methodology is explained in Chapter 6.

### **3.8.3 Selection and Aggregation**

Once the source database has been enriched in terms of partonomic relationship we can perform database transformations based on selection and aggregation operations. These operations are carried out based on the taxonomic or thematic similarity and the partonomic relationships. A set of rules are required to carry out the aggregation process. These sets of rules ensure topological consistency in the resultant database. These rules and the process of aggregation are presented along with results and the implementation in detail in Chapter 6.

## **3.9 Summary**

The process of generalisation can be categorised into model and cartographic generalisation. Model generalisation focuses on transformation of the spatial database from a detailed spatial model to a more abstract model. This differs from the cartographic generalisation which focuses on the graphic display of the database and is subject to scale constraints. Over the last two decades the importance of the database has increased from a mere storage unit to one that supports spatial and

statistical analysis. This has increased the importance of model generalisation. Moreover model generalisation processes have become critical to the automatic population of MRDBs. This chapter has described critical factors that affect the model generalisation process. The chapter has also explained different strategies of model generalisation (aggregation methods) for thematically and geometrically partitioned datasets, highlighting the importance of functionally driven generalisation that uses partonomic relationships as a basis for object aggregation in order to achieve database transformation. This forms the basis of the proposed methodology. Classes of the target data model were presented along with an overview of the different stages of the proposed methodology. The next three chapters discuss these different stages in detail.

## **CHAPTER 4: Specification and Description of the Source Datasets**

The last chapter presented an overview of different stages of the proposed methodology for transformation of a detailed source database into instances of classes of target data model (Table 3.1). This chapter presents the first stage of the methodology beginning with a description of the datasets used in this research (section 4.1). This section presents the important properties of the primary source dataset and also the additional datasets that were used in the research. Section 4.2 presents the classification stage of the methodology.

### **4.1 Datasets**

The data model specifies the content of the resultant database. However, the quality of the source dataset is very important since it affects the quality of the resultant database. Moreover generalisation operations, such as selection and aggregation, requires that the dataset is properly structured and has enough level of thematic and spatial detail. The proposed aggregation approach also requires that the dataset is geometrically and thematically partitioned.

In this research we have used two datasets; OS MasterMap and Land-Form PROFILE Plus. The important properties of these two datasets are explained in the following sections.

#### **4.1.1 Primary Dataset: OS MasterMap**

The topographic (categorical) dataset used in this research is OS MasterMap. OS MasterMap data forms a complete coverage of Great Britain (GB). It is object oriented and stores data in a seamless form. This means that any set of OS MasterMap is available by area or thematic attributes, not just via fixed tiles of data as in previous generations of Ordnance Survey products (Ordnance Survey, 2007d). It is regularly updated by ground and aerial surveys. The basic units of OS MasterMap data are called features (spatial objects in a database). OS MasterMap features are representations of real-world objects such as buildings, roads, tracks,

paths, railways, rivers, lakes, ponds, structures, and land parcels (Ordnance Survey, 2007d). The data also includes non-topographic features such as administrative and electoral boundaries, cartographic text and symbols, and postal addresses. A complete list of the feature representations is given in the OS MasterMap real-world Object Catalogue (Ordnance Survey, 2007b).

OS MasterMap is divided into four sub datasets (referred as layers by OS). These are Address layer, Imagery layer, Topography layer and Integrated Transport Network (ITN) layer (Ordnance Survey, 2007d). The address layer provides national grid coordinates and a unique reference for approximately 26 million residential and commercial postal addresses in GB (Ordnance Survey, 2007d). It establishes links between features and addresses in other layers of OS MasterMap via a unique feature identifier. The Imagery layer comprises aerial images of GB. The images are orthorectified so that the features in the other layers are aligned with their counterparts in the image. OS MasterMap, with the exception of the Imagery Layer is supplied in TIFF, JPEG, ECW or MrSID format. All other layers are supplied in compressed GML format. Softwares such as ESRI MapManager, Snowflake GOloader can be used for loading gml files into GIS readable formats or for loading into spatial databases. In this research only the Topography and ITN layers were used and are explained in more detail in the next sections.

### **Topography Layer**

The topography layer was the first layer, for OS MasterMap, to be produced by OS, in November 2001 (Ordnance Survey, 2007c). The features in this layer are captured at a high level of detail (1:1250 scale in urban areas, 1:2500 scale in rural areas and 1:10,000 scale in mountain and moorland areas) and are stored in vector format. The features within this layer are features that appear in the landscape, such as buildings, land, water and roads. The topography layer contains around 450 million uniquely identifiable geographic features. Each feature has a geometric representation in the form of a point, line or polygon. Figure 4.1 shows an extract of the Topography layer. In this research, the Topography layer is used as the primary source dataset.



Figure 4.1: Extract of OS MasterMap Topography layer (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved)

Each feature in the Topography layer has a set of thematic attributes. These can be categorised into referencing attributes, life cycle attributes describing the creation and updating dates, feature description, physical description describing the make of the feature, height properties and some attributes that refer to cartographic symbols associated with a feature (Ordnance Survey, 2007c). Attributes that are particularly important to this research are listed in Table 4.1.

Table 4.1: Important attributes of the Topography Layer of OS MasterMap used in this research

Attribute	Description	Examples
Toid	Every feature in the topography layer is identified by a TOID which is unique and acts as an object identifier in the database (Figure 3.1). As well as ensuring unambiguous identification of features, TOIDs enable full integration across the layers of OS MasterMap. For example, features in the address and integrated transport network (ITN) layers are explicitly linked by TOIDs to the respective building and road carriageway features in the Topography layer.	1000000157011914  1000000167901215
Descriptive Group	This is the primary classification attribute of a feature. Each feature has at least one or more values for this attribute. These values are typically categories of real-world topographic objects.	Building Road or Track Railway Building or Structure General Surface Natural Environment
Descriptive Term	This attribute, if present, gives further classification information about the feature. A feature may have multiple descriptive term attributes. A situation where multiple descriptive term attributes are present is where features have a descriptive group with the value of 'Natural Environment'. Such features can have one or more descriptive term attributes specifying the natural land cover types present in the selected region.	(Null) Coniferous Trees Non Coniferous Trees Scrub; Rough Grassland
Area	This is the calculated area of a polygon feature in square metres.	5000.25

### Geometric Properties

Each feature in the Topography layer has one of three geometrical structures – a point, a line or a polygon. Only polygon features are considered in this research. The point and line features represent real world objects such as post boxes, a telephone or electric pole, walls, fences etc (Ordnance Survey, 2007b). Such features were not required for creation of objects in the resultant classes. Removal of



these features does not result in creation of holes in the dataset. Moreover for many line features they have a polygonal representation as well.

The selected polygon features from the source dataset (topography layer) sit adjacent to each other rather than on top of each other. In Figure 4.2, if building or road features are removed they leave holes in the land feature; the land feature does not exist beneath these features. Also these features cover the complete area of interest. Thus they form a geometric partition (i.e. a space exhaustive tessellation of space).



Figure 4.2: Polygon features form a complete geometric partition i.e. there is no overlap between features

#### 4.1.2 Additional Datasets

In addition to the Topography layer (source dataset) two additional datasets were required for classification of the source dataset and database enrichment stages of the methodology. The important properties of these two datasets are now discussed.

##### **Integrated Transport Network (ITN) Layer**

The source dataset (Topography Layer) contains road features but it does not have detailed thematic attributes that can be used for the classification of roads. This classification is needed to select objects of road classes defined in Table 3.1. This classification is also required during the aggregation and selection process in order to distinguish between roads that can be aggregated with other objects from the road objects that are required to be kept at the resultant level of detail. OS MasterMap ITN Layer is used for classification of road objects in the source dataset.

The ITN layer currently contains the road network and road routing information for GB. Future extension of this layer is expected to include rail, water, track and path networks (Ordnance Survey, 2007d). This dataset contains approximately 13 million

road features and 1.5 million items of road routing information. Figure 4.3 shows a selection of the road network from ITN. The features in the ITN data set are topologically structured and the geometry is stored in the form of graph theoretic elements (nodes and segments). The features in ITN have a detailed ‘descriptive term’ that defines the class of each feature. In Figure 4.3 the ITN road features (segments) have been classified using their descriptive attribute value.

The road features in the ITN layer lie within their corresponding road polygon features in the source dataset. This layer also includes a cross reference (or link) table that contains the ‘Toid’ of each ITN feature and the ‘Toid’ of the corresponding road features in the source dataset. This table is used during classification to link the road features in the source dataset with ITN features (discussed in section 4.2)

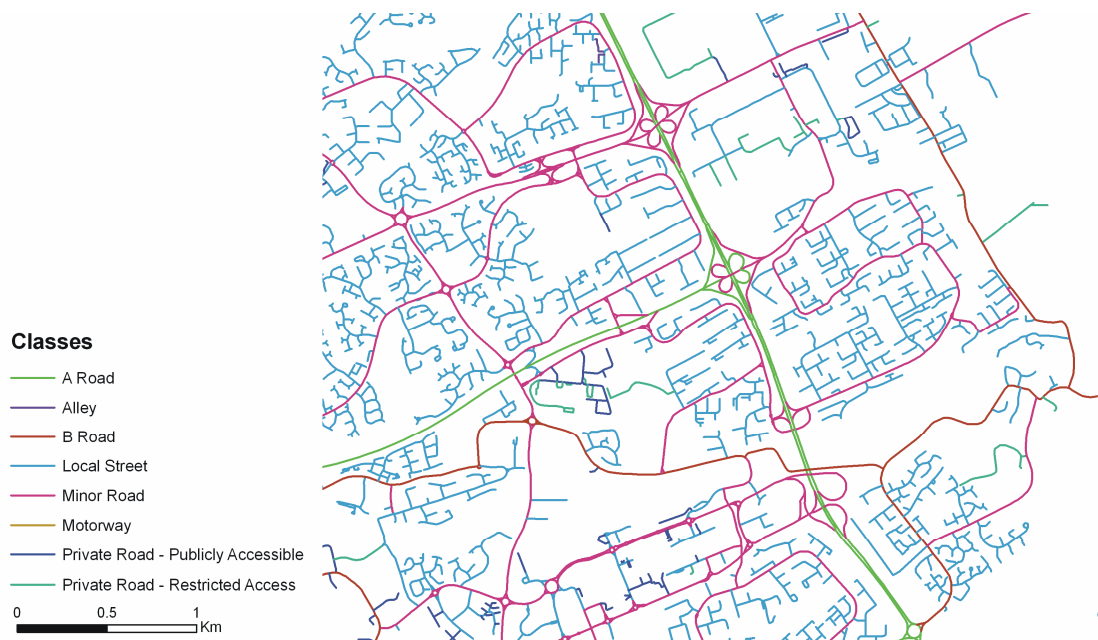


Figure 4.3: Extract of ITN data segments are classified based on their descriptive term (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved)

### Land-Form PROFILE Plus

In addition to the above OS MasterMap datasets, another dataset was required in order to create hills and range class objects in the target database. For the identification of these objects continuous elevation dataset was required. OS Land-Form PROFILE Plus is a digital terrain model. It models the height of ground

points, including those below surface features such as buildings and vegetation (Ordnance Survey, 2007a). The dataset is supplied in 5km by 5km tiles. Land-Form PROFILE Plus has three accuracy and three resolution levels (2m in urban, 5m in rural and 10m in mountainous area). A single tile may have data of more than one accuracy; in which case the resolution is determined by the most accurate data. The height data is structured as a raster dataset where the height values are calculated at the centre of the pixel. This dataset is used in the creation of 'container' boundaries for hill and range class objects (discussed in Chapter 5).

## **4.2 Pre-processing: Classification**

Each feature in the source data set (Topography layer) can have multiple values for descriptive attributes (descriptive group and descriptive term). In order to unambiguously identify features during the process of database transformation the features in the source data set are classified into classes. The classification allows selection and aggregation of objects that are common in source and target data models. The classification is based upon the common values, for descriptive attributes, of features in the source dataset. This classification system ensures that all objects in the source database belong to exactly one class and at most one class (thematic partition of the source dataset).

The source dataset features along with corresponding ITN features, for a region of interest, are loaded into a spatial database (Oracle Spatial 10g). The features loaded in the database are referred to as objects of the source database. Table 4.2 lists the classes for the source database object. The class name is stored as an attribute value for each object (data classes).

Table 4.2: Classes of the source data model

Class Name	Descriptive Group	Descriptive term
Building	Building or Structure or Glasshouse	
Railway	Rail	
Motorway	Road or Track	Motorway
A Road	Road or Track	A Road
B Road	Road or Track	B Road
Minor Road	Road or Track	Minor Road
Junction	Road or Track	Multiple descriptive term values
Other Roads	Road or Track	Alley or Local Street or Pedestrianised Street or Private Road
Path_Roadside	Path or Roadside	
Tree	Natural Environment	Coniferous trees or Non Coniferous trees or Coniferous trees and Non Coniferous trees
Vegetation	Natural Environment	Not (Coniferous trees or Non Coniferous trees or Coniferous trees and Non Coniferous trees)
Water	Tidal Water or Inland_Water or Fore_shore	
Garden Area	General Surface	Multi Surface
Land Cover	All objects that are not member of any of above classes are classified as instance of Land Cover Class	

During the classification of each road object in the source database its descriptive term attribute value is checked from the corresponding road object in the ITN dataset via the cross reference table (Figure 4.4). Road objects that have their descriptive term “Alley”, “Local Street”, “Pedestrianised Street” or “Private Road” are too detailed for the resultant level of detail and are thus classified as instances of “Other Roads Class” (Table 4.2). Whereas road objects having descriptive terms “Motorway” or “A Road” or “B Road” or “Minor Road” are classified into separate classes (Table 4.2). The relationship between ITN objects and road objects in the

source database is many to many (m:n) (Figure 4.4). If there is more than one ITN object with different descriptive term attribute values is related to a source road object. This means that the source road object acts as a connecting object (junction) between two or more road objects (Figure 4.4). Such an object is thus classified as an instance of a 'Junction' class object (Table 4.2). This process is illustrated in Figure 4.4 using a few road objects (in grey polygons) from the source database and corresponding objects in the ITN dataset (in red segments) in Figure 4.4. The resultant classification of the source objects is shown in Figure 4.5.

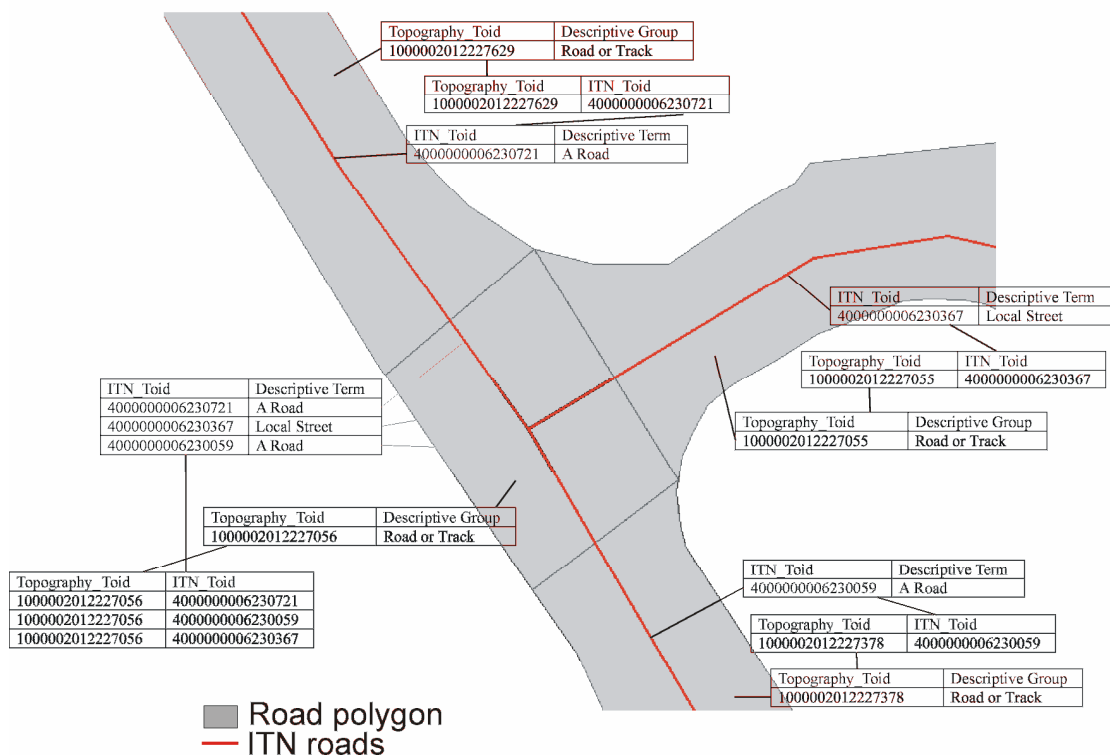


Figure 4.4: Link between ITN roads and road objects in the source database. The descriptive term of ITN roads is used for classification of corresponding road objects in the source database. The objects in the two datasets are linked via cross reference table that contains object identifier (Toids) of both datasets

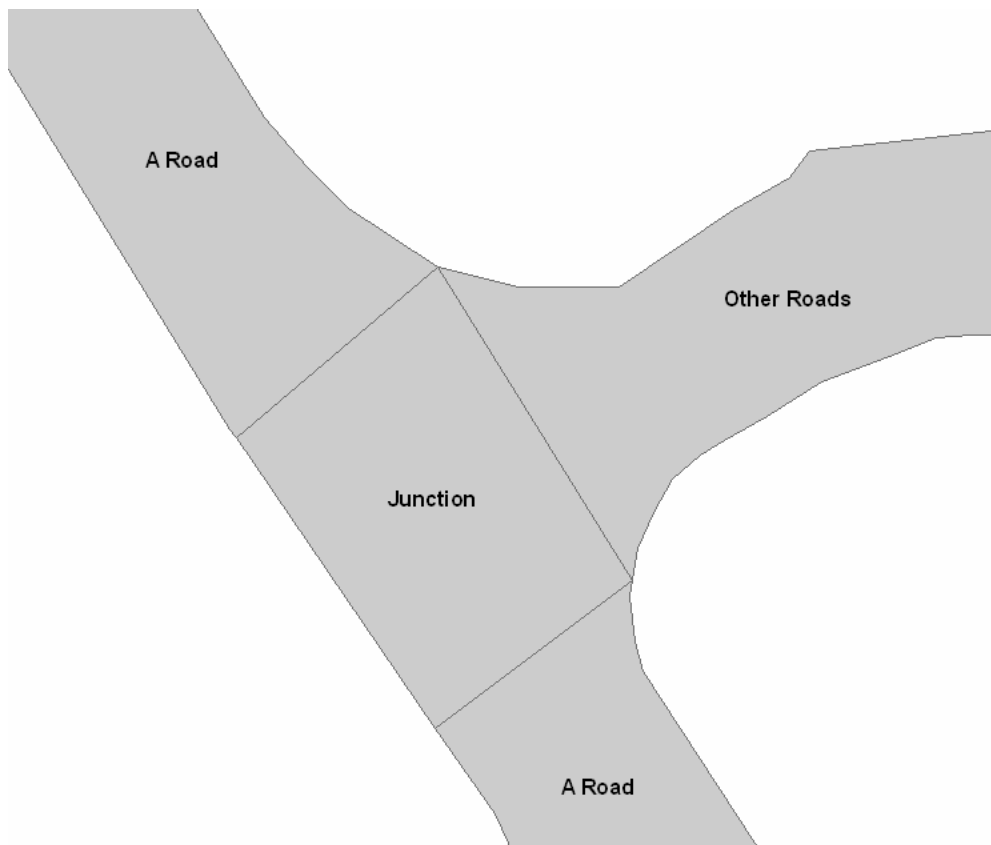


Figure 4.5: Classification of road objects in the source database shown in Figure 4.4 based on the descriptive term values of corresponding ITN roads and classification rules

### 4.3 Summary

The quality of the source dataset has a significant effect on any generalisation process. This includes how detailed the dataset is, how detailed the classification is, and whether the data set is properly geometrically structured. This chapter presented the important thematic and geometric properties of the three datasets used in this research. Out of these three, the OS MasterMap Topography Layer is the source dataset whereas the ITN and Land-Form PROFILE Plus are additional datasets that are required for classification and database enrichment of the source dataset. The polygon features of the source dataset are used in this research. These features, for a given region of interest, are selected from the source dataset and are loaded into a spatial database. Each object in the database is classified into a specific class based on the values of its thematic attributes. The resultant data forms form a thematic and geometric partition of the region of interest.

## CHAPTER 5: Identification of 'Container' Boundaries

As discussed in previous chapters, the creation of objects of the required composite classes (settlement, forest, hills and range) from the source database objects requires database enrichment in terms of partonomic relationships. Enriching the source database with these relationships requires that we determine the 'container' boundaries for the target composite objects. This chapter presents the second stage of the proposed methodology i.e. container boundary detection. The chapter starts with a discussion on types of boundaries of geographic phenomena (section 5.1). The subsequent sections (5.2, 5.3 and 5.4) then describe the methodology developed for each container type boundary. The term 'container' is intentionally used to separate it from the idea of the boundary of a resultant composite object created via aggregation (Chapter 6). A few results by the application of each methodology are presented in each section.

### 5.1 Modelling Boundaries

Boundary is a concept inseparable from that of spatial objects (Bian, 2007; Mark *et al.*, 1999). The boundary that separates the entity from its environment, and is one of the marks of its individuality (Casati *et al.*, 1998). But the degree to which we can precisely define the boundary of a geographic object varies enormously (Burrough & Frank, 1996; Campari, 1996). This observation is reflected in research on the modelling of fuzzy boundaries (Burrough & Frank, 1996; Clementini & Felice, 1996; Cohn & Gotts, 1996). Certain spatial entities tend to have well-bounded boundaries, for example, instances of buildings, streets or a lake. But there are certain geographic features that are continuous and do not have well bounded boundaries but which can still be perceived as objects. For example, a settlement or a hill (Bian, 2007; Smith & Mark, 2001). A systematic treatment of boundaries has been attempted by Smith (1995), who argues that boundaries can be divided into two basic types: bona-fide boundaries and fiat-boundaries (Smith & Varzi, 1997; Smith & Varzi, 2000). A boundary that is 'bona fide' is one that is a 'thing in itself' and exists even in the absence of all delineating or conceptualizing activity (buildings, river-

banks or coastlines are examples of bona fide boundaries). In that sense they are boundaries which exist independently of all human cognitive acts and 'are a matter of qualitative differentiations or discontinuities in the underlying reality' (Smith 1995, p476). The other type of boundary is a 'fiat' boundary in which the boundary owes its existence to acts of human decision or decree, in some way related to human cognitive phenomena. Thus 'fiat boundaries are boundaries which exist only by virtue of the different sorts of demarcations effected cognitively by human beings' (Smith 1995, p476). They are delineations which correspond to no genuine heterogeneity on the side of the bounded entities themselves. Examples would include political borders, property-lines and administrative boundaries. Geographic boundaries that are estimated by some definition or mathematical function also fall under the category of fiat boundaries (Bian, 2007). This dichotomy of boundaries is scale dependent as well. Fiat boundaries are related to spatial objects present at lower levels of detail whereas the bonafide boundaries represent spatial objects at finer scale (Bittner & Geoffrey, 2001). The required composite classes, in this research, are fiat types and their boundaries are fiat boundaries. The following sections present approaches for determination of fiat (container) boundaries for each composite class. Since each of the composite classes are quite distinct from each other, different methodologies are presented for determination of container boundaries for each class. These approaches are presented in subsequent sections.

## **5.2 Methodology for Settlement Container Boundary**

There are many ways by which we might define settlements. They may be large and densely populated regions, often having special administrative, legal, or historical status. Generally speaking large settlements (cities) are places of trade, in which benefits arise through reduced transportation costs, and the sharing of natural resources (Fujita *et al.*, 1999) – places with 'physical, social, economic and cultural dimensions' (Esnard & Yang, 2002). There is no standard international definition of a city or urban settlement and many of the administrative definitions of city are rather circular in their definition (Angel *et al.*, 2005; Heikkila *et al.*, 2003; Small *et al.*, 2005). Yet having systematic measurements of the extent of a city is important in comparator studies, in the measurement of change over time, in governance and



urban planning, environmental impact assessment and in mapping and land classification.

Various measures and techniques have been proposed for measuring characteristics of settlement (such as sprawl, compactness, contiguity) (Crouch, 2006). These include modelling the size and number of districts and wards, and the use of remote sensing and classification techniques. Examples include analysing anthropogenic light (Small *et al.*, 2005) and classifying Landsat imagery (Heikkila *et al.*, 2003; Mesev *et al.*, 1995). Population density is commonly used (Bulger & Hunt, 1991) though research has shown there to be disparity between population growth and land consumption (Esnard & Yang, 2002). Similar to measures of population density, what is proposed in this research is a definition of the settlement extent based on the area and density of buildings (Ruas & Mackaness, 1997) – the assumption being that the built environment reflects a set of social and economic activities that define ‘settlement’. Building objects are typical members or good parts of a settlement object and can be thus used here to define the extent of a settlement.

Initial ideas in the creation of a settlement container boundary focused on distance-based clustering of building objects (section 5.2.1). This proved to be limiting but led to an approach in which the density of building objects was modelled (section 5.2.2). Section 5.2.3 illustrates how the settlement container boundaries were formed using the density values. The methodology is summarised in Figure 5.1. Section 5.2.4 presents a few results for different input regions of interest selected from the source dataset, obtained from the application of this methodology.

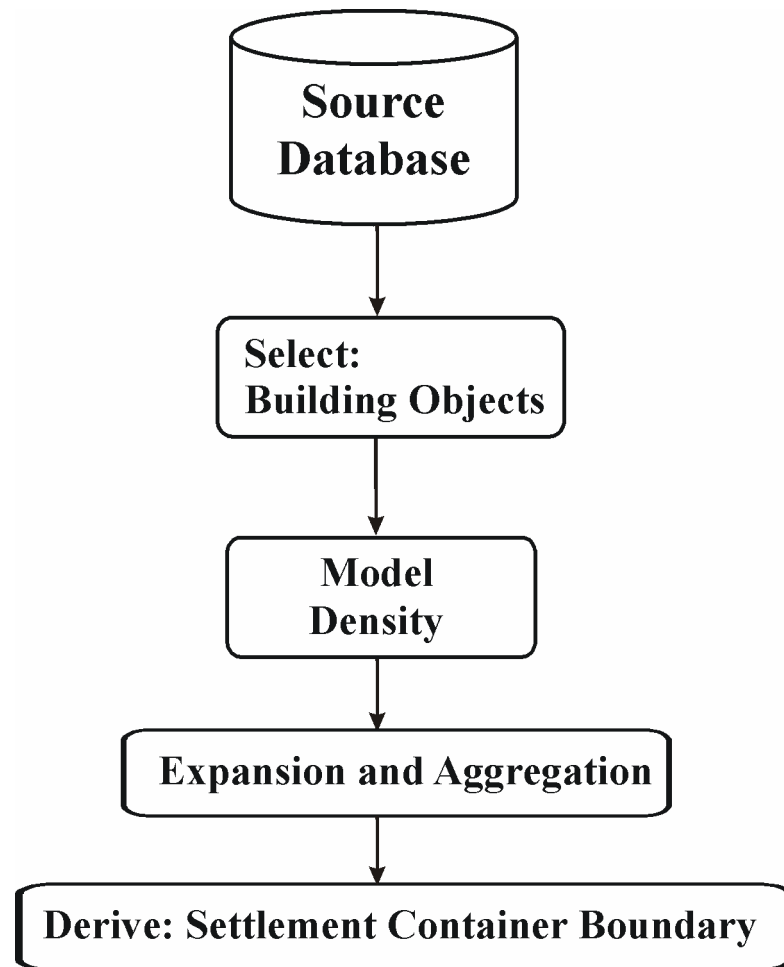


Figure 5.1: The major stages of methodology used for determination of settlement container boundaries

### 5.2.1 A Distance based Clustering Approach

From our empirical evaluation of existing settlement boundaries at 1:250 000 scale (Strategi dataset) it became clear that settlement boundaries were created ‘around regions with a high concentration of buildings’ (as stated in 1:250,000 data set specification (Ordnance Survey, 2005)). Our initial approach was to cluster the objects based on a (fixed) distance threshold and create boundaries around each cluster. This is illustrated in Figure 5.3 (Chaudhry & Mackaness, 2005b). This approach resulted in small sparse building groups at the city periphery (‘noise’ or ‘outliers’ –Figure 5.2 and Figure 5.3) which tended to make the boundaries overly large and resulted in the stringing together of different cities or towns. Additionally, it proved inadequate since it could not separate low density areas from high density areas. A possible solution was to use a distance threshold of varying size. But this

makes the threshold specific to the area selected and it was difficult to determine a value that is applicable to all areas. It was therefore necessary to devise a method that could group buildings in a way that took into account their local concentration.



Figure 5.2: Typical objects (buildings) selected from the source database. Three low dense areas are ringed (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved)



Figure 5.3: The output boundaries generated by distance based clustering algorithm using a 100m threshold. The low dense areas ringed in Figure 5.2 have become part of the output boundaries

### 5.2.2 Calculation of Citiness

As a solution to this problem (Figure 5.3) we formulated an equation that ascribes a value to each building in terms of its areal footprint, the total area of the buildings in its neighbourhood and the sum of the distances from that building to all the buildings within that neighbourhood. This gives us a value of ‘citiness’ denoted by  $c$  (Equation 5.1).

$$c_j = \frac{\sqrt{a_j} \sqrt{\sum_{\substack{i=1 \\ i \neq j}}^n a_i}}{\sum_{i=1}^n d_{ij}^2} \quad \text{Equation 5.1}$$

In Equation 5.1 we calculate  $c_j$ — the citiness value for building  $j$ . Where  $a_j$  is the area of the building  $j$ ,  $a_i$  is the area of building  $i$  and  $d_i$  is the distance of building  $i$  from building  $j$ . The denominator acts as a decay function such that the citiness value  $c_j$  will be high if the building  $j$  is located in a dense neighbourhood and will be

low if it is at the periphery or in a low density area. It is not necessary (nor desirable) to calculate the value of  $c$  by taking into account all the buildings in the database. The neighbourhood is not defined as a fixed radius from building  $j$ , but as a count of the  $n$  closest buildings. Where  $n$  is small, localised dense regions are identified. Conversely where  $n$  is large, a generalised view is created. Through empirical testing it was found to be sufficient to consider the fifty closest neighbouring buildings. Figure 5.4 shows the normalised output surface created based on the values of citiness calculated for each building shown in Figure 5.2. These values of citiness were then used to create a settlement container boundary - a boundary around the settlement used to identify all the objects that are part of the settlement (section 5.2.3).

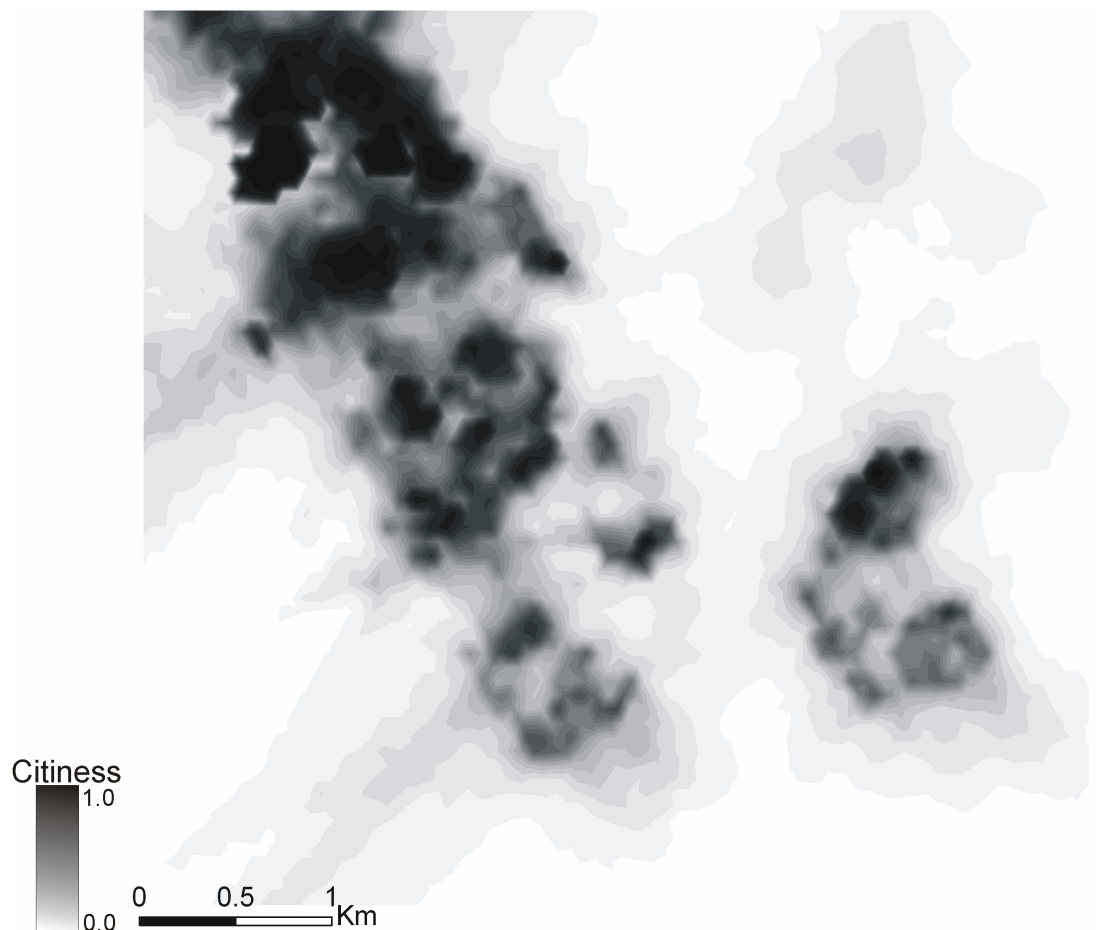


Figure 5.4: Interpolated and normalised surface based on citiness values (using equation 1) for buildings in Figure 5.2. The three prominent settlements are 'Livingston', 'Mid Calder' and 'East Calder' in Scotland

### 5.2.3 Creation of the Boundary of the Settlement Container

From the surface (Figure 5.4) we wish to identify a discrete region that is deemed to 'contain' the settlement. This is done by expanding and then aggregating the resulting overlapping building objects. The areal extent of each building is expanded by the value  $e$  according to the value of  $c$ . In this manner regions are aggregated according to their density and proximity. The amount of expansion is calculated using Equation 5.2.

$$e_a = k \cdot c_a \quad \text{provided } e_a \leq k \quad \text{Equation 5.2}$$

Where  $e_a$  is the amount of expansion for building  $a$ ,  $k$  is a constant determined empirically and is the upper limit of expansion and  $c_a$  is the citiness value of building  $a$ . Large changes in the level of detail require larger value of  $k$ . This idea of expansion is illustrated in Figure 5.5. In Figure 5.5a objects are grey scaled according to their value of  $c$ . These are used to expand the objects according to Equation 5.2 resulting in Figure 5.5b. Where buildings overlap, they are aggregated into one boundary polygon (Figure 5.5c). After aggregation the next step is the selection or elimination of the resultant aggregated object. Here we use the area of the resultant object as a basis for selection. Using the intension of settlement class (Table 3.1) the area of each resultant boundary object has to be equal to or greater than  $0.01 \text{ km}^2$ . Thus small boundary objects and small open spaces within the boundary objects are removed or absorbed into their containing object (Figure 5.5d). The idea of elimination of holes and small boundary polygons is based on the principle of generalisation that it should lead to elimination rather than addition of detail (Müller & Wang, 1992).

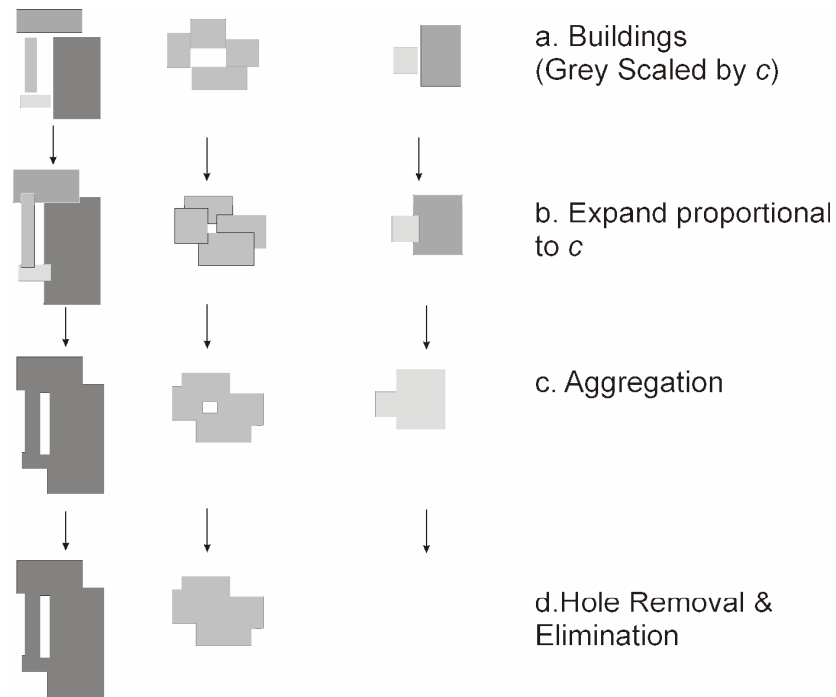


Figure 5.5: Process of expansion, aggregation and elimination for three examples (a). (b) expansion of objects proportional to their citiness value. (c) Aggregation of overlapping objects. (d) Elimination of small boundary objects and holes

The resultant boundary objects for the data in Figure 5.2 is illustrated in Figure 5.6. Note that the low density objects ringed in Figure 5.2 are no longer part of the resultant boundary (Figure 5.6).

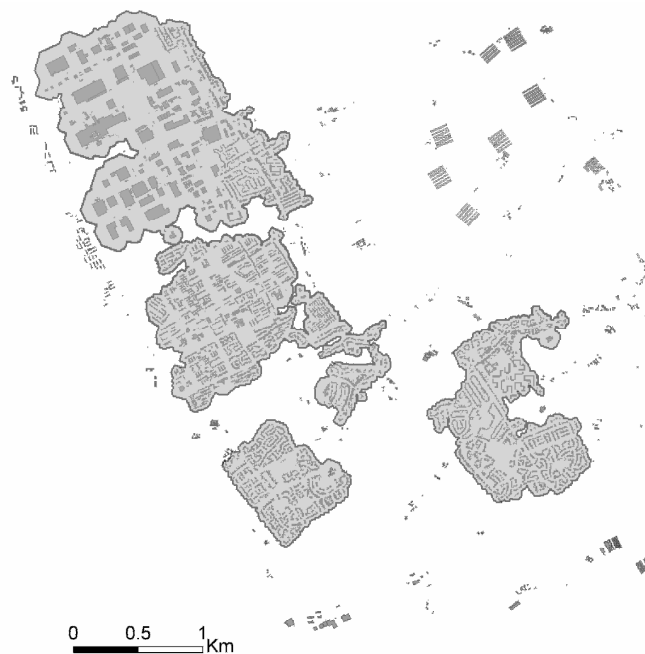


Figure 5.6: Resultant settlement container boundaries generated by the algorithm superimposed on the input buildings (Figure 5.2)

#### 5.2.4 Implementation and Results

The platform selected for the implementation of this methodology was Java, SQLJ and Oracle 10g. Results of the approach on areas around Livingston in Scotland (Figure 5.2) are presented in Figure 5.6. Figure 5.7, 5.8 and 5.9 show settlement container boundaries for three different areas within Great Britain (GB) selected from the source database. It is important to point out that all the parameter settings are kept the same in each instance. Results for a few more regions and datasets are presented in Appendix III.



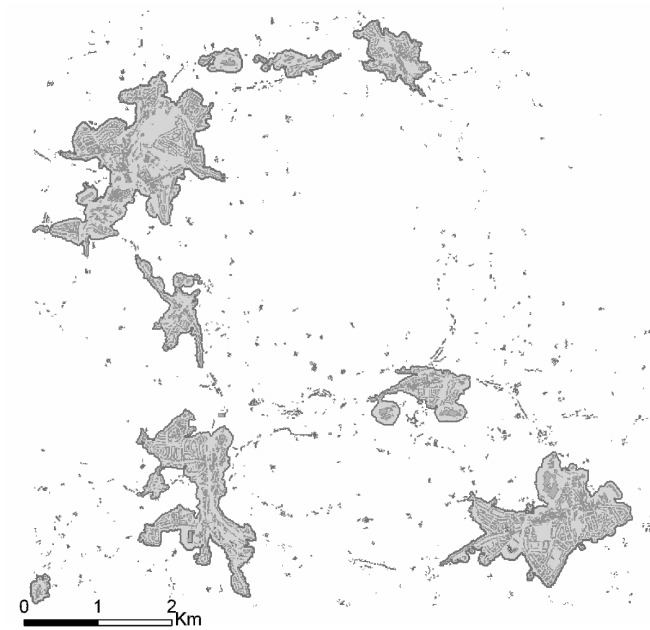


Figure 5.7: Resultant boundary generated by the algorithm with input buildings selected from the source database. The region is in the Peak District in England

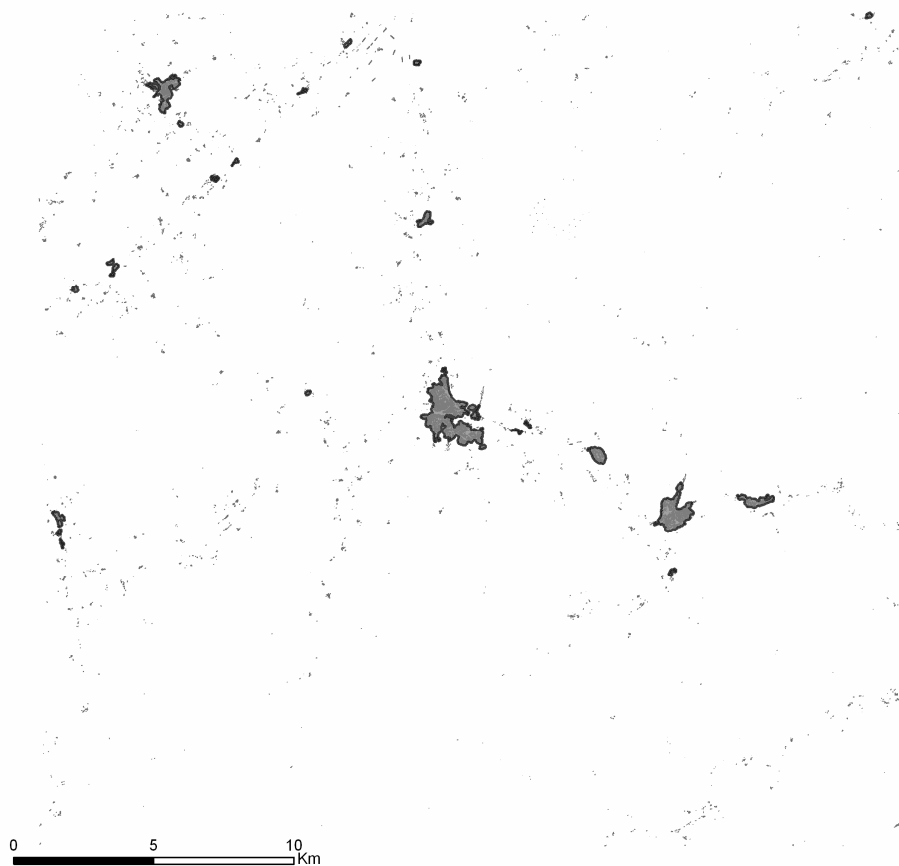


Figure 5.8: Resultant boundary generated by the algorithm with input buildings selected from the source database. The region is around the town of Peebles in Scotland

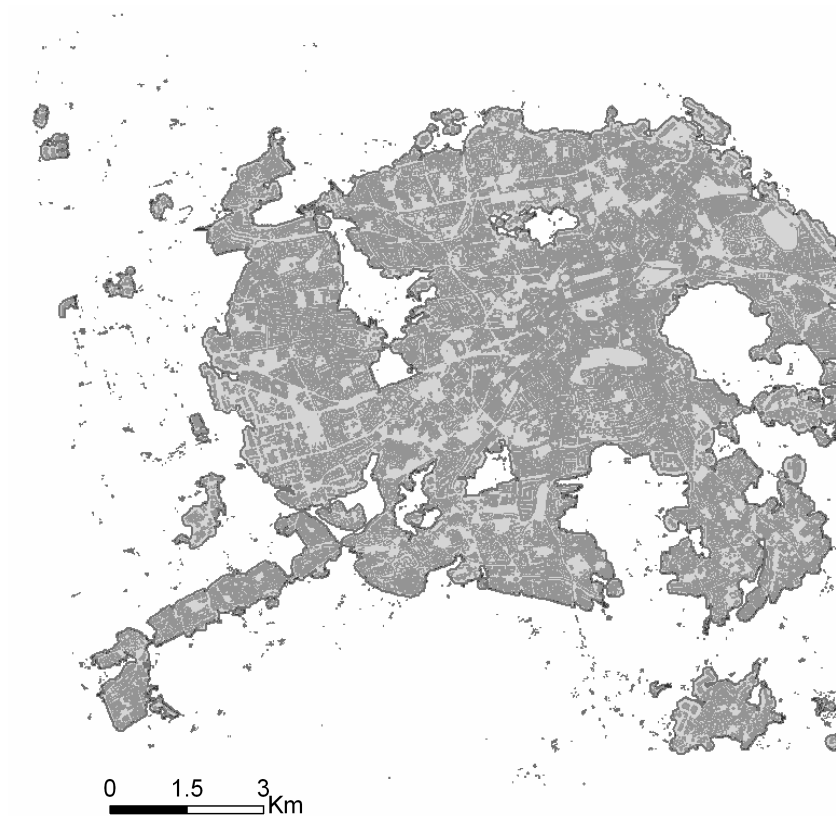


Figure 5.9: Resultant boundary generated by the algorithm with input buildings selected from the source database. The input dataset consists of approximately 130,000 buildings. The region is Edinburgh city and its surroundings in Scotland

### 5.3 Methodology for Forest Container Boundaries

Forest or woodland generalisation has tended not to receive much attention (notable exceptions being (Gold, 1998; Gold *et al.*, 1996; Revell, 2005)). Gold (1998) developed a digitizing technique for rapid capture of forest polygons based on digitizing around the interiors of each polygon and creating a Voronoi diagram, and extracting the Voronoi boundaries between points (Gold *et al.*, 1996; Gold 1998). More recently Revell (2005) outlined a technique for the generalisation of tree polygons using OS MasterMap Topography Layer for representation at 1: 50,000 scale. This approach involved clustering of tree polygons into independent groups, amalgamation of these clusters into single woodland and simplification of the output boundary.

This section presents an approach developed for creating a forest container boundary at required levels of detail. Just as in the case of settlement here too the typical or

good parts i.e. tree objects were used for the creation of these boundaries. But unlike buildings, where each object is a representation of a real world entity (i.e. building), a tree object in the source database does not represent an individual tree but a collection or group of trees. Moreover unlike the building objects, tree objects in the source database are delineated by other objects such as roads, rivers and railways, that run across these objects. The tree objects also have much larger footprints as compared to building objects. In the case of settlement boundaries it was the combined effect of groups of buildings that were used to create the boundaries. But in the case of the trees due to their much larger footprint the primary selection criteria was their area and dense neighbourhood is used as secondary criteria. The initial work drew upon the work of Müller and Wang (1992) who developed an algorithm to generalise groups of lakes (section 5.3.1). This proved to be limiting for forest boundaries but led to an approach in which density and area of tree patches were modelled to generate expansion or contraction values (section 5.3.2). The methodology is summarised in Figure 5.10. Section 5.3.3 presents a few results for different regions obtained from the application of this methodology.

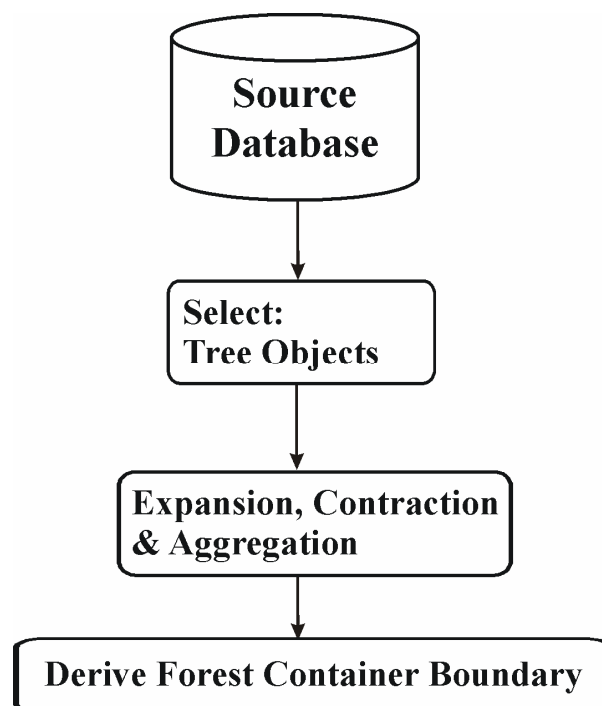


Figure 5.10: Methodology for creation of forest container boundaries

### 5.3.1 Area Patch Methodology

Müller and Wang (1992) proposed an ‘area patch’ algorithm to generalise groups of lakes. The methodology proposed by them in essence is based on clustering in which large objects are expanded and retained whereas the smaller objects are contracted and eliminated. This methodology is analogous to the idea that ‘the rich get richer, and the poor get poorer’. The main stages of this methodology are summarised in Figure 5.11. Here area patches that have area larger than threshold are expanded and those below are contracted (Figure 5.11b). In the next stage the overlapping area patches are aggregated (Figure 5.11c). The resultant objects are selected based on the selection criteria (minimum area) which is dependent on the resultant level of detail.

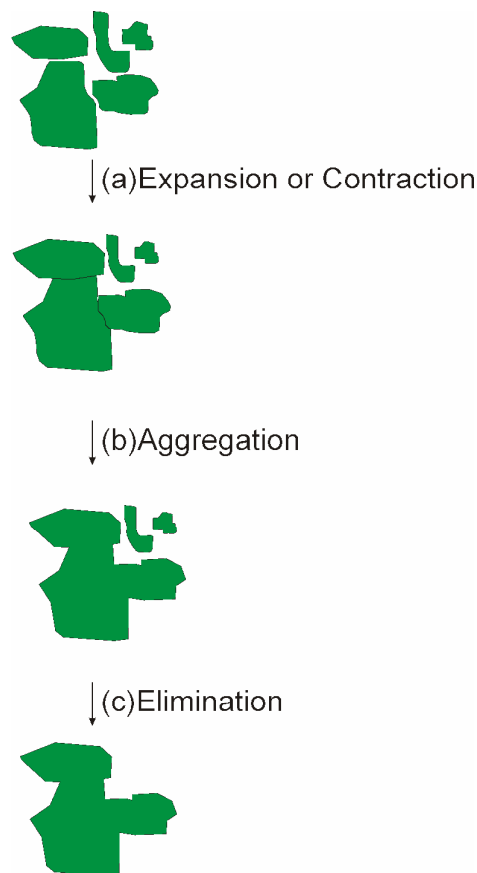


Figure 5.11: Main stages of area patch methodology

This methodology was re-implemented and applied to tree objects. An example of output generated using this methodology for tree objects is presented in Figure 5.12.

The variability in size, and pattern of distribution of tree objects (being ‘broken by network structures’) meant that the above area patch methodology produced results that did not correlate with the input dataset (Figure 5.12). It was not possible to set a single parameter that accommodated the variance in size among the tree objects. Importantly small objects within dense areas were not selected for expansion and were eliminated. Though the proposed approach retains the idea that “the rich get richer and the poor get poorer” it was extended to take into account the neighbourhood density of each tree object. The different stages of the methodology are presented in the next section.

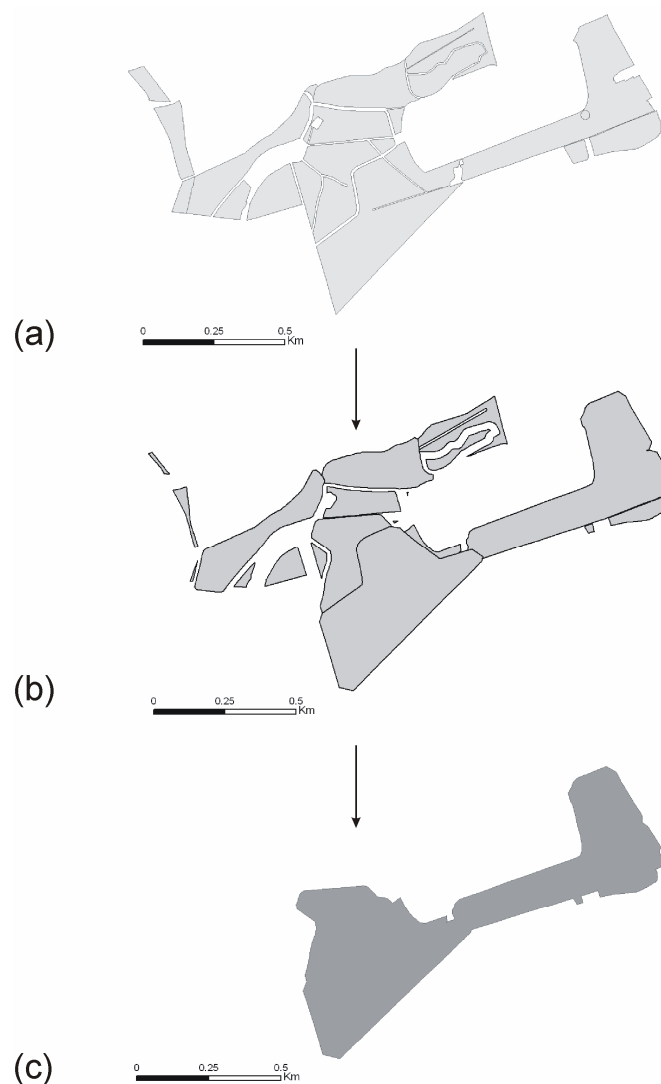


Figure 5.12: Application of Müller and Wang (1992) area patch methodology on tree objects (a). Result of expansion and contraction (b). Resultant forest boundary object (c)

### 5.3.2 Extension to Area Patch Methodology

All objects that are instances of the tree class (typical members) are selected from the source database. These objects are then compared against an area threshold (FT). Any object larger than FT was expanded by a blanket width defined by Equation 5.3 (Müller & Wang, 1992) and any patch smaller than T was either expanded or contracted by the blanket size depending upon its density (Equation 5.3). The value of threshold FT was determined empirically by overlaying the source database tree objects from the source database with forest objects at 1:250,000 scale dataset (Strategi) and identifying the area at which patches began to be eliminated. The algorithm was sufficiently robust that once set the threshold did not need to be changed for other regions of the dataset. Small changes in scale require lower value of FT (i.e. more patches are allowed to expand) whereas large changes in detail require large value of FT. Here in this research for all study regions FT was determined empirically and was set to 35000m<sup>2</sup>.

$$\text{Blanket Width: } t_i = (c_i)(K)/\sqrt{|a_i - FT|} \quad \text{Equation 5.3}$$

Where  $t_i$  is blanket width used for expansion or contraction of tree object  $i$

$K$  is the constant for scaling the blanket width

$$K = \hat{t} / \sqrt{(Max - FT)} \quad \text{Equation 5.4}$$

$\hat{t}$  = constant (the larger the changes in the level of detail, the larger the constant should be)

$Max$  = Maximum area of a tree object in the input dataset

$FT$  = Threshold for expansion or contraction

$c_i$  = Compactness of objects  $i$

$$c_i = a_i / (p_i / 4\pi) \quad \text{Equation 5.5}$$

$a_i$  = Area of object  $i$

$p_i$  = Perimeter of object  $i$

In order to avoid the problem illustrated in Figure 5.12 tree objects whose area was less than FT were subject to a ‘density check’. For each such object the summation of area of all tree objects that are within a given distance (set to 50m) of that object was calculated. This summation area illustrates how dense is the object’s neighbourhood and is compared against the threshold, FT. If greater than FT the object is part of a dense neighbourhood and is therefore expanded (Equation 5.3). On the other hand if the summation area is below the threshold then the object is contracted by a blanket size calculated using Equation 5.3.

After expansion and contraction of all objects the overlapping objects are aggregated into boundary objects. The next step is the selection (or elimination) of the resultant boundary objects. Using the intension for forest class (Table 3.1) the area of each resultant boundary has to be equal to or greater than 0.25 km<sup>2</sup> (Ordnance Survey, 2005). Small boundary objects are removed and small open spaces within these objects are absorbed into their containing boundary. The resultant objects are the required forest container boundary objects. An example is given in Figure 5.13. Note that in Figure 5.13 the resultant boundary includes the objects that were not included in Figure 5.12. The next section presents a few outputs for different regions in GB selected from the source database.

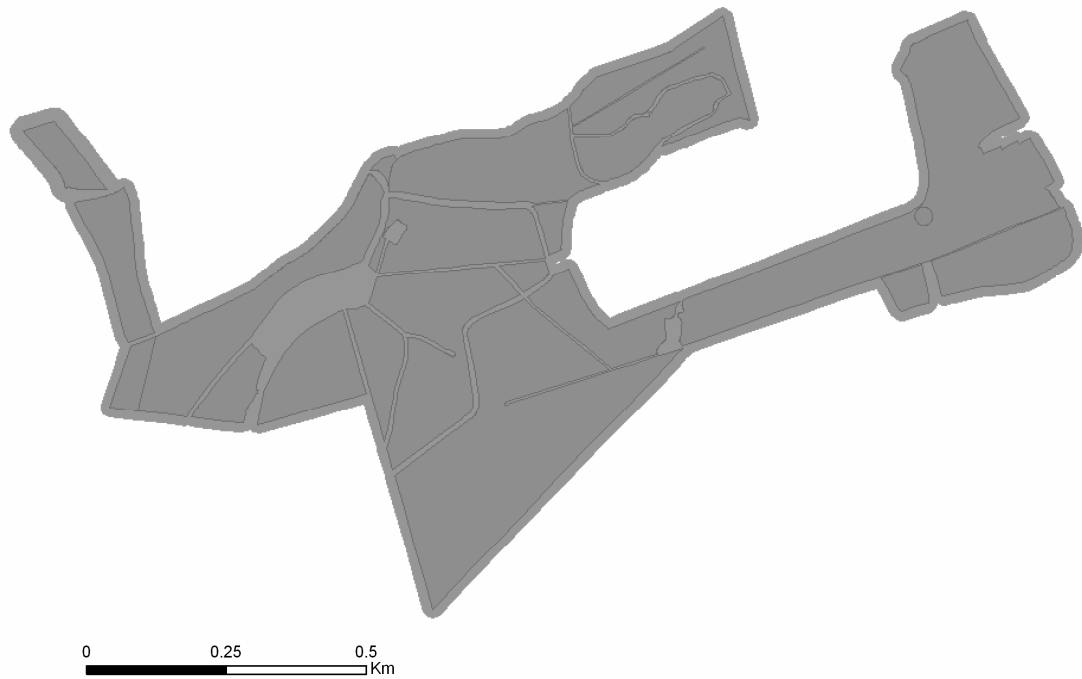


Figure 5.13: Forest container boundary (dark grey) overlaid over original tree objects (light grey) selected from source database. The region is south of the town of 'East Calder' in Scotland

### 5.3.3 Implementation and Results

The platform selected for the implementation of this methodology was Java, SQLJ and Oracle 10g. Figure 5.14 and Figure 5.15 show forest container boundaries for two different areas within GB selected from the source database. These regions were selected because of the variation of tree objects, in terms of area, density and compactness. It is important to point out that all the parameter settings are the same in each case.





Figure 5.14: Output forest container boundaries (dark grey) overlaid on source database tree objects (light grey). The region is Livingston and its surrounding in Scotland

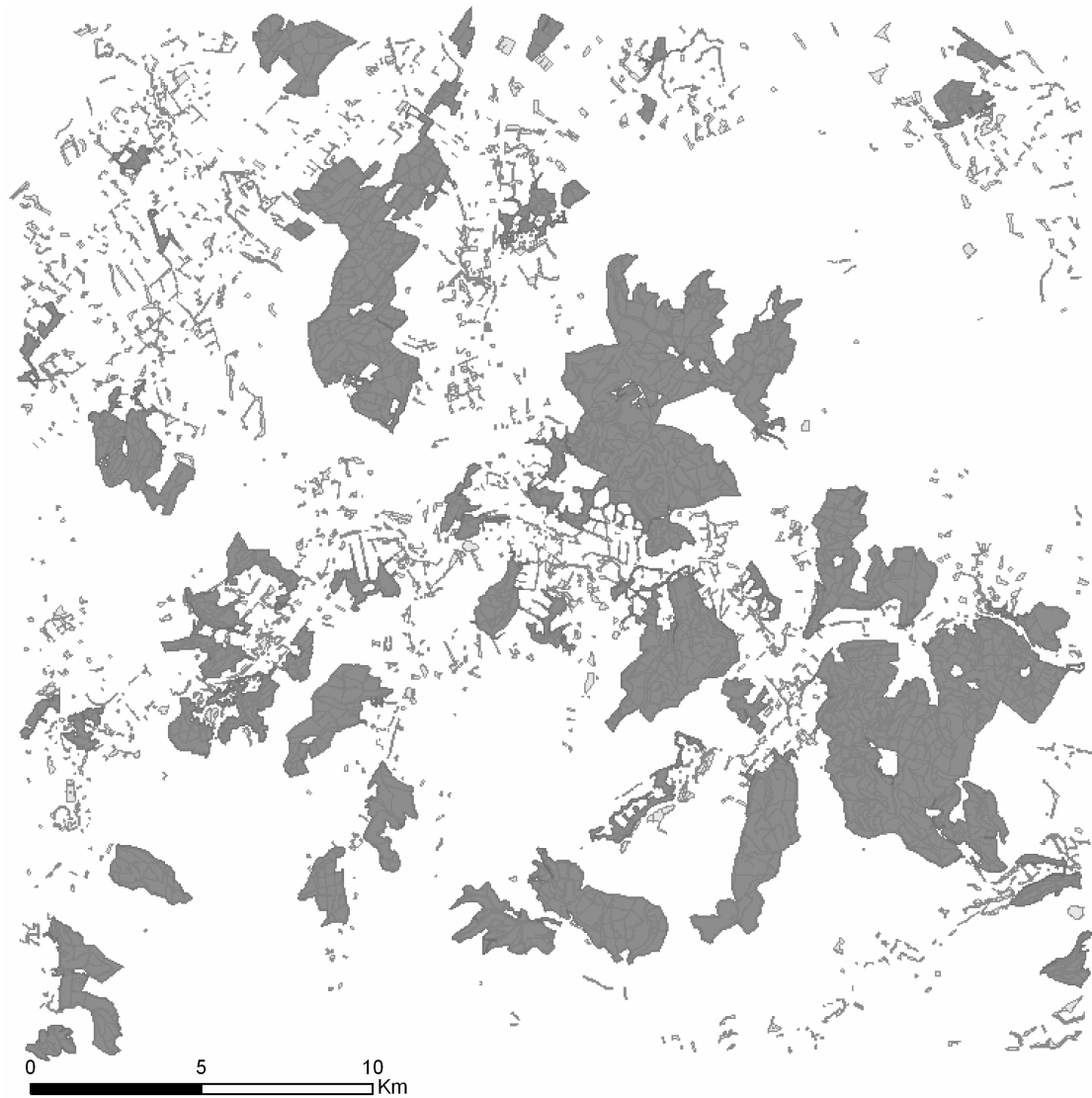


Figure 5.15: Output forest container boundaries (dark grey) overlaid on source database tree objects (light grey). The region is Pebbles and its surroundings in Scotland

#### 5.4 Methodology for Hill and Range Container Boundaries

This section of the chapter focuses on the derivation of the fiat container boundaries for hills and ranges. Many attempts have been made to mathematically define (and thus automatically identify) different types of landform features. What constitutes a hill; a mountain or a range is a very scale-dependent issue. The person asking the question may have a very vague prototypical view of a particular region (Kuhn, 2001) and that view will alter depending on the context. There again, someone may have a precise mathematical definition that, in the context of a spatial query, returns

a definitive answer. Many researchers have arrived at different definitions of what a mountain or hill is (Bonsall, 1974; Campbell, 1992; Cohen, 1979; Purchase, 1997; Usery, 1996) – the definitions often reflecting localised understandings of the landscape (for example, that the notion of a mountain in Scotland is very different when viewed in a Himalayan context). One example of an attempt to define the mountains of Scotland is reflected in the ‘Munros’ of Scotland, named after Sir Munro who compiled and published in 1891 a list of all mountains over 3000ft in Scotland (he identified 277 separate mountains). He did not define how prominent the mountain should be, only that there be ‘sufficient separation’ from neighbouring tops ([www.smc.org.uk](http://www.smc.org.uk)). The subjective definition of what constitutes a Munro is reflected in revision to the list in 1995 resulting in something that was defined as a ‘Murdo’ (Scottish hills at least 3000 feet in height with a drop of at least 30 metres on all sides) of which there are 444 (Dawson, 1995).

At its simplest we might use absolute height to define a hill or a mountain. But caricature (important to cartography and our conceptual grouping of things) has much to do with observable difference and being able to differentiate between prototypical views of things. For example each of us has a conceptual understanding of plateau, delta or mountain and our labelling of these features reflects a shared agreement and understanding. Prominence (the amount by which a hill rises above the local area) clearly influences people’s perception of whether something deserves the epithet ‘hill’. Additionally its morphological variability as compared with its surroundings is also a critical factor (Fisher & Wood, 1998). The morphological variability can be measured in terms of the frequency of peaks, passes and ridges, and additionally in terms of its pits, channels and planes (Fisher *et al.*, 2004). These descriptors are useful in modelling variability and can thus help characterise a region. The methodology proposed here reflects these two essential ingredients *prominence* and *morphological variance*. These were derived from a generalised digital terrain model (DTM), and combined to create bounded regions that demarcated the individual hills. This provided the basis by which the hills and ranges could be defined. The end result is a morphologically nested description of the region (Figure

5.16) which acts as a framework for model generalisation and spatial query. In the following sections we will present different stages of the approach in more detail.

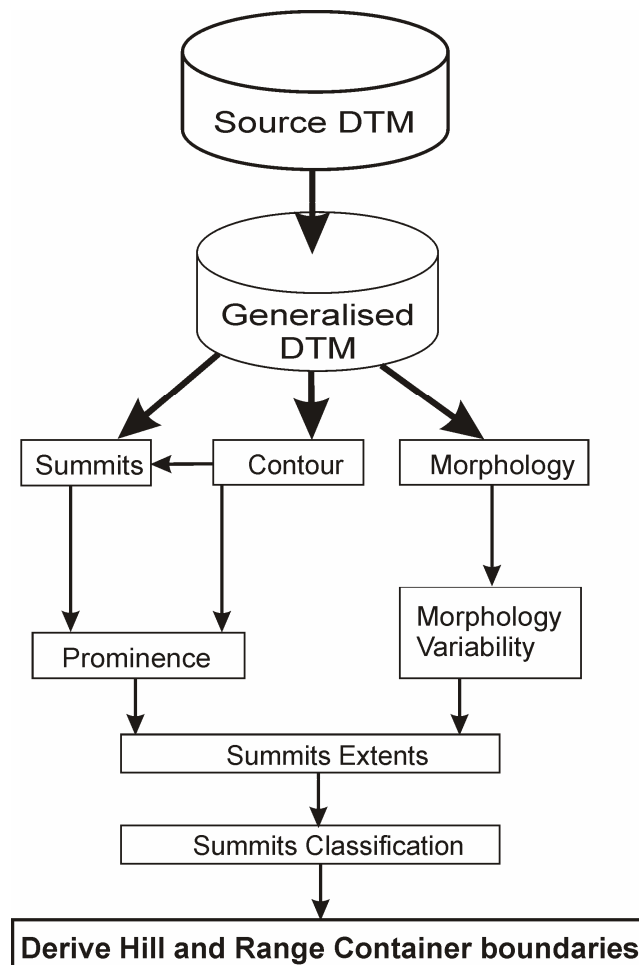


Figure 5.16: The overall method by which hills and ranges are identified

#### 5.4.1 Calculating the prominence of a Summit

In topography, prominence, may be referred to in terms of ‘relative height’, ‘shoulder drop’ (in America) or ‘prime factor’ (in Europe), or simply ‘relief’ (Press & Siever, 1982; Summerfield, 1991) and is a measure of the independent stature of a summit. There are different methods for calculating prominence. Here it is defined as the elevation difference between the summit and the lowest closed contour that encircles that summit but which contains no summit(s) of higher elevation than itself (reflecting the idea of elevation difference with respect to the surroundings). This lowest contour that encircles the summit and no higher summit is called the *key contour* of the summit. To be sure that we identify the correct key contour we must

consider an extended area of the DTM that includes the continent that the summit resides within. For a summit such as Ben Nevis that extent should in theory incorporate the entire DTM of GB. From a computational perspective it is not practical to adopt this approach. Instead it is appropriate to choose a “sufficiently” large extent such that the region of interest lies well within that extent. By way of an example, a “sufficient” extent for Ben Nevis might be a centred square 50km by 25km and for the Pyrenees that extent might be a rectangle of 600km by 300km. The main point is that the results for any given region are meaningful within the outermost contour that is “closed” within the selected extent. Once the key contour for each summit has been identified, prominence is then calculated as the elevation difference between the elevation of the key contour and the elevation of the summit.

The prominence for each summit was calculated by firstly creating contours from the source digital terrain model (DTM) (Land-Form PROFILE Plus). The contours created using ArcGIS or Landserf (URL: <http://www.soi.city.ac.uk/~jwo/landserf/landserf220/>) from the source DTM were not appropriate for processing because ‘spikes’ were present around the edges of some contours or contours were attached to other contours or were broken (Figure 5.17a and Figure 5.17b). To avoid these problems, the input DTM was filtered using a smoothing algorithm (Wood, 1996b) with a given kernel size (set to 25 cells (25\*25)). The resultant DTM and contours (contour interval 5m) are shown in Figure 5.18a and Figure 5.18b respectively.

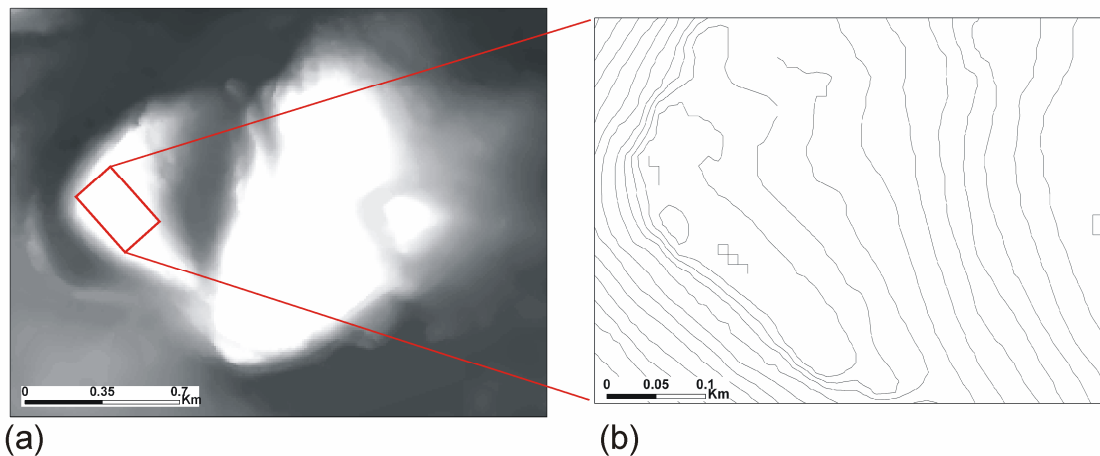


Figure 5.17: (a) Source DTM (b) Contours at 5m interval from source DTM. Contours have 'spikes' and are broken at certain places

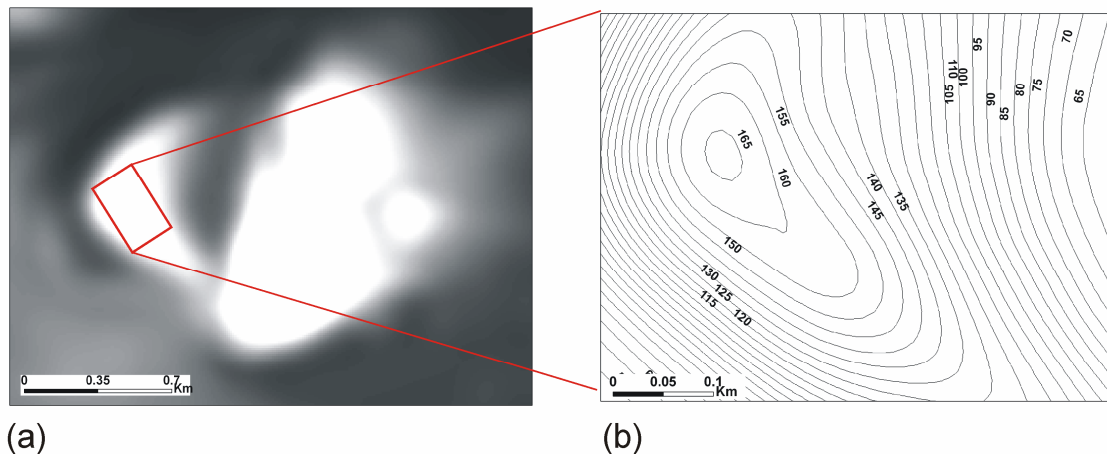


Figure 5.18: (a) Smooth or generalised DTM (b) Resultant contours at 5m interval including height labels

The resultant contours from the generalised DTM were used by the algorithm to identify the summit points and their prominence. The summit points were identified by using the highest closed contours (contours that contain no other contour) and finding the cell from the DTM that has the maximum elevation within each of those contours. For such cells a summit point (vector geometry) is generated that stores the location and its elevation (Figure 5.19). The second step is the calculation of prominence for each summit point. The algorithm finds the key contour for each summit point using the above definition of key contour. It then calculates the

prominence by subtracting the summit's elevation from the elevation of the key contour (Figure 5.19 and Figure 5.20).

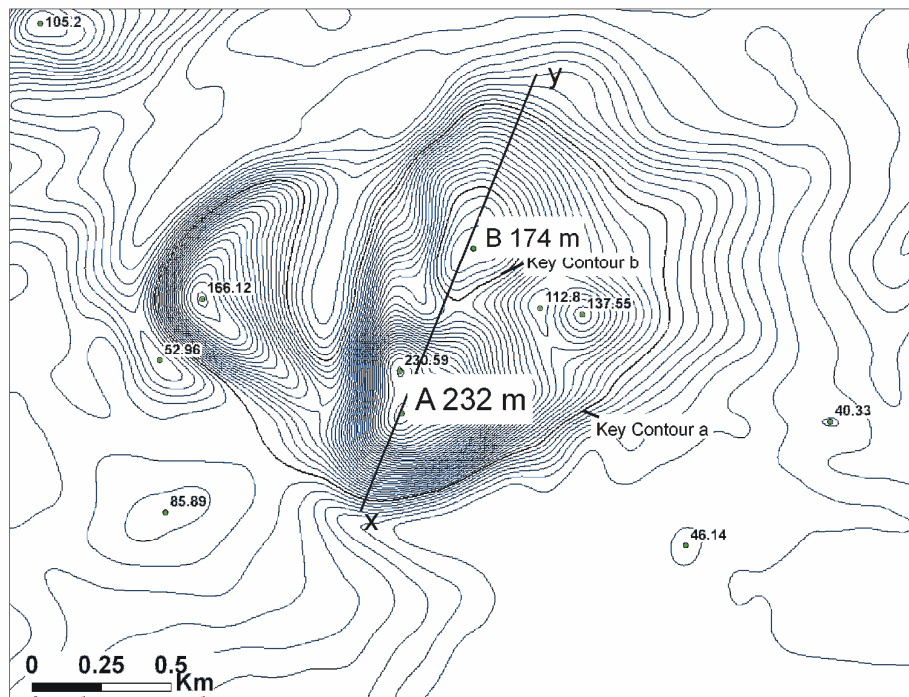


Figure 5.19: Contours created from a smooth DTM (Figure 5.18a). Summit points along with their elevation values are shown within each of the highest contours. Key contours of summit A and B are highlighted in bold. Note that all the summit points that are inside the key contour 'a' of summit A are of lower elevation than summit A (232m)

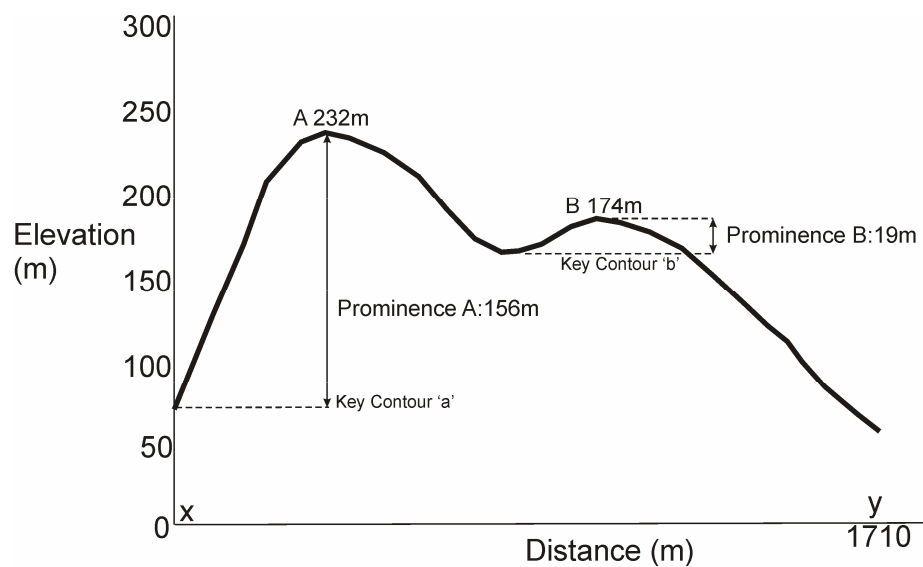


Figure 5.20: Approximate profile of transect from Figure 5.19 showing the Prominence of summit A and B

### 5.4.2 Modelling morphological variance

To determine the areal extent of a hill or mountain's summit along with its prominence we also need to take into account the surface variability between the peak and the key contour. This is because it is not meaningful to define extent purely in terms of the key contour. Theoretically such a rule would make the coastline of Great Britain (GB) the key contour for its highest peak (Ben Nevis 1174m). This does not accord with our own perception of the extent of the region that contains this peak because the surface between the summit and this key contour is not changing sufficiently. Thus in addition to prominence we also need to model the amount of change in the surface in order to identify the extent of a hill, or mountain range and use this information to bound the region.

This change in elevation of a surface can be modelled based on its morphology. One approach is to classify the surface in terms of its morphometric features or classes (pits, peaks, passes, ridges, channels and planes). Several methods exist for the identification of these morphometric features (Evans, 1972; Peucker & Douglas, 1974; Tang, 1992). Here we have used a technique developed by Wood (1996a) that uses an approach based on the quadratic approximation of a local window or kernel of given size in order to find the first (slope) and second derivative (curvature) of the DTM. This method assigns each location of the generalised DTM to one of the six morphometric classes. Due to the scale-dependent nature of the phenomena there is a degree of fuzziness in a location's classification and extent (Fisher *et al.*, 2004; Wood, 1996b). This means that a location classified as a peak at one scale may be viewed as a ridge at another scale, or a plane at some other scale. There has been a great deal of research dealing with modelling the fuzziness of a landform (Fisher, 2000; Robinson, 1988, 2003; Robinson *et al.*, 1988; Usery, 1996; Wood, 1998). In this research, the fuzziness in classification was modelled by using the method developed by Wood (1996a, 1996b), where the generalised DTM is modelled at different scales using different kernel sizes (ranging from 3\*3 to 51\*51). Each location, for each kernel size is classified, into one of the six morphometric classes. The final class of each location in the resultant surface is the one which is most dominant over all kernel sizes (Figure 5.21).



The resultant morphometric units shown in Figure 5.21 are converted into polygons. All polygons that are non plane (i.e. pit, channel, pass, ridge or peak) depict areas with change in morphology. These polygons are called morphologically variable polygons (Figure 5.22) and are used by the algorithm for the identification of extent of each summit as explained in next section.

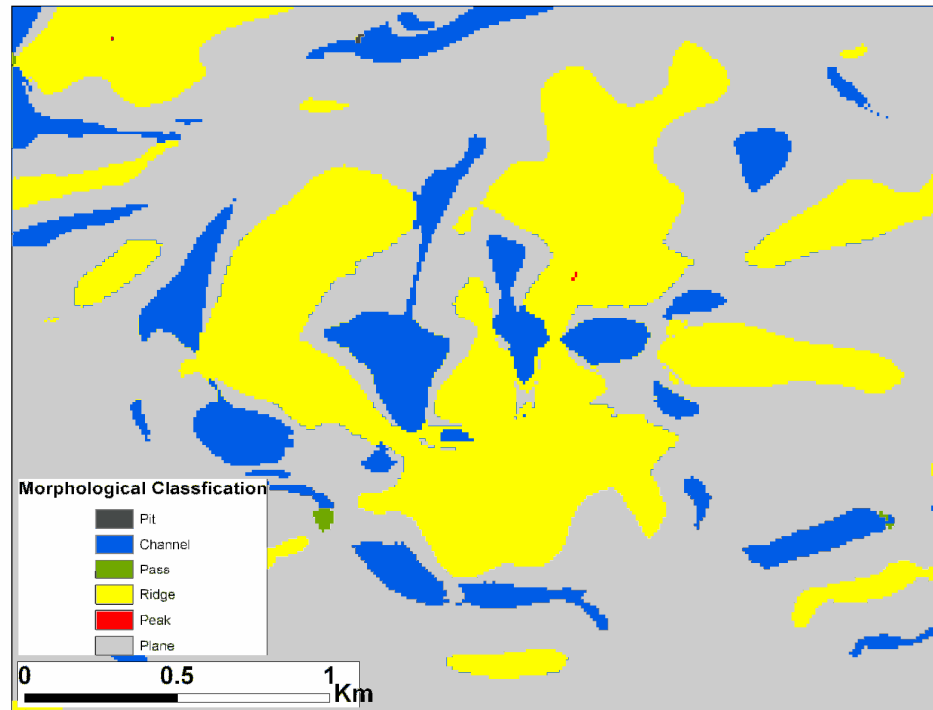


Figure 5.21: Multiscale morphometric classification of DTM in Figure 5.19. The kernel used ranges from 3\*3 to 51\*51

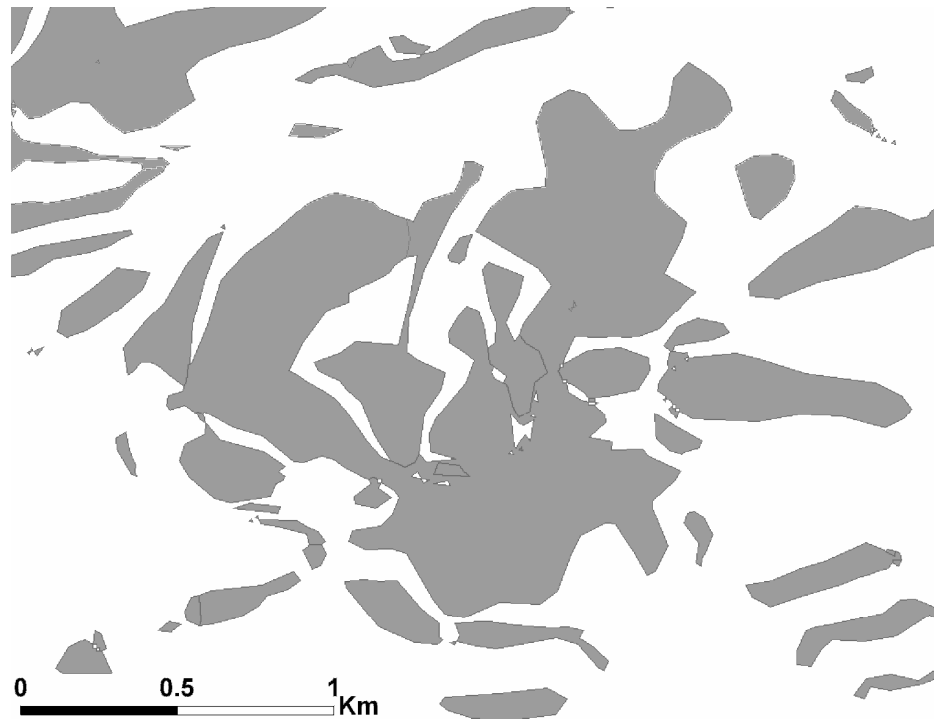


Figure 5.22: Morphologically variable polygons of morphometric features as depicted in Figure 5.21

#### 5.4.3 Calculating the extent of each summit

We can now combine the information of prominence, the key contours and the morphologically variable polygons in order to identify the extent of each summit. In essence we identify the contour that best overlaps with the morphologically variable polygons. We start with the key contour polygon of a summit and intersect it with each morphological variable polygon and calculate the area intersection. The total area intersection divided by the area of the contour polygon gives the percentage of variability within that contour. The percentage is compared against a threshold called the minimum morphologically change threshold (MMC). If the percentage is below this threshold it indicates that the variability in the surface is too low (or area is too plane) and so the next higher contour of the given summit is selected. This process is repeated until the percentage is above or equal to a MMC. The value of MMC was determined empirically and was set to 65%. Reducing the value of this threshold would include more plane region as part of the summits extent and increasing it would result in an extent with least non plane region. Depending upon the target application and scale the value of MMC can be altered accordingly. The

contour polygon selected from this process is assigned as the extent of the given summit. This sequence of events is illustrated in Figure 5.23 in which we start with the key contour for summit A. The percentage of variability is below the MMC (Figure 5.23a). In Figure 5.23b and 5.23c the next higher contour is selected and the same process is repeated and again the percentage is found to be below MMC. In Figure 5.23d the percentage of variability for summit A is found to be greater than MMC so this contour polygon is assigned as the extent of summit A (Figure 5.23d). Figure 5.24 illustrates the extents of all summits identified in Figure 5.23a and Figure 5.19.

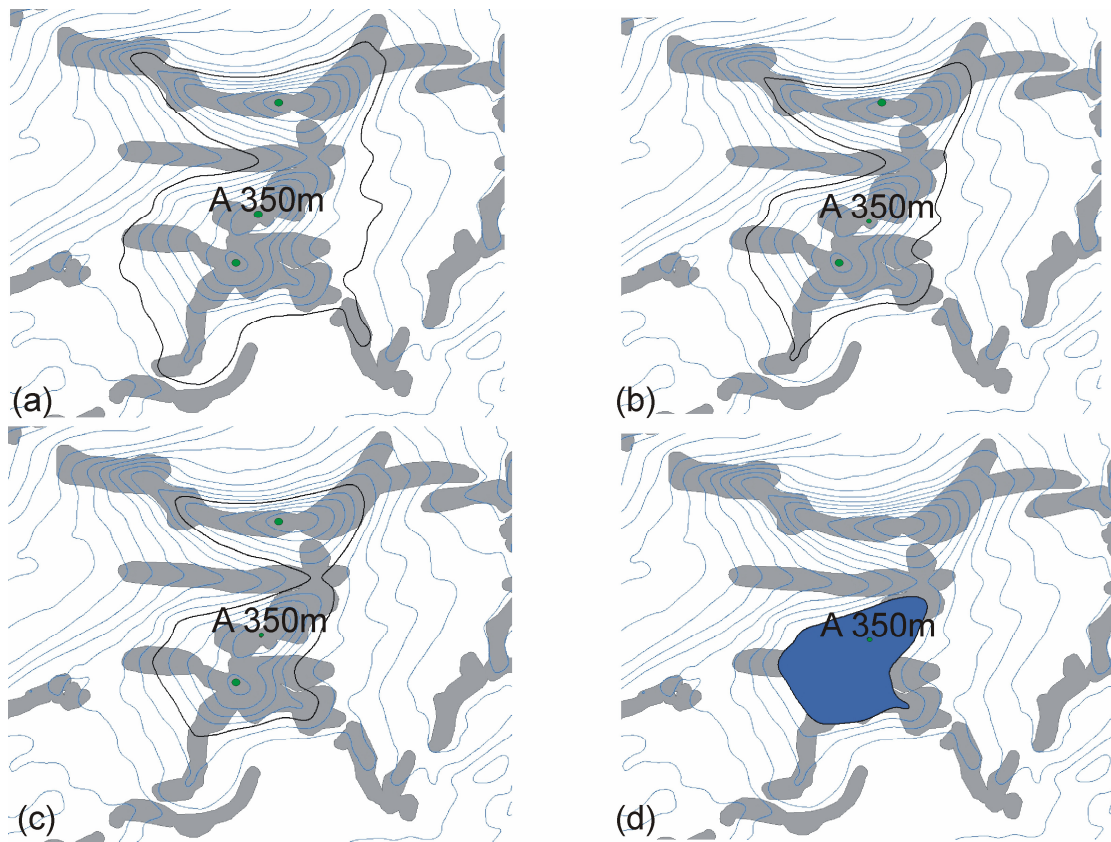


Figure 5.23: Determining the extent of a summit A.(a): Key contour A, morphologically variable polygons (b) next higher level contour is selected (c) next higher level contour is selected (d) the resultant extent of summit A

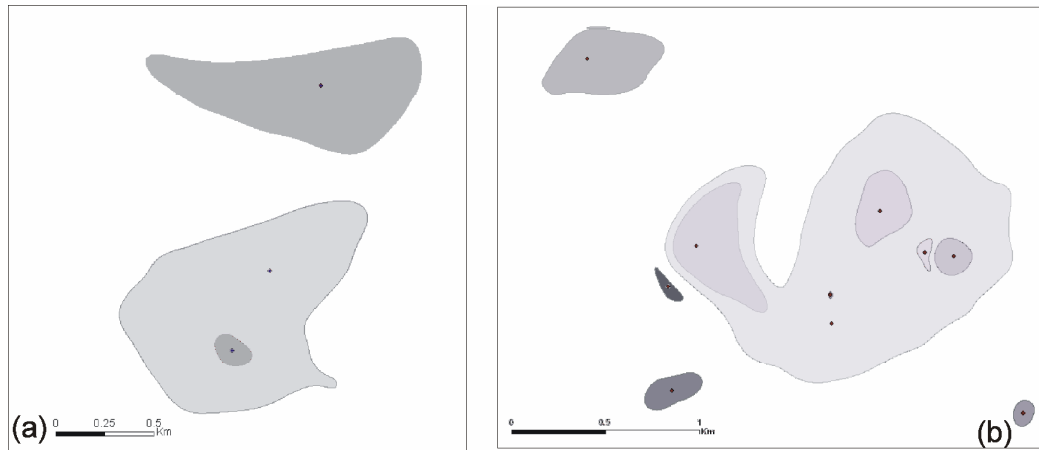


Figure 5.24: (a) Summits and their extents identified in Figure 5.23a (b) the summits and extents of Figure 5.19

#### 5.4.4 Classification into Hills and Ranges

Once the extents of the summits have been identified we can classify these into hills and ranges. These concepts enable us to group summits on a landmass into a hierarchy showing which summits are ‘sub-peaks’ of others. In this way ranges can be identified from groups of individual hills. If a summit has child summits within its extent then it is a range. On the other hand if a summit doesn’t have any summits within its extent it can be classified as an instance of hill class. In Figure 5.25 summit A is categorised as a range since its extent contains summit B whereas B and C are both classified as hills since their extents do not contain any other summit.

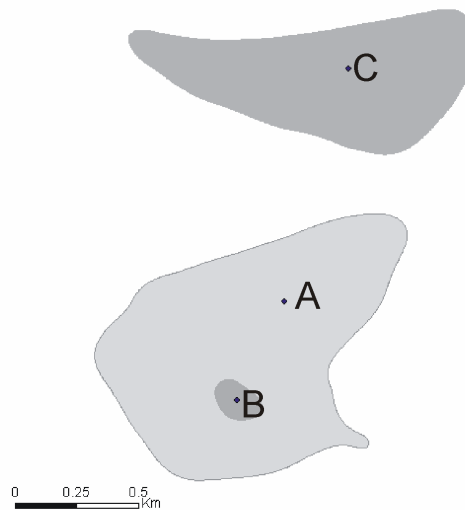


Figure 5.25: Summit A is a range since its extent contains summit B which on the other hand is a hill. Similarly summit C is a hill because it does not contain any summit in its extent

Once the classification process is complete, the resultant extents are selected based on hill and range class intensions (Table 3.1) i.e. the prominence of the object needs to be at least to 35m for resultant level of detail (Figure 5.26). The next section presents results for a few regions on which this methodology was subsequently applied.

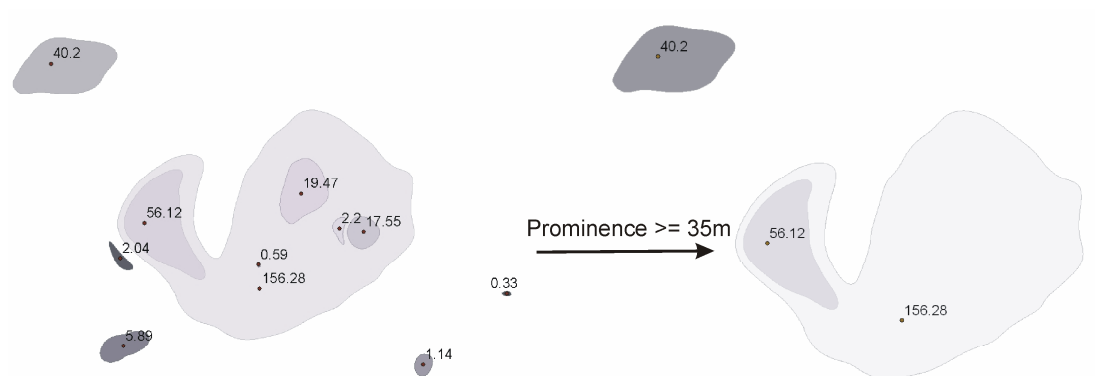


Figure 5.26: Hill and range boundaries above threshold of 35m. The summit B (prominence 19m) in Figure 5.19 has been aggregated into its Peak A (prominence 156m).

### 5.4.5 Implementation and Results

The algorithm was implemented in Java, and used functionality from ArcGIS 9.0 and LandSerf. Landserf is free source software that was required to carry out digital terrain model analysis which included creation of generalisation (or smoothing) of the DTM and morphometric classification. Landserf provides Java libraries that could be called from our code written in Java. The methodology (Figure 5.16) was applied in the derivation of hill and range container boundaries using OS Land-Form PROFILE Plus. Figure 5.27 shows the source DTM for Edinburgh, in Scotland. Figure 5.28 shows the resultant container boundaries for hills and ranges with a prominence of greater or equal to 35m. Figure 5.29 shows the DTM for the Peebles region in Scotland. Figure 5.30 shows the resultant boundaries for hills and ranges in this region with prominence above 35m. A few additional results for some other regions are presented in Appendix V.

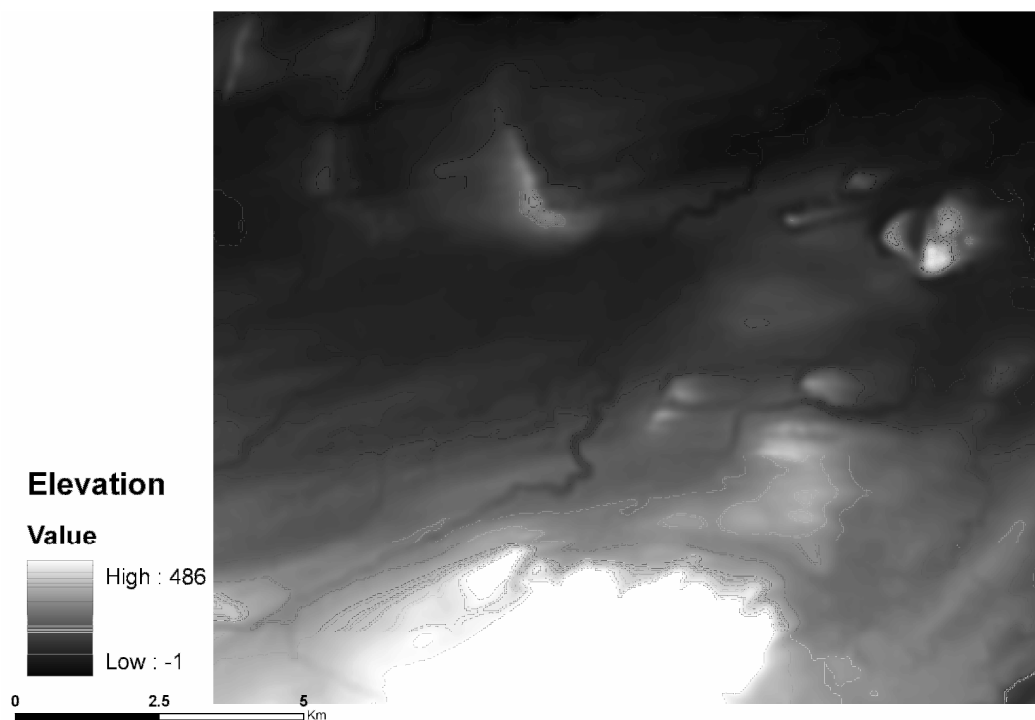


Figure 5.27: Input DTM for Edinburgh and surrounding in Scotland

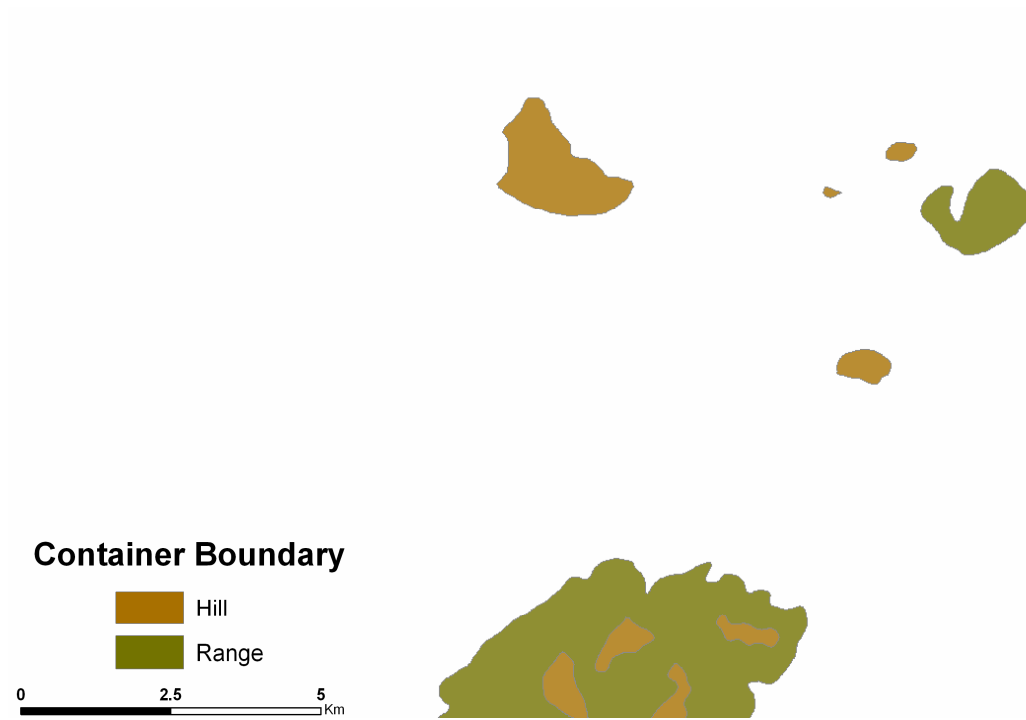


Figure 5.28: Resultant hills and range container boundaries for Figure 5.27 above 35m prominence

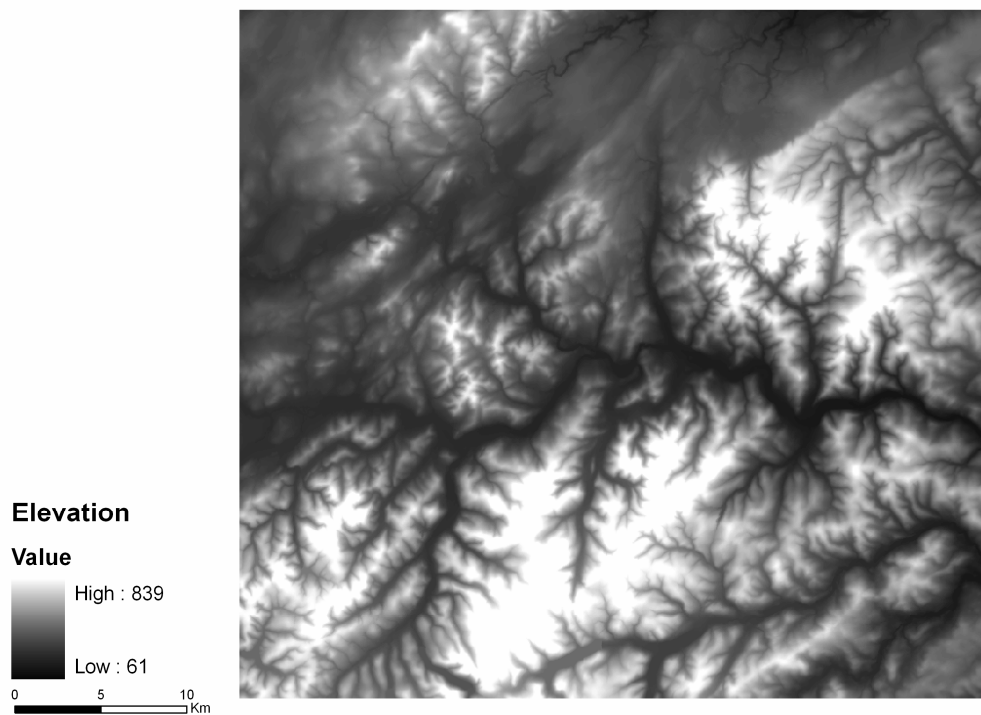


Figure 5.29: Input DTM for region surrounding Peebles in Scotland

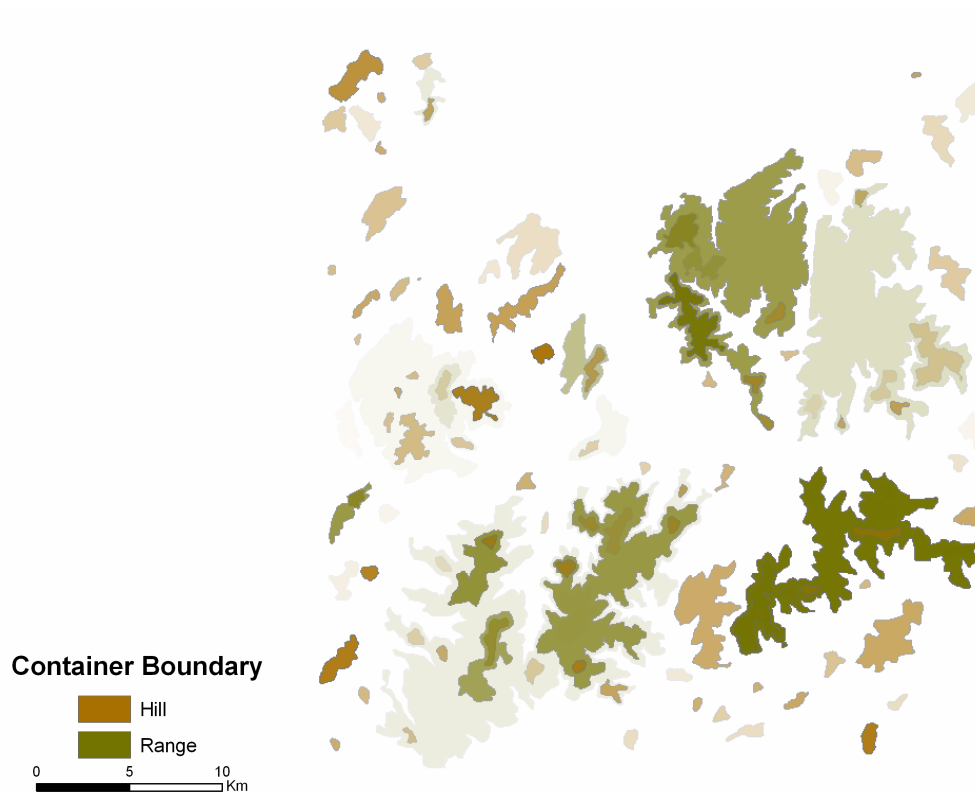


Figure 5.30: Output container boundaries of hills and ranges for Figure 5.29 above 35m prominence

## 5.5 Summary

As discussed in Chapter 2, geographic concepts are scale dependent thus the boundaries of their instances are also ‘level of detail’ dependent. The boundaries, presented in this chapter are all examples of fiat boundaries. Such boundaries are not explicit but owe their existence to human cognition or mathematical definitions. The three target concepts (composite classes) (settlement, forest, hill/range) in the target data model are fiat. Each of these concepts have specific properties that differentiate them from one another. Different approaches are thus required for the detection of boundaries of objects of these classes. Three approaches were presented in this chapter - one for each composite class. For settlement and forest classes we exploit the property of typical or necessary parts of partonomic relationships, as discussed in Chapter 2 (section 2.5.3). We hypothesise that the typical parts of a settlement and forest, i.e. buildings and trees respectively, play a critical role in defining the extents of a settlement and forest objects. On the other hand, prominence and morphological properties were used for defining the extents of hills and ranges.



The boundaries presented here act as containers for all objects in the source database. These (container) boundaries can be used to determine the partonomic relationships which are used in the aggregation process. The next chapter presents the database enrichment and aggregation stages of the methodology for database transformation.

## **CHAPTER 6: Database Generalisation via Aggregation based on Partonomies**

Chapter 5 presented the methodologies and results for creating container boundaries for composite classes (settlement, forest, hills and ranges). As explained in Chapter 5, these boundaries were generated using specific properties and specific objects of each composite class. In order to transform the source database objects completely we need to bring these separate boundaries together. This chapter presents the process of source database enrichment in terms of creating partonomic relationships using the container boundaries of Chapter 5 (section 6.1). This section also presents an approach for modelling the non exclusive nature of these relationships. The next section (6.2) presents the approach used to create the geometries of the required objects of the target database via selection and aggregation operations. The last section (6.3) discusses the implementation of the methodology.

### **6.1 Making Partonomies Explicit**

In Chapter 3 (section 3.6.4) we presented the importance of partonomic relationships in order to transform the source database objects, via aggregation, into objects of the resultant classes. As discussed earlier, such relationships are not explicitly defined in the source database and need to be made explicit before aggregation can be performed. The boundaries (Figure 6.1) are used to determine the partonomic relationships of source objects in terms of the resultant objects. These boundaries act as ‘containers’ – all objects within are classified as ‘part of’ the composite classes. The topological relationships between the source objects and resultant boundaries are identified. If the source objects are completely ‘covered by’ or ‘inside’ the resultant boundary they are deemed to be part of the composite object (Figure 6.1). If it is an overlap operation then area intersection is calculated. This is then divided by the area of the source object which gives the percentage or degree of membership (or partonomy) (Figure 6.2). If the topological relationship is ‘disjoint’ or ‘touch’ the objects are not part of the resultant composite object.

It is important to point out that we can determine the parthood, without having the overhead of determining the topological relationship, directly by calculating area intersection between each source object and each container boundary. For small datasets this is fine but for large datasets determining the area intersection for each source object with each container boundary is computationally intensive. If the objects in the spatial database are spatially indexed, as is the case in this research using Oracle (Oracle, 2005), then determining the topological relationship is much faster and area intersection only needs to be performed for objects that have an ‘overlap’ relationship. In the case of ‘inside’ or ‘cover’ relationships the area intersection does not need to be calculated separately because the source object’s interior is completely contained by the container boundary. The degree of parthood in these cases is directly set as 100%. This degree of parthood is used in the selection of objects for aggregation (explained in section 6.2.2). As discussed in Chapter 2 (section 2.4.4) the parthood relationships are non exclusive i.e. m:n. The parthood relationships identified for the source objects using the container boundaries can result in multiple parthoods. This needs to be modelled and is explained in the following section.



Figure 6.1: Container boundaries (forest and settlement) using the approach proposed in Chapter 5 overlaid on source database objects. The container boundaries are labelled with their object identifier. A portion of overlapping area has been highlighted in red and is shown in Figure 6.2

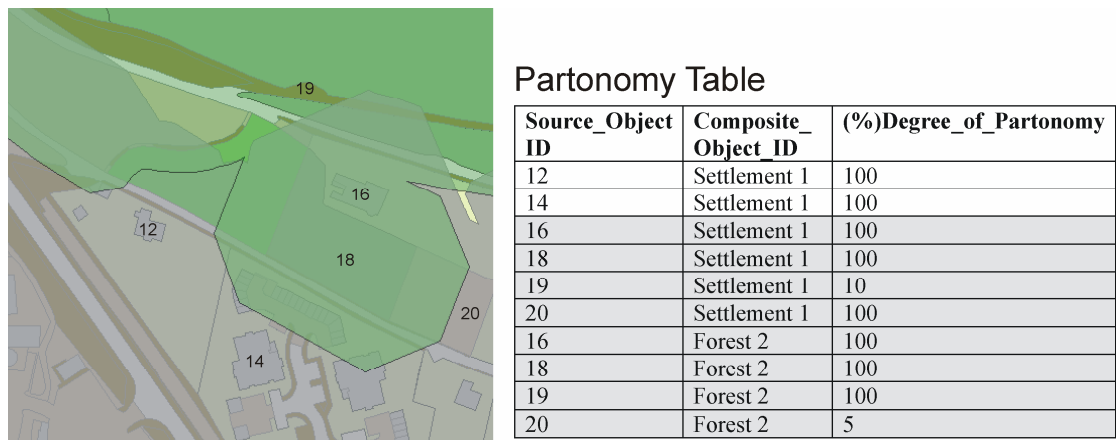


Figure 6.2: Partonomic relationships for a few source database objects in the region highlighted in Figure 6.1. The rows highlighted in the table are those objects having multiple partonomies

### 6.1.1 Modelling a palimpsest of partonomies

The container boundaries for the composite classes presented in Chapter 5 were created using specific properties and objects of specific classes of the source database. As a result the container boundaries are not consistent with each other and may well result in overlap at certain places (as shown in Figure 6.1 and Figure 6.2). In such cases some of the source objects will have topological relationship (inside and overlap) with all boundaries that either cover them completely or partially. These overlapping containers can be envisaged as a palimpsest of containers. Thus the resultant partonomic relationships for these source objects will be multiple (non exclusive). This is modelled in the database by storing the partonomic relationship information for each source object stored in a separate table (partonomy table Figure 6.2). The primary key of this table is not a single attribute but a collection of attributes – object identifier of source object and object identifier of the composite object. Modelling of multiple partonomies is useful for spatial analysis and for further enhancement of the target data model (Figure 6.3) (discussed in Chapter 7).

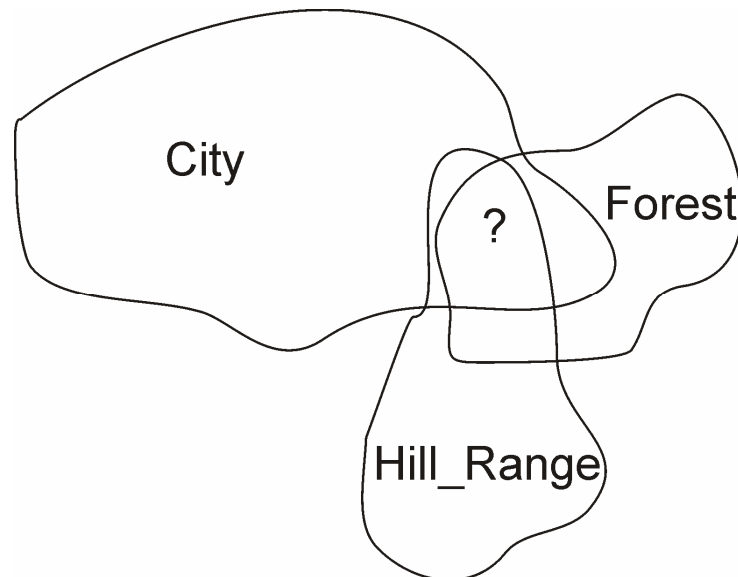


Figure 6.3: Using multiple partonomic information to support multi criteria spatial analysis. For instance finding all objects that are part of leafy-suburban and hilly region.

Any source database object that does not interact (degree of partonomy 0) with any of the container boundary objects (settlement, forest, hill and range) is modelled as part of the 'General Land' class object. This is required so that after aggregation no holes are present in the resultant database (i.e. it becomes geometrically partitioned). But once more classes such as lakes, costal area and rivers are added into the target data model such source objects would have partonomic relationships with objects of these classes. This would result in creation of composite objects of new classes. But for the current research a 'General Land' class is considered.

## 6.2 Database Transformation

Database transformation involves the creation of objects that are instances of the classes of the target data model. Objects of the target data model classes (Table 3.1) are created using aggregation and selection operations. This section presents the process for the creation of the resultant objects. As explained in Chapter 3 (section 3.6.3 and 3.6.4) depending on the type of resultant classes (super or composite), the aggregation of the source objects can be based on taxonomic relationships and partonomic relationships. Thus there are two aggregation processes that are being carried out. These are explained with examples in the following sections.

### 6.2.1 Selection and Aggregation of Road and Rail class objects

It is not necessary that all classes of the target data model are different from the classes of the source data model. Depending upon the resultant application and level of detail it is possible that there are a common set of classes in the source and the target data models. As pointed out by Müller (1989) the change in geometry and class occur at different scales (or levels of detail) for different objects and is determined by the application of the resultant database or map. In this research important road classes (Motorway, A road, B road, Minor road and Junction) and the railway class are common in both the target and the source data models (Table 3.1 and Table 4.2). Objects of these classes, for the target database, can be directly selected from the source database.

The 'Other Roads' class in the source data consists of objects such as local streets, alleys, alleyways and pavements, which are unimportant (too detailed) with respect to the intended level of detail (1:250,000). Objects of this class would be aggregated with other source objects into composite objects based on paronomic relationships (section 6.2.2). Because of this, certain junctions, (between an important road class object and an 'Other Roads' class object in the source database), are redundant in the target database. Such junction objects can be aggregated with adjacent road objects there by reducing the spatial detail, in terms of the number of objects, in the resultant database (as illustrated in Figure 6.4). Here in Figure 6.4a the junction object (ringed in red) can be aggregated with the adjacent 'A Road' class objects without loss of any information since the 'Other Road' class object would not be present in the resultant database. On the other hand the junction object that is kept in Figure 6.4b serves as a link between three important road objects (two A Road class objects and a Minor road object) and thus carries useful semantic information. Therefore it is not aggregated with adjacent road objects and is kept as separate object in the resultant database (Figure 6.4b).

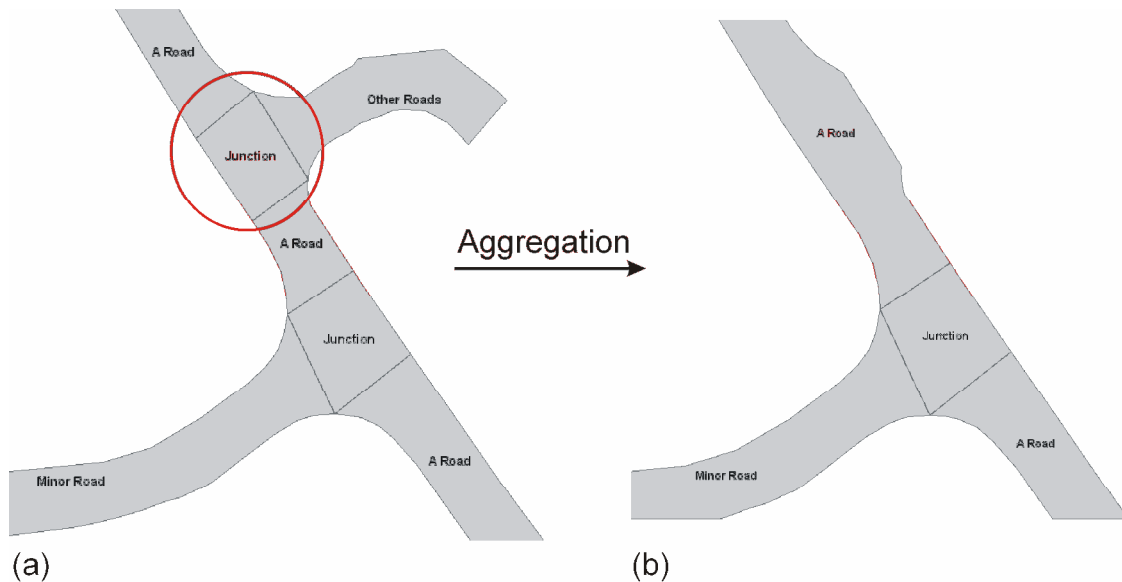


Figure 6.4: Aggregation of road segments based on adjacency and being of the same class

### 6.2.2 Aggregation via Partonomy

Creation of instances of composite classes (settlement, forest, hills, ranges and general land) requires us to aggregate source objects belonging to different classification. Such objects are created from source database objects by aggregation based on partonomic relationships. Once these relationships have been explicitly identified (section 6.1), we can use them relationships to create the geometry of composite objects via aggregation. As defined earlier, a composite object can have disconnected geometry (non contiguous), but its boundary should not overlap with the boundary of another object (topological constraint) since the partonomic relationships between the source objects and composite objects can be many to many (section 6.1.1), aggregation, without constraints, can result in overlapping geometries between composite objects (Molenaar, 1998). This is illustrated in Figure 6.5. The figure illustrates the geometries of composite objects (settlement and forest) objects for the region shown in Figure 6.2.



Figure 6.5: Overlapping composite objects due to m:n partonomic relationships for the region highlighted in Figure 6.2. The forest (light green) and settlement object (light grey) overlap because of aggregation based on multiple partonomies without any constraints

The light grey object in Figure 6.5 represents the settlement object created by aggregation using partonomic relationships. The light green object in Figure 6.5 represents a forest object also created via aggregation based on partonomic relationships. As illustrated in Figure 6.5 because of multiple partonomies the resultant composite objects overlap, resulting in violation of topological constraint i.e. spatial objects in the resultant database should not overlap (Peng, 1997).

Molenaar (1998) suggests in such cases, where objects have multiple partonomies, it is valid to define aggregation rules such that the component (source) objects can be aggregated into only one composite object. These aggregation rules ensure topological consistency in the target database. These aggregation rules are explained using a model database objects in the next section.



### 6.2.3 Selection and Aggregation via specific rules

Figure 6.6 shows a sample set of source database objects. These objects are overlaid by two container boundaries (forest and settlement in Figure 6.6a) in order to illustrate the partonomic relationships for source objects. In the first step, before aggregation, all source database objects that are part of the same settlement object Figure 6.6(b1) and forest object Figure 6.6(b2) are selected. As shown in Figure 6.6(b1) and (b2) some of the marginally connected objects (land cover objects) and important object (A road) objects are part of this selection. The marginally connected objects, (i.e. objects having a degree of partonomy too low), are removed from the selection. This is because their degrees of partonomy illustrates a very weak relationship between the source and composite object. This is done via examination of the degree of partonomy against a degree threshold (DT). Important road objects (Motorway, A Road, B Road, Minor and Junctions) and railway objects are also removed from the selection since they are populated using the approach presented in section 6.2.1. The resultant selected objects for settlement and forest are shown in Figure 6.6(c1) and (c2).

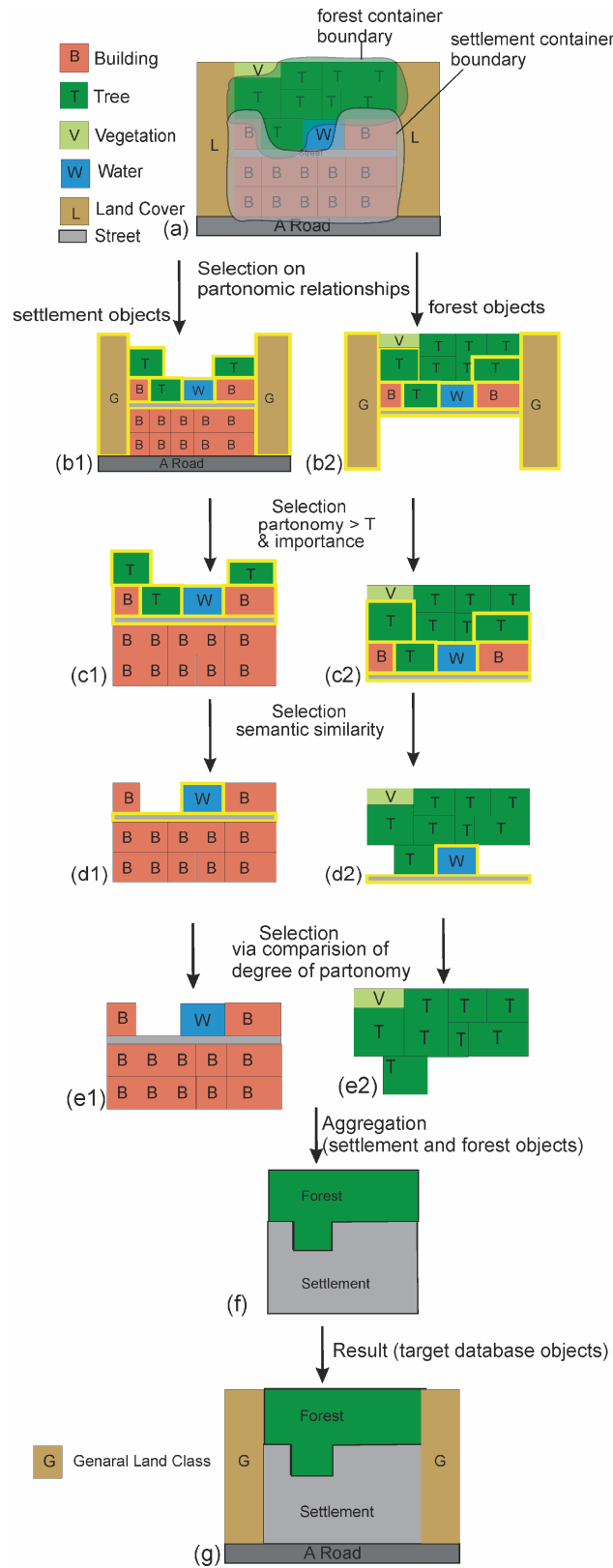


Figure 6.6: Aggregation stages for the creation of (non overlapping) composite objects from component objects in the source database

In the next stage, for each selected object, we check if that object is also part of some other composite object with a degree of partonomy above DT. For instance a few objects in Figure 6.6(c1) and (c2) (highlighted in yellow) are part of both settlement and forest. In such cases the taxonomic similarity between the source object and the other composite object class is used for selection or elimination. For instance, if a tree object is part of a settlement and is also part of a forest object then it is removed from the selection of settlement object (Figure 6.6c2). This is because a tree or vegetation class object is more (taxonomically and partonomically) similar to a forest class than a settlement class. Similarly, using the same principle, if a building object is part of a forest object and settlement object, as in Figure 6.6c1, then it would be removed from forest object's selection. The resultant objects are shown in Figure 6.6d1 and Figure 6.6d2.

But this approach is limited to only those source and target class objects that we know are related (tree, vegetation –forest, building – settlement). It is because we do not always know the the taxonomic or thematic similarity between all source and target classes (as stated in section 3.6.4). For example for a water class object that is part of both forest and settlement class (as in Figure 6.6d1 and Figure 6.6d2). There is no similarity between the source class and the composite class. Thus even after similarity check there can be common selected objects, as shown in Figure 6.6(d1) and (d2). In order to resolve these specific cases we use the degree of partonomy to make the selection exclusive. For each such common source object we compare its degree of partonomy for one composite object with its degree of partonomy for the other composite object. The object is removed from the composite object's selection if its degree of partonomy is less than the degree of partonomy for the other composite object. In Figure 6.6(d1) and (d2) both the street object and the water object have a degree of partonomy higher for the settlement object as compared to the forest object. Thus they are removed from the forest object's selection (as shown in Figure 6.6e1 and Figure 6.6e2). In a very few cases, it is possible that the degree of partonomies is exactly the same for the common object. In such cases both the similarity and degree of partonomy were not able to resolve the non exclusive selection issue. For such special cases a pre-defined order that defines the

importance of classes in terms of aggregation is used. The pre-defined order could be changed depending upon the intended application. In this research the precedence order is settlement than forest, hill, range and lastly general land class. But as stated earlier this pre-defined order is not absolute and can be altered depending on the application of the resultant database.

Using the above steps the selected objects are now exclusive to each composite object and can thus be aggregated into a resultant composite objects (Figure 6.6f) without any overlaps. Figure 6.6(g) shows the resultant database objects for the corresponding source objects shown in Figure 6.6a.

Figure 6.7b illustrates the output objects for regions shown in Figure 6.7a. These objects have been created using selection and aggregation operations following the steps described above. All objects in the target database form a thematic and geometric partition. Unlike the objects in Figure 6.5, objects in Figure 6.7b do not overlap.

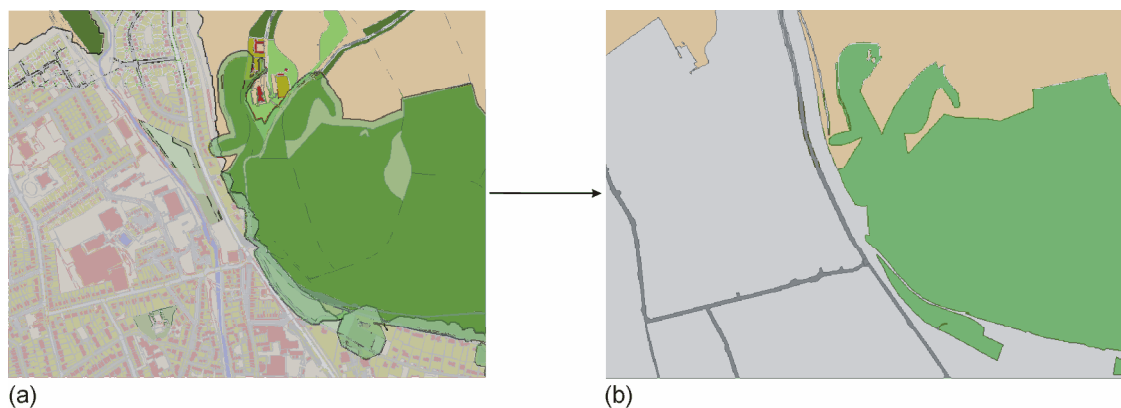


Figure 6.7: (a) Figure 6.1 (b) Resultant objects created by the proposed aggregation approach. The resultant objects are thematically and geometrically partitioned

### 6.3 Implementation

All the above stages of the methodology were implemented using Oracle spatial 10g and Java. Oracle spatial not only provides a database management system for storing and retrieving spatial objects, it also provides all the spatial functions as defined by OGC (van Oosterom *et al.*, 2002). The spatial functions used in this research were

nearest neighbours, within distance, topological relationships and aggregation (Oracle, 2005).

Another reason for using Oracle spatial 10g is that it provides functions to implement spatial indexing on the spatial objects in a table. Spatial indexing enhances the performance of spatial queries. For instance, in order to find all objects adjacent to a particular object in a given table with 7000 entries takes 5 seconds to execute without spatial indexing. The same query takes 0.8 seconds to execute on the same table with the same data but using spatial indexing. This is quite significant for spatial databases containing a large number of spatial objects. For instance the same query performed on a table with 57,000 objects without spatial indexing takes 45 seconds to execute whereas it takes 1.67 seconds to execute on the same table with spatially indexed geometry. Oracle Spatial provides functions for implementing R-tree indexing (the default) or Quad-tree indexing (Kothuri *et al.*, 2002; Oracle, 2005). In this research we have used R-tree indexing for all spatial functions. This is because for most spatial functions, such as distance queries, overlap and aggregation, R-tree indexing out performs Quad-tree indexing. Moreover Quad-tree indexing requires careful fine tuning, depending upon the size of the input dataset, in order to attain better performance (Kothuri *et al.*, 2002).

The performance of the aggregation function to create the geometry of a composite object from its component objects is affected adversely as the number of component objects increases. The performance was improved by dividing a single aggregation operation into multiple nested aggregate functions that are called iteratively. It is important to point out that the proposed methodology could be implemented using a different database management system (DBMS). For instance, one could use PostGIS (<http://postgis.refractory.net/>) which is the spatial extension to Postgres SQL – an open source object relational DBMS. PostGIS also implements the spatial functions that were used in this research and also supports spatial indexing for handling large spatial datasets. Similarly, other open source libraries such as the Java Topology Suite (<http://www.vividsolutions.com/jts/jtshome.htm>) could also be used, but these open and free platforms when implemented on a standard PC, consumed large memory and processor time as the datasets became larger.

A programming language was required in order to call the spatial functions provided by Oracle Spatial when and where desired. In this research the algorithms for the proposed methodology were implemented using the Java programming language. The spatial functions were embedded into Java's code using SQLJ. SQLJ is a set of programming extensions that allows the user to embed SQL statements within Java code. It offers flexibility over traditional JDBC (Java Database Connectivity) in terms of fewer lines for similar programs, easier debugging and stronger type checking of SQL query results (Cline, 2004). ArcGIS and Geomedia were used to visualise the input and output data and for creation of figures presented in the thesis. The next chapter presents results, for three different regions, using the proposed methodology. The chapter also discusses the evaluation and the utility of these results.

#### **6.4 Summary**

The container boundaries for composite objects generated in Chapter 5 are used by the database enrichment stage of the methodology. In this stage the partonomic relationships between source objects and the composite objects are made explicit. The proposed approach is implemented such that it models multiple partonomic relationships for source objects. Once the database has been enriched the next stage of the methodology creates objects of the classes of the target data model. This is achieved by aggregation of adjacent rail objects and road objects belonging to the same class. But for composite classes (settlement, forest, hill, range and general land) this is achieved by aggregation using the partonomic relationships. The aggregation makes use of the partonomic relationships along with a set of aggregation rules in order to create a geometrically partitioned target database. The proposed methodology is implemented using Oracle Spatial 10g and Java.

## CHAPTER 7: Results and Discussions

This chapter presents the results of the proposed methodology. Section 7.1 presents three case studies illustrating the flexibility of the approach on application to different regions of interest selected from the source dataset. Finding appropriate evaluation techniques for the results is difficult because of the lack of existing solutions against which to compare. Nevertheless the chapter also presents a few methods (both qualitative and quantitative) for evaluation (section 7.2). The chapter also presents methods illustrating the utility of the proposed methodology (section 7.3). The chapter also discusses the generalisation errors and degree of fuzziness in the results. The last section highlights the potential of results in the automated generation of representations at higher abstraction levels.

### 7.1 Case Studies

The proposed methodology was tested on a number of regions selected from the source database. Here we present results of three different regions of interest selected from the source database. These regions were selected because they contained a variety of topographic features and presented some interesting challenges. In all three regions the source database objects have been classified according to the classification stage of the methodology (section 4.2). The interim results (container boundaries) are also presented along with the final results illustrating the objects in the target database for each case study.

Figure 7.1 shows the first region selected from the source dataset. The region consists of approximately 55,000 objects. The region has a high density of buildings in certain areas and a few large tree objects. The container boundaries for this region are shown in Figure 7.2. There are no hills or ranges found in this region. Figure 7.3 shows the output objects of the resultant database.



Figure 7.1: Source Database objects classified according to the pre-processing stage of the proposed methodology (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved). The region is Livingston in Scotland



Figure 7.2: Settlement and forest container boundaries for the input region shown in Figure 7.1





Figure 7.3: Objects of the resultant database classified according to the target data model

Figure 7.4 shows the second case study, region of ‘Peebles’ and its surroundings in Scotland, Great Britain (GB). This region consists of approximately 70,600 unique objects. Most of the region is populated with forest, hills, ranges and a few small settlements. The container boundaries are shown in Figure 7.5. There were many places where the container boundaries overlapped thus resulting in multiple partonomies for various source objects. Figure 7.6 illustrates the objects of the target database.

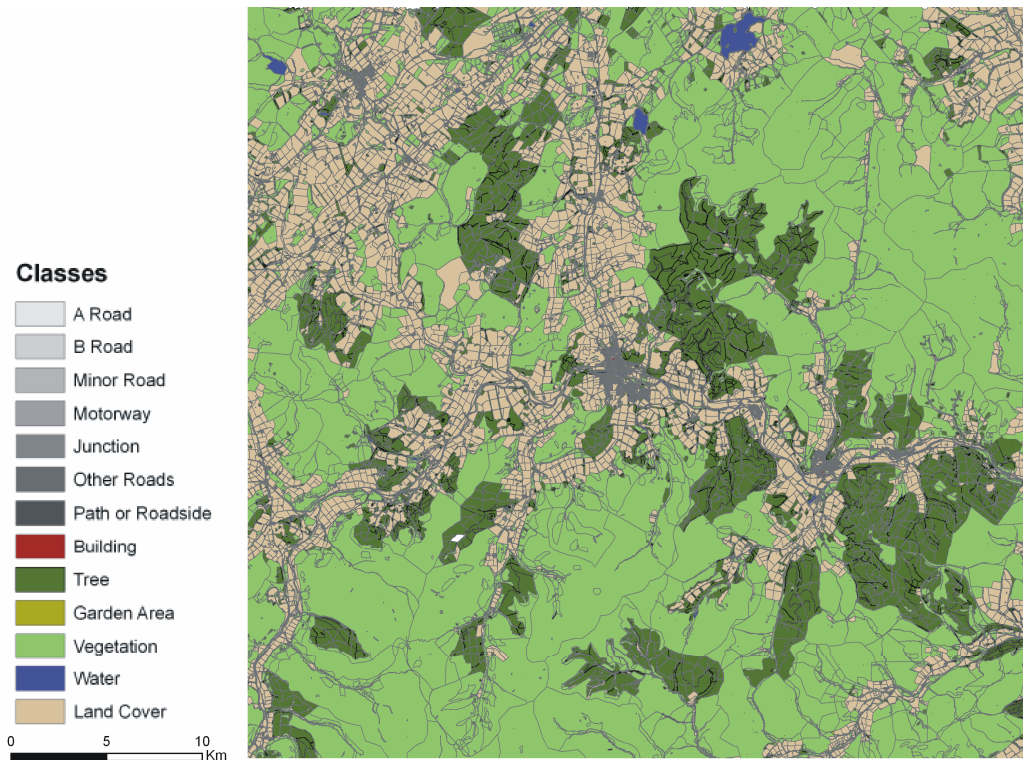


Figure 7.4: Source Database objects classified according to the pre-processing stage of the proposed methodology (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved). The region is Peebles in Scotland

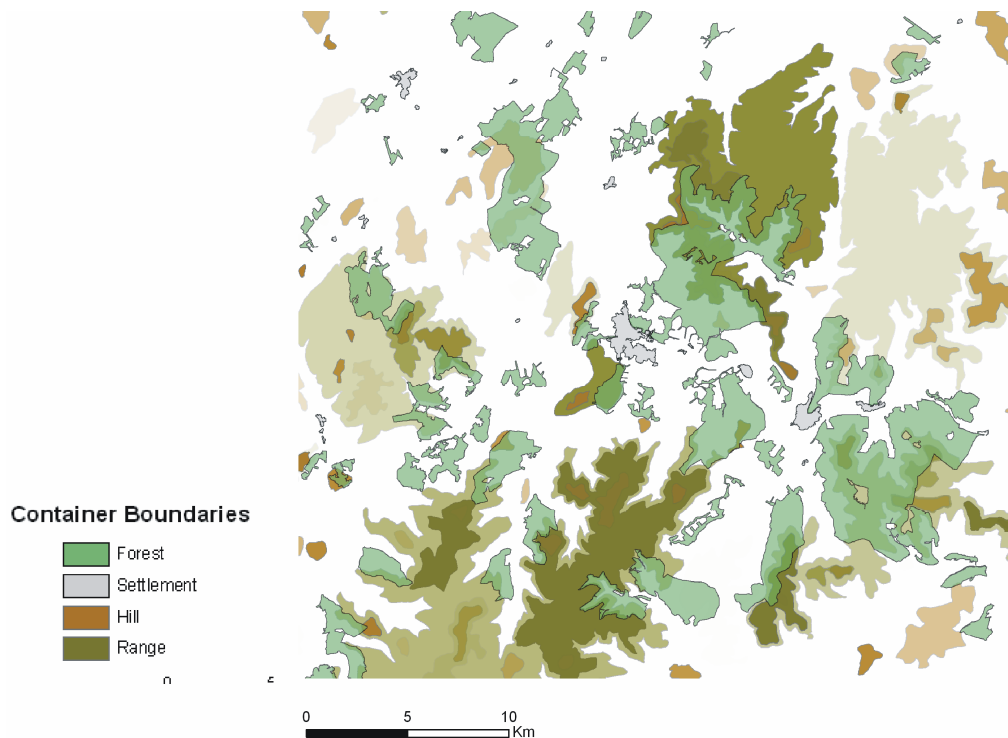


Figure 7.5: Settlement, forest, hills and range container boundaries for the input region shown in Figure 7.4

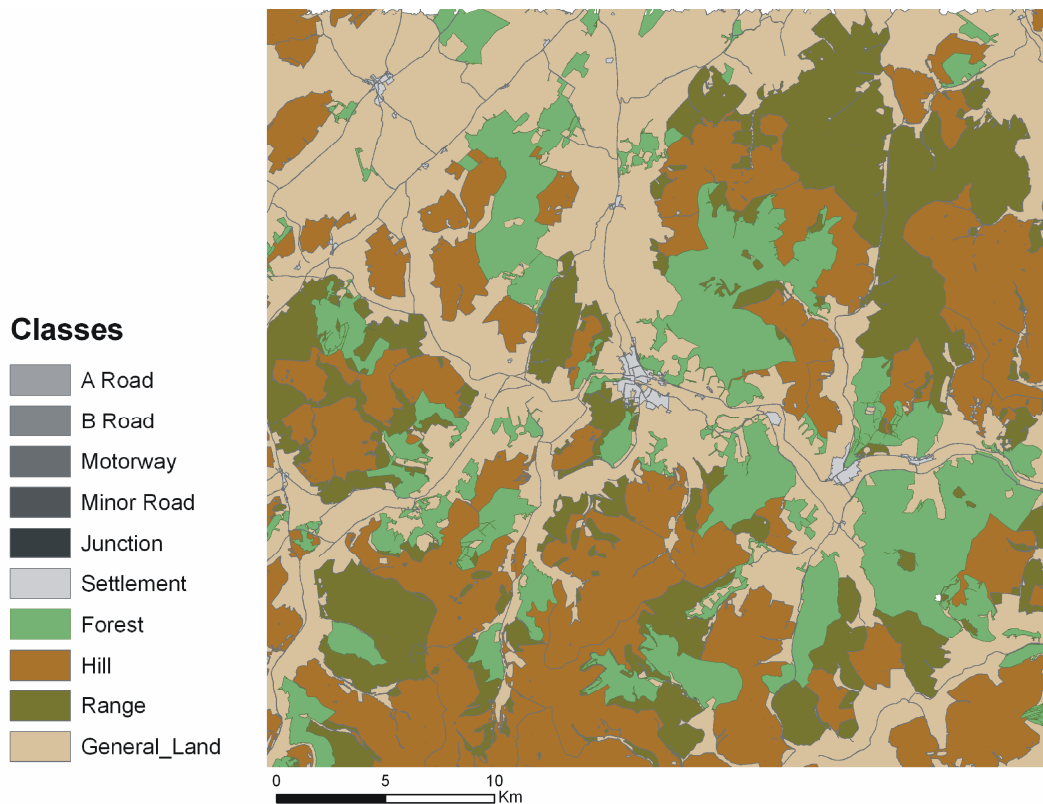


Figure 7.6: Objects of the resultant database classified according to the target data model

The last study region was the city of Edinburgh, Scotland GB and its surrounding area (Figure 7.7). The number of objects in this region is around 400,000. The primary reason for selection of this region was to check the scalability of the approach on large datasets (discussed in section 7.2). The region is mostly dominated by a high density of building objects. A few hills and ranges are also present in this region. The container boundaries for this region are illustrated in Figure 7.8. Figure 7.9 shows the resultant objects from the target database.



Figure 7.7: Source Database objects classified according to the pre-processing stage of the proposed methodology (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved). The region is the city of Edinburgh and its surroundings in Scotland

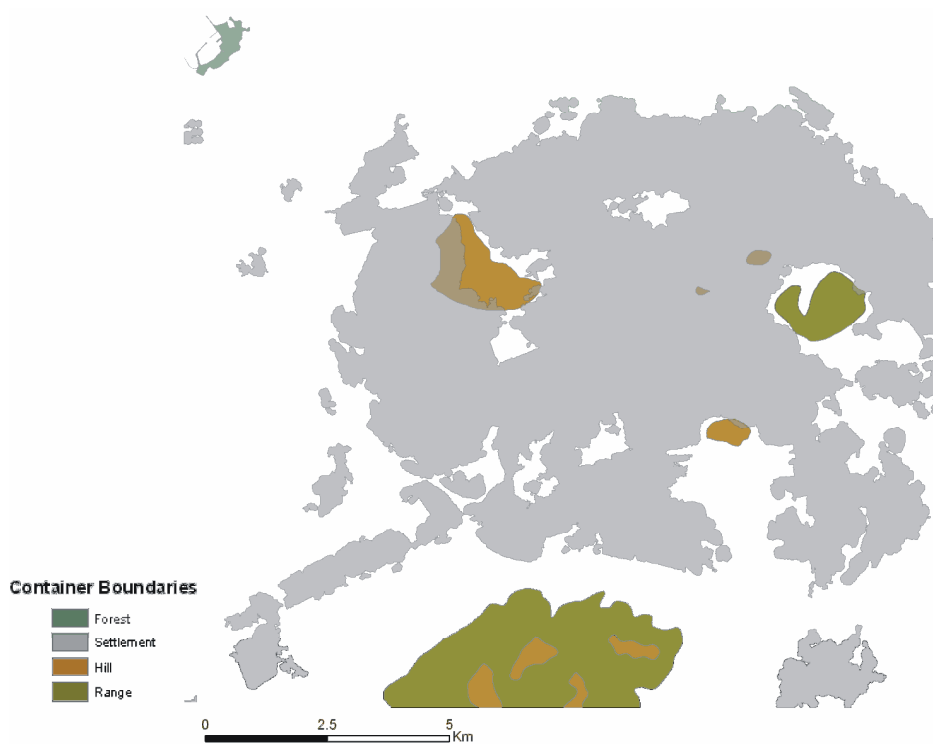


Figure 7.8: Settlement, forest, hills and range container boundaries for the input region shown in Figure 7.7

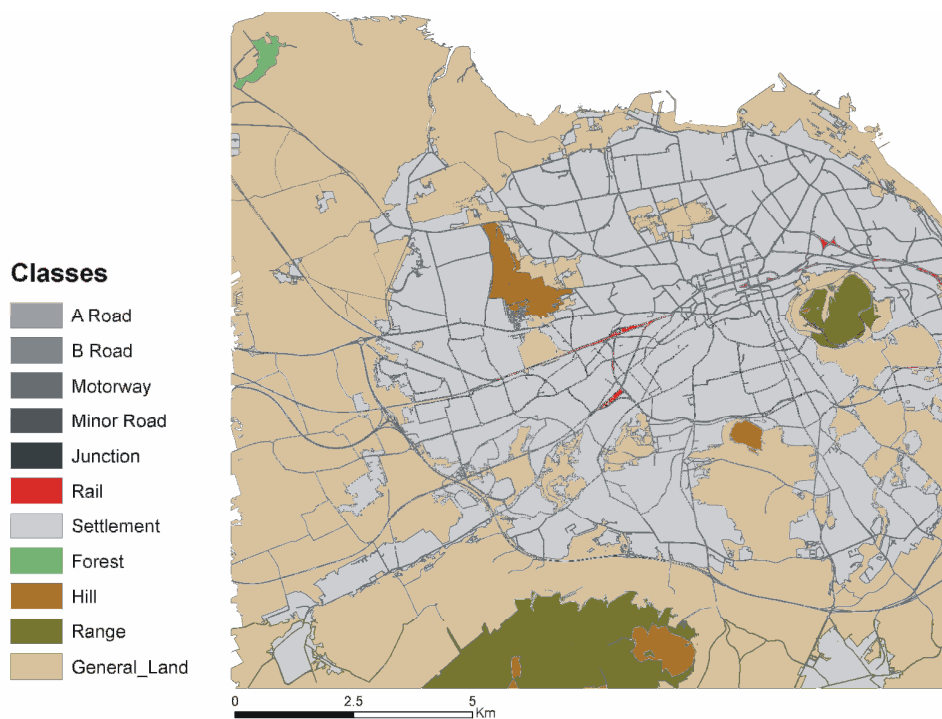


Figure 7.9: Objects of the resultant database classified according to the target data model

## 7.2 Evaluation

The evaluation techniques proposed in generalisation research are usually appropriate for cartographic results and are appropriate over small changes in scale where the phenomena before and after generalisation are quite similar. But over large changes in the level of detail there are fundamental changes in the types of objects and such techniques are not appropriate (Mackaness & Edwards, 2002). Thus finding appropriate methods for evaluation of the results obtained from the proposed methodology is quite difficult. This difficulty was also observed by van Smaalen (2003) for the evaluation of his proposed model generalisation methodology.

In this section we present three possible evaluation techniques. All these techniques use Strategi dataset for evaluation purpose. As stated earlier Strategi is a *manually* generalised cartographic dataset at a scale of 1:250,000. The dataset was obtained in vector format which has separate layers for each class. The layers used in this research for comparison were settlement, forest, roads (Motorway, A road, B road, Minor road), railway, land use and settlement text points. The evaluation was carried out for the three case studies presented in the previous section. These evaluation techniques are presented in subsequent sections.

### 7.2.1 Visual Comparison

The results obtained from the model generalisation approaches are not meant as final cartographic quality products in terms of traditional mapping, as discussed earlier. It is certainly the case that such results can act as an input to cartographic generalisation processes and thus can be used to create an appropriate visual output in the form of a map (digital or paper). The visual comparison of the result against a manually generalised cartographic datasets can indicate the favourability of the results as input for cartographic generalisation (either automatically or manually). Such a comparison can also indicate the appropriateness of thresholds and rules used for creation of the results. Comparison is made more difficult given the subjective nature and thematic variability of maps. As Bertin (1967) illustrated a broad number of interpretations/representations for the same data (Figure 7.10) can exist at lower

levels of detail (Figure 7.11). This is mainly due to the application of different rules and criteria used to generate these results. As pointed out by Müller and Wang (1992) it is more appropriate to check the results against the set of rules used to generate these results.

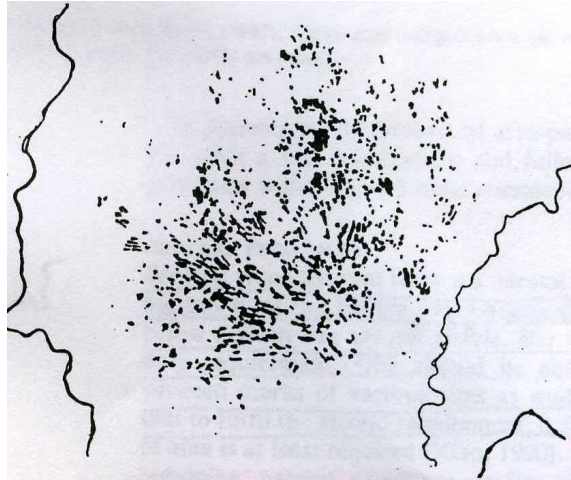


Figure 7.10: Collection of lake patches. Taken from Bertin (1967) scanned from Müller & Wang (1992). The region is Les Dombes, located North East from Lyon (France)

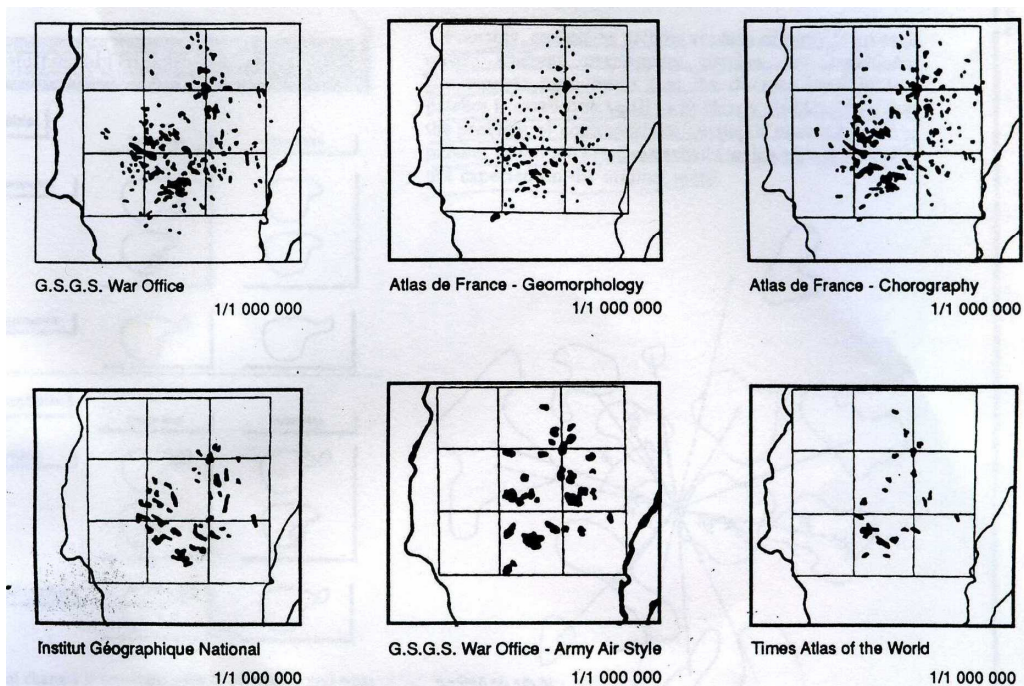


Figure 7.11: Various generalised solutions (at same scale) for data shown in Figure 7.10, from Bertin (1967) (scanned from Müller & Wang (1992))

In this research, the intensions of the classes of the target data model (listed in Table 3.1) was either determined using a set of rules from OS 1:250,000 dataset specifications or by empirical observations using of Strategi dataset as the comparator. The objects of these classes can thus be visually compared against Strategi dataset classes for the same regions. It is important to point out that such a visual comparison cannot be done for hills and ranges since these are not modelled as objects in cartographic products (such as Ordnance Survey Strategi). A visual comparison for each case study with Strategi dataset is presented here.

### **Visual Comparison for Case Study 1**

Figure 7.13 shows the Strategi dataset (settlement, forest, roads and railways) in comparison to the resultant objects shown in Figure 7.12. As can be observed the overall structure for most objects is quite similar. But there are a few differences. The forest objects are different in a few places (ringed in Figure 7.12). There are two possible reasons for this. Firstly, the Strategi dataset is not updated as frequently as the source dataset (OS MasterMap Topography Layer). Thus locations, such as those highlighted in Figure 7.12, are occupied with a high concentration of large tree objects in the source database (Figure 7.1) yet there are no forest objects in OS Strategi dataset at these locations. Secondly, different criteria than the ones used in this research, have been used by the cartographers to create forest objects at locations where there is a very low concentrations of small tree objects (Figure 7.1 and Figure 7.13.). In other words cartographers have worked ‘outside’ the published specifications.



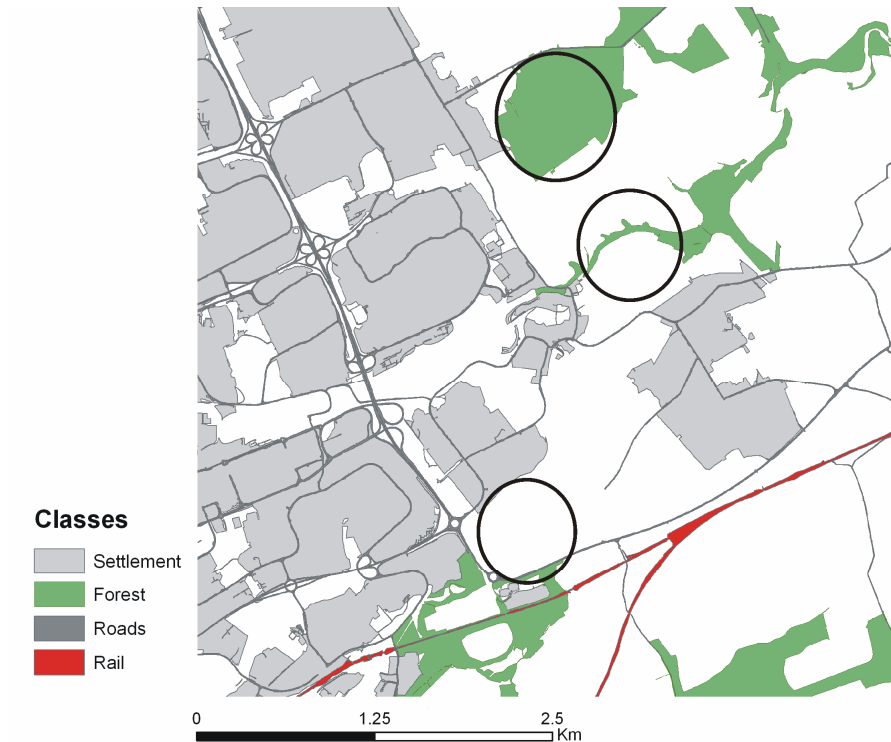


Figure 7.12: Settlement, forest, road and railways objects selected from the resultant database (see Figure 7.3 for comparison). A few differences with Figure 7.13 are ringed

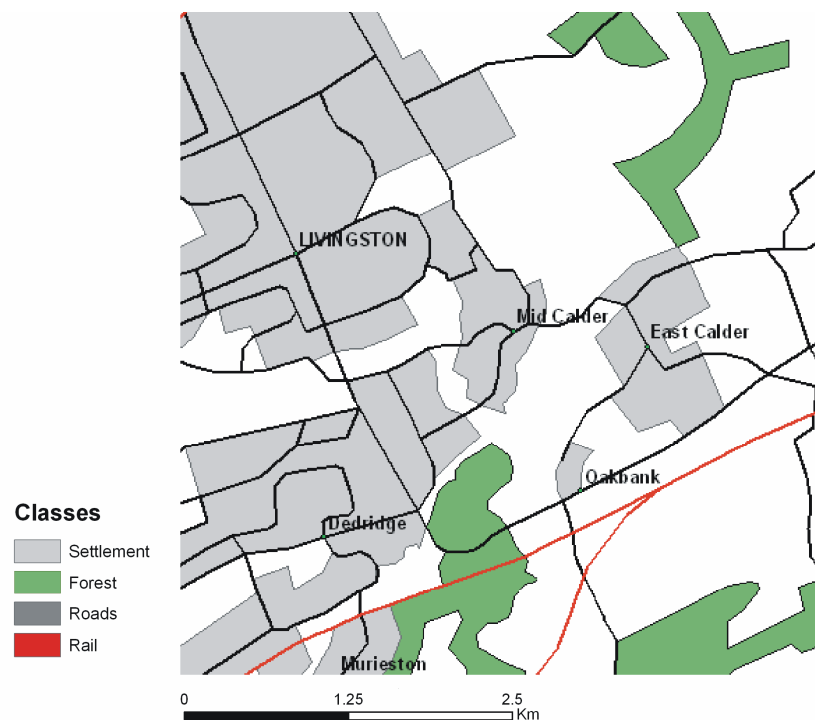


Figure 7.13: Stragati dataset illustrating settlement, forest, rail and roads in the region shown in Figure 7.1 (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved)

Being a cartographic product, geometries of objects in the OS Strategi dataset are made smooth and symbolised as compared to the resultant objects in Figure 7.12. Also the dual carriageway roads have been collapsed into single road objects in the Strategi dataset, whereas no such (cartographic) operations have been applied to road objects in this research. Automated solutions have been proposed in research (summarised recently by Regnauld and McMaster (2007)) that can be applied to the resultant objects to achieve similar results as in the cartographic product.

Note that, in some cases, the settlement objects have been combined with other close settlement objects in Strategi whereas the corresponding settlement objects are shown separate in the results obtained. For instance the 'Dedridge' settlement object in Figure 7.13 has been combined with a settlement object on its right but which is shown as a separate object in the result (Figure 7.12). This might be due to the symbolisation of the two objects. The cartographer may have chosen to exaggerate the objects by enlarging the extent of the towns. Another reason might be that the surveyor has captured additional information for the extent of 'Dedridge' for small scale representation which is not available in the source dataset.

Another noteworthy difference is that certain small settlement objects (for instance Oakbank in Figure 7.13) are present in the Strategi dataset whereas there is no corresponding settlement object in the results. The criteria used in this research for a settlement or forest object to be retained or removed is based on the area of the resultant object (Table 3.1). This criteria was determined using OS 1:250,000 dataset specifications (Ordnance Survey, 2005). The cartographer might have used some additional criteria in order to retain objects below the area threshold. On the other hand in some cases additional objects (settlement and forest) are present in the result (Figure 7.3) as compared to the Strategi dataset (Figure 7.13). This may be due to competition for space at smaller scales in which not all objects in the database can be represented in the resultant cartographic product and are thus removed. Similar observations were made by Harrie and Hellström (1999) in their comparison of automated generalised results with manually generalised cartographic datasets.

## Visual Comparison for Case Study 2

Figure 7.15 shows the Stragegi dataset's settlement, forest, roads and text point layers for the source region shown in Figure 7.4. Figure 7.15 is visually compared with the corresponding resultant objects shown in Figure 7.14. These objects have been selected from the resultant database (Figure 7.6). Visual comparison indicates the corresponding objects are mostly similar. The minor differences in appearance are due to cartographic operations (smoothing, enhancement and collapse) or usage of different rules as compared to the ones used in this research as discussed previously.

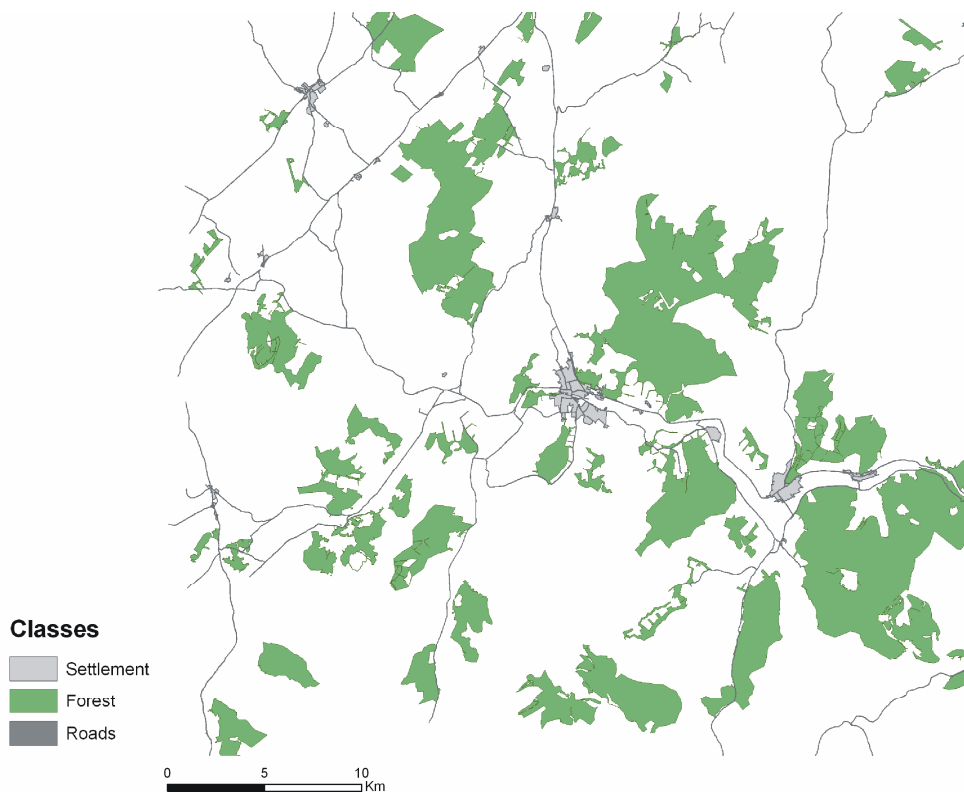


Figure 7.14: Settlement, forest and roads objects selected from results shown in Figure 7.6 for comparison against Figure 7.15

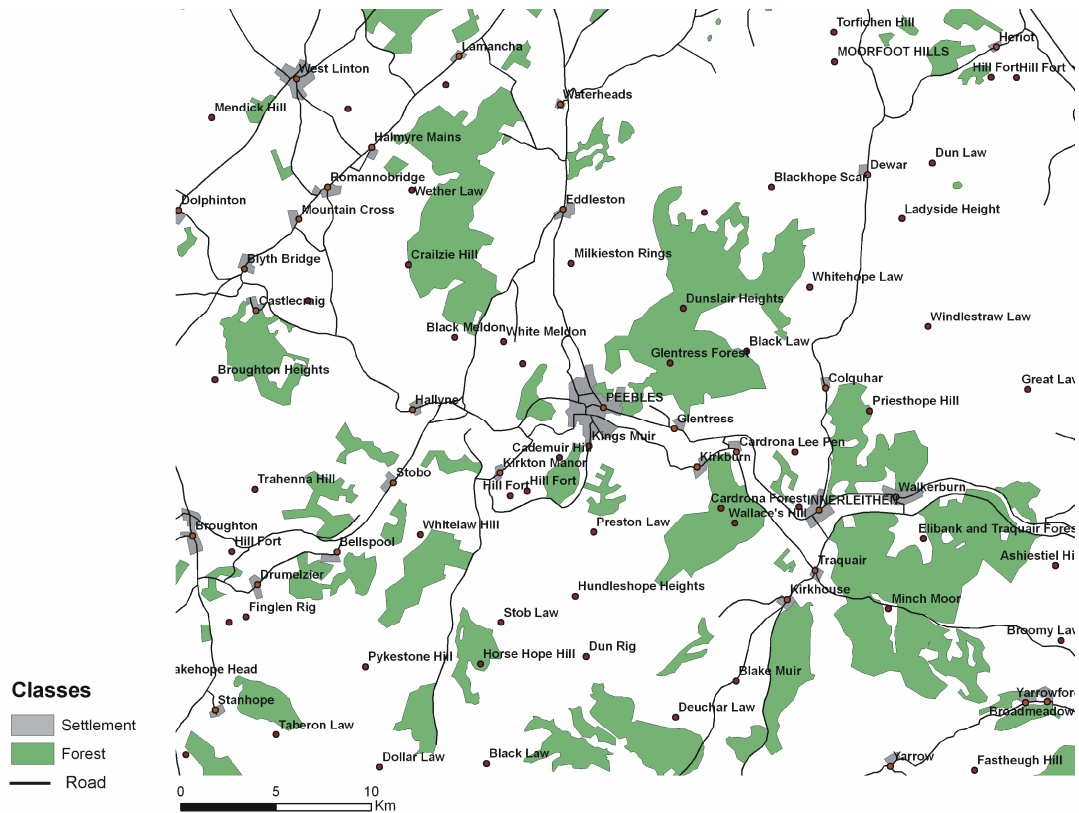


Figure 7.15: Strati dataset illustrating settlement, forest, rail and roads in the region shown in Figure 7.4 (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved)

### Visual Comparison for Case Study 3

Lastly, the visual comparison was done for the Edinburgh region (Figure 7.7).

Figure 7.16 shows settlement, forest, road and railways objects selected from the resultant database for this region (Figure 7.9). This is visually compared against corresponding features selected from Strati illustrated in Figure 7.17. In similarity with the previous two visual comparisons, here too the overall structure of objects is also comparable. The major difference is again due to the cartographic considerations as previously discussed.

Additionally, another noteworthy difference is that additional open spaces (or holes) are present in some settlement objects in the Strati dataset (Figure 7.17), as compared to settlement objects in the results (Figure 7.16) though the corresponding locations are populated with dense buildings (Figure 7.7). This is because of the infrequent updating of the Strati. The results therefore reflect a more up-to-date

view of the real world features. It is important to note that such open spaces are populated by 'General Land' class objects in the resultant database (as shown in Figure 7.9) and there are no holes (geometrically partitioned). This contrasts with Strategi which is not a geometrically partitioned dataset and thus holes are present in settlement and forest objects.

Note that the resultant railway objects in Figure 7.16 are 'disconnected'. This is not due to the result of any operation performed during the generalisation process but due to inconsistency in the source database. This is explained more in section 7.4.



Figure 7.16: Settlement, forest, road and railways objects selected from resultant database shown in Figure 7.9 for comparison

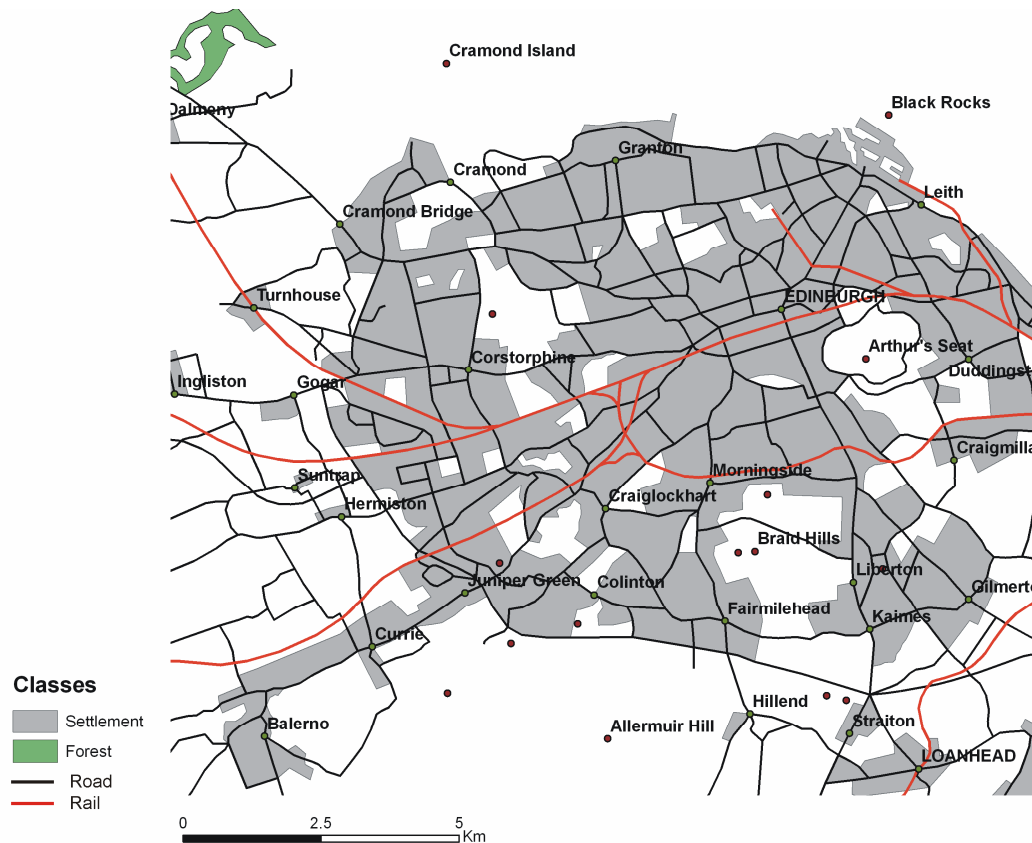


Figure 7.17: Strati dataset illustrating settlement, forest, rail and roads in the region shown in Figure 7.7 (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved)

## 7.2.2 Using text points from cartographic dataset

Evaluation by visual comparison cannot be carried out for hills and ranges because these are not represented as objects in the Strati dataset. But the Strati does contain text points in order to annotate such features. These text points provide an alternative approach for evaluation, as was used by Fisher et al (2004) for evaluation of their landform objects. If the resultant objects are ‘properly’ identified their extents should contain these text points. Two text point layers (settlement\_point and land\_use\_point) were used from the Strati dataset for such an evaluation. The settlement point layer, in the Strati dataset, contains text points to annotate settlements and the land use point layer has text points for naming important hills, ranges and a few large forest objects.

The text points for the three case studies are shown in Figure 7.13 and Figure 7.15 and Figure 7.17 respectively. The evaluation was done by first selecting the text

points for the given region of interest from the two Strategi text layers. The distance for each settlement text point from the closest settlement object in the resultant database was computed. The same process is repeated for forest text points and hill/range text points selected from the 'land use point layer'. The distances calculated, from this process, for the three cases studies (Figure 7.3, Figure 7.6 and Figure 7.9) are given in Table 7.1, Table 7.2 and Table 7.3 respectively.

Table 7.1: Settlement points distances from the closest settlement objects in Figure 7.3

<b>Settlement Point</b>	<b>Closest Target Settlement Object ID</b>	<b>Distance in meters</b>	<b>Result Interpretation</b>
Livingston	Settlement5	0.00	Object found
East Calder	Settlement15	0.00	Object found
Mid Calder	Settlement5	8.52	Due to displacement
Dedridge	Settlement9	45.87	Due to displacement
Murieston	Settlement4	151.41	Due to displacement
Oakbank	Settlement15	630.30	Displacement or Object does not exist

Table 7.2: Text points distance from the closest settlement, forest, hill and range objects in Figure 7.6

<b>Settlement Point</b>	<b>Closest Settlement Object ID</b>	<b>Distance in meters</b>	<b>Result Interpretation</b>
Peebles	Settlement1	0.00	Object found
Innerleithen	Settlement13	0.00	Object found
Eddleston	Settlement2	0.00	Object found
Cardrona	Settlement5	0.00	Object found
Walkerburn	Settlement11	0.00	Object found
Blyth Bridge	Settlement20	0.00	Object found
West Linton	Settlement19	0.00	Object found
Traquair	Settlement12	12.81	Due to displacement
Broughton	Settlement15	54.51	Due to displacement
Halmyre Mains	Settlement24	78.33	Due to displacement
Kings Muir	Settlement1	161.47	Due to displacement
Glentress	Settlement6	195.13	Due to displacement
Romannobridge	Settlement25	823.17	Displacement or Object does not exist
Lamancha	Settlement23	950.75	Displacement or Object does not exist
<b>Forest Points</b>	<b>Closest Forest Object ID</b>		

Glentress Forest	Forest36	0.00	Object found
Cardrona Forest	Forest29	0.00	Object found
Elibank and Traquair Forest	Forest38	125.34	Due to displacement
<b>Hill and Range Points</b>	<b>Closest Hill/Range Object ID</b>		
Dollar Law	Hill2694	0.00	Object found
Byrehope Mount	Hill453	0.00	Object found
Taberon Law	Range3114	0.00	Object found
Lee Pen	Hill1796	0.00	Object found
Black Law	Hill2690	0.00	Object found
Trahenna Hill	Hill1934	0.00	Object found
Pykestone Hill	Hill2334	0.00	Object found
Stob Law	Hill2359	0.00	Object found
Dun Rig	Hill2360	0.00	Object found
Hundleshope Heights	Hill2362	0.00	Object found
Broughton Heights	Range1556	0.00	Object found
Whitehope Law	Hill1228	0.00	Object found
Great Law	Hill1419	0.00	Object found
Preston Law	Hill2027	0.00	Object found
White Meldon	Range1359	0.00	Object found
Whitelaw Hill	Hill2080	0.00	Object found
Fastheugh Hill	Hill2702	0.00	Object found
Black Meldon	Hill1383	0.00	Object found
Blakehope Head	Hill2440	0.00	Object found
Mendick Hill	Hill751	0.00	Object found
Cademuir Hill	Range1884	0.00	Object found
Black Law	Hill1569	0.00	Object found
Wether Law	Hill887	0.00	Object found
Minch Moor	Range2290	0.00	Object found
Windlestraw Law	Range1356	0.00	Object found
Ladyside Height	Range1356	0.00	Object found
Dun Law	Range1357	0.00	Object found
Deuchar Law	Hill2551	0.00	Object found
Hill Fort	Hill1935	0.00	Object found
Broomy Law	Hill2365	0.00	Object found
Dunslair Heights	Range e889	0.00	Object found
MOORFOOT HILLS	Range 889	0.00	Object found
Horse Hope Hill	Hill2446	2.91	Due to displacement
Finglen Rig	Range 3114	21.72	Due to displacement
Priesthope Hill	Hill1677	32.76	Due to displacement
Crailzie Hill	Hill1174	98.51	Due to displacement
Hill Fort	Range1356	415.82	Displacement or Object does not exist
Wallace's Hill	Hill2046	698.52	Displacement or



			Object does not exist
--	--	--	-----------------------

Table 7.3: Text points distance from the closest settlement, hill and range objects in Figure 7.9

Settlement Text Point	Closest Settlement Object ID	Distance in meters	Result Interpretation
Turnhouse	Settlement17	0.00	Object found
Loanhead	Settlement29	0.00	Object found
Juniper Green	Settlement15	0.00	Object found
Colinton	Settlement15	0.00	Object found
Fairmilehead	Settlement15	0.00	Object found
Kaimes	Settlement15	0.00	Object found
Gilmerton	Settlement15	0.00	Object found
Craiglockhart	Settlement15	0.00	Object found
Morningside	Settlement15	0.00	Object found
Corstorphine	Settlement15	0.00	Object found
Duddingston	Settlement15	0.00	Object found
Dalmeny	Settlement5	0.00	Object found
Craigmillar	Settlement28	0.00	Object found
Currie	Settlement15	0.76	Due to displacement
Balerno	Settlement1	1.11	Due to displacement
Granton	Settlement15	1.31	Due to displacement
Hermiston	Settlement10	3.96	Due to displacement
Edinburgh	Settlement15	7.90	Due to displacement
Straiton	Settlement29	26.83	Due to displacement
Leith	Settlement15	27.85	Due to displacement
Liberton	Settlement15	100.44	Due to displacement
Cramond Bridge	Settlement15	195.62	Due to displacement
Bilston	Settlement29	309.67	Displacement or Object does not exist
Cramond	Settlement15	384.16	Displacement or Object does not exist
Hill and Range Points	Closest Hill/Range Object ID		
Allermuir Hill	Range423	0.00	Object found
Arthur's Seat	Range218	0.00	Object found
Braid Hills	Hill284	763.96	Displacement or Object does not exist

It was observed from these tables that 71% of all text points are inside the resultant objects (distance =0) and 90% are within 200m of the closest resultant object. Text points such as 'Murieston' (distance =151.4m Table 7.1) or 'Glentress' (distance

=195.1m in Table 7.2) or 'Liberton' (distance =100.44m Table 7.3) are quite far (within 200m) from the closest resultant object. This is because the text point layers are part of a cartographic product (Strategi) so these text points have been subject to displacement in order to maintain clarity in the resultant map.

Certain text points such as 'Oakbank' (distance=630.30m Table 7.1), 'Preston Law' (distance=301.12m Table 7.2) or 'Cramond' (distance=384.15m Table 7.3) are hundreds of meters from the closest resultant object. This might again be due to displacement of text points. Alternatively for such text points, that are so far from the closest object, there may be no corresponding objects in the resultant database. This may be due to some additional criteria besides minimum area (for settlement and forest) and prominence (for hills and ranges) that has been used by the cartographer. Thus the distances are very large in certain cases. The important thing to note here is that in the Strategi there is no link between the text points and the places they represent. There is no inherent understanding of the link between the feature and the name. But once the objects such as those generated here are identified, they can then be used to create such links. This will help the cartographer in making decision on how much displacement can be tolerated. Such information will be useful for automatic text placement in cartographic generalisation (Barrault, 1995; Petzold *et al.*, 2003).

The results from this methodology were used by an algorithm to link labels to resultant objects. The algorithm simply selected the text points that are 'close' to the resultant objects (settlement, forest, hills and ranges). The distance tolerance was determined empirically and was set to 200m. The selected text points and the selected composite object's id were then stored in a database table. The database table is designed such that more than one label can be assigned to a resultant object. This is because as observed in the above tables it is possible for a resultant object to be associated with more than one text point. This proved useful in carrying out more meaningful spatial analysis on the results (section 7.3.1).

### 7.2.3 Quantitative Evaluation

Another possible way of evaluation is via quantitative comparison between the total number of objects in the source database and total number of objects in the target databases. This is termed the ‘global reduction factor’ (Richardson, 1993). Table 7.4, Table 7.5 and Table 7.6 show the total number of objects in the source and the target database for the three case studies illustrated in Figure 7.1, Figure 7.4 and Figure 7.7 respectively.

Table 7.4: Total number of objects in the source database, number of objects per class in the resultant database and in the Strategi for the case study shown in Figure 7.1

Total Number of objects in source database		55,000
Output Class	No of Objects	No of Objects in OS Strategi
Settlement	7	4
Forest	5	3
Hills/Ranges	0	0
Roads	93	80
Rail	5	8
General Land	2	(as holes) 1
<b>Total</b>	<b>112</b>	<b>96</b>

Table 7.5: Total number of objects in the source database, number of objects per class in the resultant database and in the Strategi for the case study shown in Figure 7.4

Total Number of objects in source database		76,000
Output Class	No of Objects	No of Objects in OS Strategi
Settlement	25	32
Forest	51	55
Hills/Ranges	85	(text points) 46
Roads	166	150
Rail	0	0
General Land	13	(as holes) 10
<b>Total</b>	<b>340</b>	<b>291</b>

Table 7.6: Total number of objects in the source database, number of objects per class in the resultant database and in the Strategi for the case study shown in Figure 7.7

Total Number of objects in source database		400,000
Output Classes	No of Objects	No of Objects in OS Strategi
Settlement	8	10
Forest	1	1
Hills/Ranges	8	(text point) 3
Roads	623	486
Rail	73	60
General Land	5	( as holes) 16
Total	718	576

These tables illustrate that the proposed methodology has reduced the number of objects, from source database to target database in each case study, significantly. The results in these tables enable us to compare the resultant number of objects obtained per class against the number of objects in the corresponding class in the Strategi dataset. The number of objects for each class in the resultant database is quite close to the number of objects in Strategi. The small differences are due to the reasons as explained in the previous section. A major difference between the number of objects is in the case of road objects. In addition to the cartographic reasons (discussed in section 7.2.1), there may be two further reasons that account for this discrepancy. Firstly, the road objects in the Strategi dataset are in the form of polylines whereas in the results they are in the form of area objects. Thus junctions are presents as separate objects in the results whereas no such objects are present in Strategi (Figure 7.18). Secondly, in this research, objects of roads classes (Motorway, A road, B road, Minor Road) were selected and aggregated based on their classification in the source database. Whereas in Strategi dataset roads are selected or eliminated based on a variety of criteria, including visual constraints, besides classification. For instance a B road object is not represented where the alignment of the road coincides with the alignment of other topographic feature (Ordnance Survey, 2005). These additional constraints result in the number of roads being greater than the number of road objects in Strategi. Such constraints could be applied as an extension of the proposed methodology.

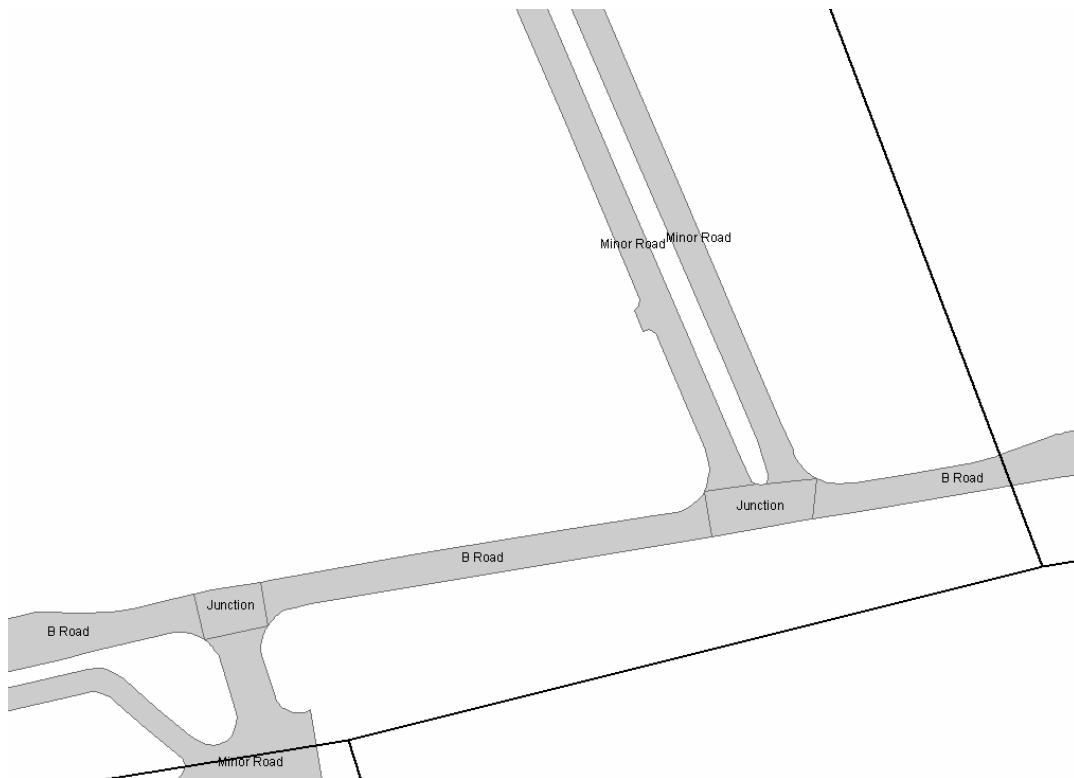


Figure 7.18: Differences between road objects in the resultant database (in grey polygons) and in Strategi (blank polylines). Junctions are not present as a separate object in Strategi, dual carriage ways have been collapsed into single road polylines in Strategi in contrast to objects in the results

Table 7.7 compares the cumulative area of settlement and forest class objects for each case study against the cumulative area of corresponding objects in Strategi dataset. This comparison illustrates the similarity between the extents of the resultant objects with the extent of similar objects in Strategi. Such a comparison is only possible for settlement and forest class because objects of these two classes are modelled as polygons both in the results and in Strategi. The table illustrates that cumulative area of objects in results closely match with the total area of objects in Strategi dataset. The minor differences are due to the cartographic reasons (exaggeration and enhancement of objects) as explained in previously. The major difference is in the total area for settlement objects in case study 2. This is because Strategi contains more settlement objects in this region as compared to objects in the results (Table 7.5).

Table 7.7: Comparison of cumulative areas for settlement and forest class objects in the results and strategi for the three case studies

Class and Region	Total Area in Results (km <sup>2</sup> )	Total Area in Strategi (km <sup>2</sup> )
Settlement (case study1)	9.19	10.23
Forest (case study 1)	2.55	2.56
Settlement (case study2)	4.77	7.60
Forest (case study2)	150.71	150.17
Settlement (case study3)	77.76	85.15
Forest (case study3)	0.41	0.66

### 7.3 Utility of Results

Making comparison with manually generalised datasets can give some indication of the quality of the proposed approach. But it is important to point out that comparison with a manually generalised dataset produced by cartographers is a subjective issue (João, 1998; Mackaness & Ruas, 2007; Weibel & Dutton, 1999). Trying to replicate human cartographic products runs the risk of creating ‘cartographic kitsch’. This is because it is based on the belief that cartographers have a superior knowledge of generalisation; an assumption that is seldom questioned (Harrie, 2001; Schylberg, 1993). Often the results of manual generalisation are themselves not up to standard (João, 1998). Visual assessment of automatically generalised results would be more appropriate if evaluated against other automatically generalised set of results of the same region (Harrie, 2001). Unfortunately no automatic solutions exist at this time for the classes and datasets used in this research.

A better way of assessing the results, and especially for model generalisation results, is to assess the results in terms of their functionality (Harrie, 2001; Mackaness & Ruas, 2007). By functionality we mean the ability of the proposed approach to satisfy user needs that were not previously possible using the source database alone. The next section presents the different functional utilities that can take advantage of the approach presented in this thesis. They illustrate the potential use of both the enriched source database in terms of partonomic relationships and also the use of the resultant database besides input to cartographic generalisation.

### 7.3.1 Spatial analysis

The utility of an enriched database lies in its ability to support spatial analysis routines that were not possible using the original source database. For instance a user might wish to know which buildings are part of a particular settlement. It is not possible to generate the result of such a query directly from the source database because that database does not have the required settlement containers and (partonomic) relationships modelled explicitly. Although we have a building class in the source database, the objects it contains have no information relating as to which city or settlement they are part of. But a simple spatial query (shown below) can be performed (using SQL) once the source database has been enriched with these partonomic relationships. In the following query all buildings that are part of the settlement 'East Calder' are selected from the source database. This is possible because the source database objects have been updated with partonomic relationships in terms of composite objects. And a name table has been created that stores the name for a composite object. The composite object id ('East Calder' in this example) is selected from the name table and corresponding building objects are selected from the source database that are linked to the composite object id via the partonomy table. The result from the query is visually presented in Figure 7.19.

```
SELECT      a.Geometry
FROM        Source_database a, Partonomy_table b, Name_database c
WHERE       a.object_id=b.objet_id and a.descgroup='Building' and
b.composite_object_id=c.composite_object_id and c.object_name='East Calder';
```

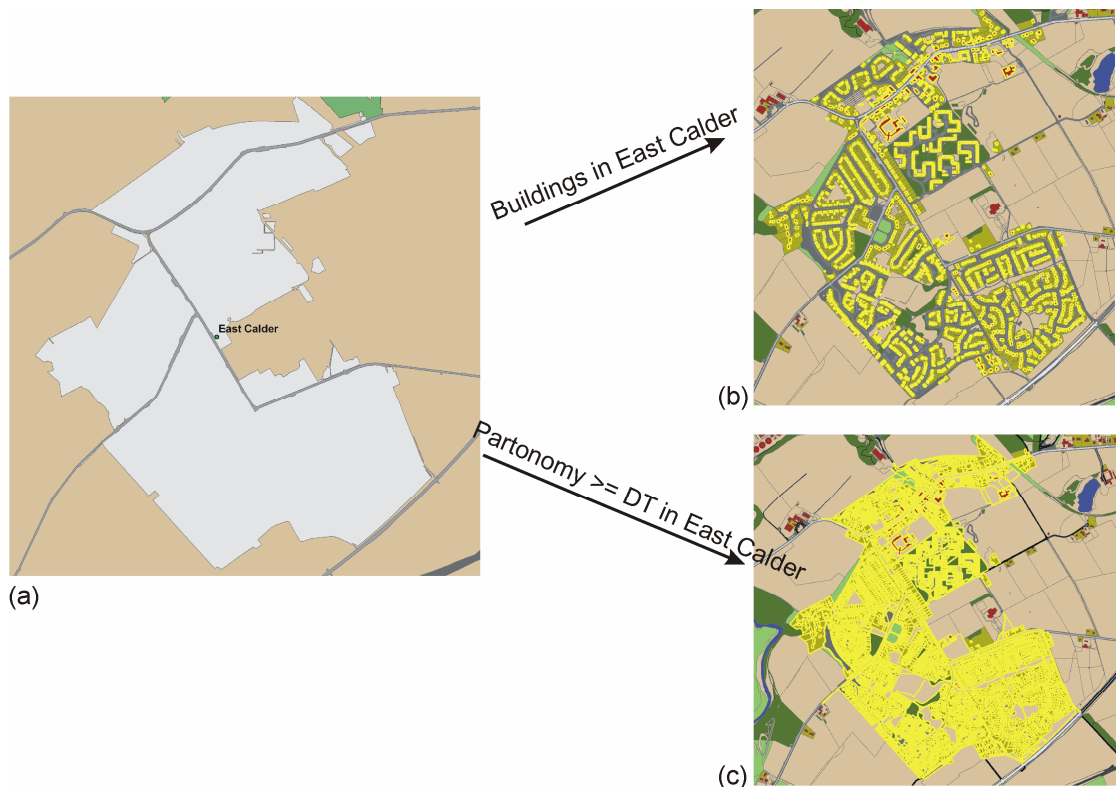


Figure 7.19: Example of spatial analysis on the enriched source database. (a) Resultant database. (b) All buildings in East Calder selected in source database (in yellow). (c) All objects that are part of East Calder with degree of parthood above or equal to minimum degree of parthood threshold (DT) used during aggregation stage (section 6.2.3)

Similarly such queries can also use the degree of parthood as a further selection criterion. For instance Figure 7.19c shows the result of a spatial query (shown below) that returns all objects (not just buildings) from the source dataset that are part of 'East Calder' with a degree of parthood greater or equal to minimum degree of parthood threshold (DT) used during the aggregation stage (section 6.2.3). Note that both of these queries could be directly performed on the enriched source database, without having a name table, but in such a case we have to know the composite object id for the settlement instead.

```
SELECT      a.Geometry
FROM        Source_database a, Parthood_table b, Name_database c
WHERE       a.object_id=b.object_id and b.degree_of_parthood>=65 and
           b_composite_object_id=c.composite_object_id and c.object_name='East Calder';
```



More sophisticated analysis such as determining the shortest road network between different resultant objects can be carried out and visualised using the proposed approach. The shortest road path between 'East Calder' and 'Livingston' is highlighted in red in Figure 7.20. This is the result of an algorithm that uses the enriched database, the name table and OS Integrated Transport Network (ITN) in order to generate the result. Firstly two road objects are selected from the source database using SQL queries similar to those shown above. One of these road objects is part of the start location ('East Calder' in the given example) and one part of the end location ('Livingston' in the given example). From these two road objects, corresponding network objects (ITN) are selected. These are passed to a function that implements a shortest path algorithm (Dijkstra, 1959) and uses a road network (ITN) to retrieve the shortest path (shown in red in Figure 7.20) connecting the source destination to target destination. Again, such a query is not possible in the original dataset because of the lack of required links (part of relationship) between source database road objects and the settlement. That is to say ('East Calder' and 'Livingston') objects that are referred to in the query. It is only once the extent of these objects has been identified that this type of spatial analysis can be performed and the results visualised.



Figure 7.20: Example of spatial analysis. Finding the shortest road path (highlighted in red) between 'Livingston' and 'East Calder'

These spatial utilities illustrate that for many applications the required information is implicit in the detailed source data and needs to be made explicit. In other words more data or detailed data does not necessarily mean more information the required information might be hidden within the data. These examples also illustrate that there is an inherent scale or level of detail associated with a spatial query. Hence the need of spatial data at different levels of detail. In these examples, we have demonstrated the utility of a *single* parthood. The next section presents the utility of *multiple* parthoods in order to create objects that are instances of other composite classes.

### 7.3.2 Identification and Enrichment of new Composite Classes

The multiple parthoods, generated from the overlapping container boundaries, can be used to aggregate objects of composite classes that were not initially defined in

the target data model. For instance, from the enriched database we are able to select source objects that are part of a hill as well as a forest. The selected objects can then be aggregated into a composite object instance of a ‘hilly forest’ composite class (Figure 7.21). Similarly instances of other types of composite objects, that are different combinations of existing composite classes, can be determined by aggregation using multiple paronomies. These composite classes can be settlement and forest, hilly and settlement, or hill, settlement and forest composite class. This enables us to identify leafy suburbs with nice views across the city.

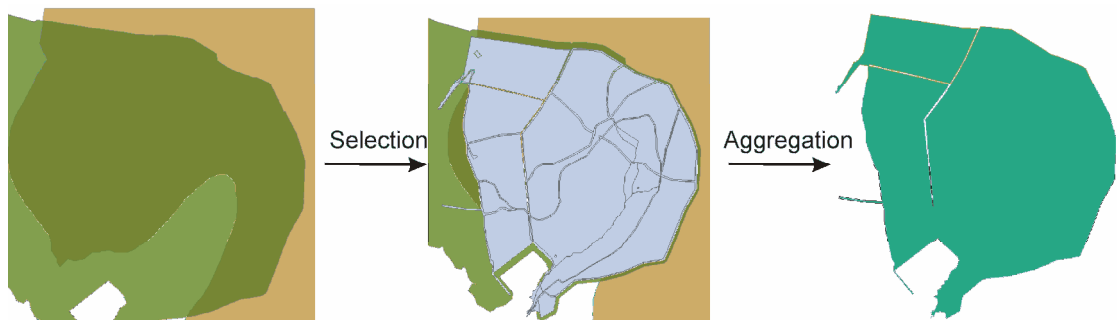


Figure 7.21: Creation of instance of new composite classes that are combinations of existing composite classes. The figure illustrates the creation of hill\_forest (composite class) object via the aggregation of objects that are part of a hill as well as a forest

### 7.3.3 Creation of links for MRDBs

In Chapter 3, MRDBs were introduced and their importance in terms of spatial analysis, creation of customised maps and for cartographic outputs at various levels of detail was highlighted. The chapter also pointed out problems associated with creation of MRDBs by establishing links using existing cartographic datasets. The more appropriate approach, as discussed earlier in Chapter 3, is generalisation of the source database by model generalisation techniques. The partonmoic relationships, such as those identified in this research, serve as links between objects in the source database (Mustière & van Smaalen, 2007). Once the links between source objects and target objects have been identified these can be used for database updates such that only the objects in the source database need to be updated. The target database object is automatically and consistently updated using the methodologies presented here. Aggregation of source objects into composite objects based on these

partonomic relationships effectively creates a multi-resolution database that allows navigation through different levels of abstraction.

It should be pointed out that in this research each source database has two representations: one at the source database level and one at the target level of detail (1:250,000). Other higher levels of abstraction can be generated using the resultant database (discussed in section 7.6).

Having discussed the evaluation and utilities of the results the next two sections (7.4 and 7.5) present a few errors and limitations of the proposed approach.

#### **7.4 Generalisation errors due to inconsistencies in the source database**

As stated in Chapter 4 the quality of the data in the source database affects the quality of the resultant database. This we would expect given the process of derivation. An example is illustrated in Figure 7.22. Here the rail objects in the resultant database are disconnected. This 'dis-connectivity' is not caused by some operation in the proposed approach but because of the nature of the data capture process performed by OS. As a result, the corresponding rail objects are not topologically structured in the source database. Research has been undertaken to create topologically connected network objects (Regnauld & Mackaness, 2006). Unlike roads, at locations where rail and road objects meet no separate junction or object is present in the source database. Thus either the road or the railway is broken. The resultant road or rail objects at these places are disconnected.

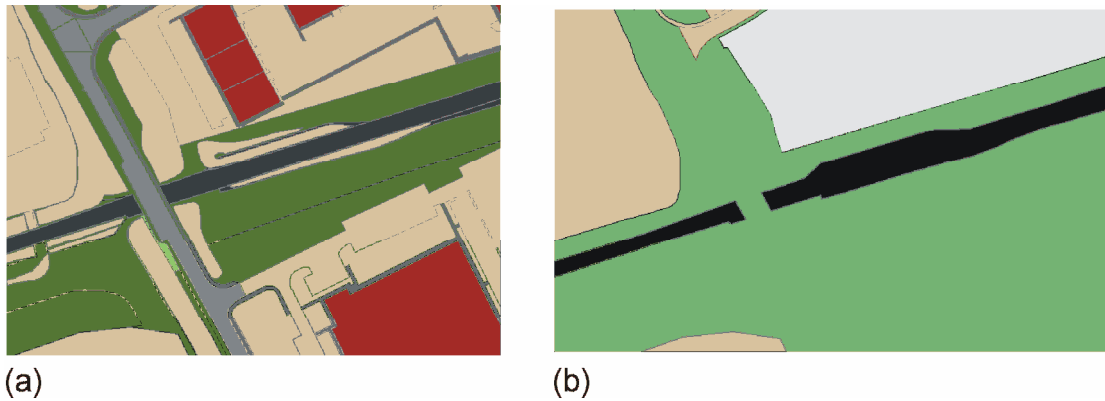


Figure 7.22: (a) Railway objects (black polygons) in the source database are disconnected (b) The resultant rail objects are thus disconnected in the target database

A precursor to the creation of partonomies would be the creation of topologically ‘correct’ network type area objects in the source database.

### 7.5 Degree of Fuzziness in Results

Although the composite objects created in this research have been modelled as objects with ‘crisp’ boundaries, we acknowledge that there is a degree of vagueness or fuzziness in the extent of the resultant composite objects. As discussed in Chapter 5, (section 5.1.1) all the container boundaries identified for composite class objects in this research are of ‘fiat’ type. The objects created from these boundaries are thus fiat objects. These fiat boundaries are by definition, the result of human cogitation (Smith & Mark, 2003) or the result of some measurement or estimation (Bian, 2007). The fuzziness is inherited in these because of the fuzziness in the concept they represent (Usery, 1996). As pointed out by Smith and Mark (2003) and Molenaar (1993), geographic phenomena often have a certain level of indeterminacy reflecting the need to define and understand the phenomena within a certain context of observation. The context and the resultant application define the parameters and rules that are used to determine the objects representing these concepts. And these parameters and rules should thus be understood within the context of use.

A lot of research has been done on dealing with fuzziness in spatial objects, crisp and non crisp boundaries (Burrough & Frank, 1996; Campari, 1996; Duckham *et al.*, 2001; Duckham & Sharp, 2005; Fisher, 2000; Goodchild *et al.*, 1998; Miyamoto,

1990; Molenaar, 1993; Molenaar, 1996b; Reinke & Hunter, 2002; Winter & Thomas, 2002). But in generalisation research, fuzziness has not received much attention. This may be due to the fact that most of the research in generalisation has been motivated from a cartographic perspective and less as a modelling process. But this fails to acknowledge cartography as a modelling process. It is in part the fuzzy nature of spatial objects that requires us to model and view the geographic phenomena at multiple levels of detail. At 1:250,000 we can create approximate boundaries as distinct and clear boundaries to convey a region – a boundary that cannot and should not be verified, precisely because of the scale of representation. Fuzziness in the results should be modelled as part of the results. Several researchers have presented different techniques for modelling fuzziness in terms of a spatial object's geometry using vector and raster data structures, modelling their relationships and also they have proposed aggregation models that combine several fuzzy sets into a single fuzzy set (Cohn & Gotts, 1996; Molenaar, 1996b; Mustière & Moulin, 2002; Robinson, 2007; Winter & Thomas, 2002). These techniques need to be incorporated as part of generalisation and data modelling processes in order to model fuzziness in the higher level, fiat objects.

## 7.6 Further Discussion

The partonomic relationships determined in this research are not only important for spatial analysis and database transformation. They can also be used for different representations in creating cartographic outputs. For instance, a major road might be modelled in a different way if its part of a city (servicing the daily commute) as compared to its role in a rural setting – in which the road more serves to *connect* cities. These different behaviours can result in different cartographic visualisations.

The approach presented here is completely automated. The parameter settings and thresholds used in different stages of the methodology are set only once and were kept constant for all the case studies. The values for most of the thresholds and parameters were derived from existing rules (Ordnance Survey, 2005) for the same classes, whilst for some composite classes such as hills and ranges were determined empirically. The algorithms developed and the parameter and threshold settings can

be further tested by using other sources of data. The minimum requirements for the dataset are; it forms a thematic and geometric partition of the region of interest and objects can be classified to basic classes such as building, tree, railway and different type of roads. Such evaluation techniques can be carried either by acquiring new dataset or by converting the algorithms into *web services*. Several researchers, in recent years, have highlighted the importance of making the generalisation algorithm available as web generalisation services (Burghardt *et al.*, 2005; Edwards *et al.*, 2007; Foerster & Stoter, 2006; Neun & Burghardt, 2005). These services provide the required functionality through a web-accessible interface in a platform and programming-language independent way. These services allow sharing of generalisation algorithms not only within the generalisation research community but their application in other areas of research in GIS such as web mapping, geo-visualisation and location based services (Edwards *et al.*, 2007). Example of a few web generalisation services is available at:

[http://www.ixserve.de/pub\\_whatiswebgen.php](http://www.ixserve.de/pub_whatiswebgen.php). These services would allow comparison of the results with other automated solutions.

The results are completely reproducible using the proposed approach, provided the same datasets are used. The results can be further used for creation of a different representation of the same geographic phenomena. For instance, an object representing a settlement object has been created (in the form of a polygon) from the source database objects using this approach. A simple spatial function (shown below) can be applied on this object that creates the centroid of the polygon. This centroid (point object) is another representation of the same object but at a different level of detail (such as 1:1m). Such a point object is not only important in terms of text and symbol placement for a small scale map, but more importantly such a point carries the same semantic information as is carried by the polygon object from which it was created. This means that it is not just a geometrical point but has meaning in terms of the objects that belong to this settlement (Figure 7.23). In Figure 7.23 the two 'dots' with the word 'London' written next to them are the same in terms of geometry but are quite distinct in terms of the semantic information explicitly stored with them.

```
Select SDO_GEOM.SDO_CENTROID(a.geom,0.000005)
From output_table a , name_table b
Where a.ID= b.OBJECT_ID and b.OBJECT_NAME= 'London'
```

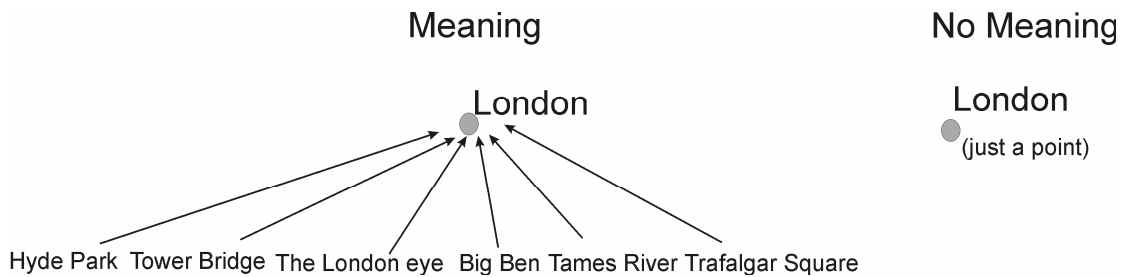


Figure 7.23: What does 'London' mean?- A phenomenological view much closer to our concept of London

Ideas of semantic modelling, and the ability to characterise the saliency of objects at different levels of detail is critical to the interpretive process. When the map reader sees a dot with the word 'London' next to it, they understand in an instant, what that dot represents, together with all the processes and phenomena that are contained within it (Mackness *et al.*, 2007). That 'dot' should be modelled with all these properties so that systems support a whole set of meaningful queries and analysis techniques. Such modelling is possible via the proposed methodology as presented by the above example.

## 7.7 Summary

This chapter presented results of the proposed methodology across different regions selected from the source database. The results were visually compared against manually generalised cartographic datasets. The visual comparison illustrated the appropriateness of the results as an input for cartographic generalisation. The results were also evaluated by calculating distance between text points, in existing cartographic products, from the composite objects of the results. Quantitative evaluation illustrated the global reduction in the total number of objects in the source database and target database. This evaluation also compared the number of objects for each class with the number of objects in existing cartographic datasets. These



evaluations highlighted the difference between model generalisation results and cartographic results.

The chapter presented the utility of the results in terms of different spatial analysis routines. The chapter argued that the partonomic relationships between source and target objects serves as links for MRDBs. The chapter highlighted the opportunity to model fuzziness and that this could be an important ingredient in future map generalisation methodologies. Lastly, the chapter presented the utility of the results in the creation of the higher abstractions of the same geographic phenomena using the results. The semantic information carried by the results would be passed to the higher abstraction. This would offer meaningful analysis and support ideas of 'intelligent zoom' (Frank & Timpf, 1994).

## CHAPTER 8: Conclusions and Future Work

Map generalisation is more than a process of creating aesthetically pleasing maps automatically. It deals with the understanding the relationships between geographic phenomena at different levels of detail. It provides ways of converting the phenomena at one level into a representation at a much higher level of abstraction. It is about preserving the salient qualities of spatial data as the level of detail changes. The main motivation of this research was to develop a model generalisation approach for a direct transformation of spatial database over large changes in level of detail. A detailed spatial database does not necessarily mean more information for every application. For many applications the required information is implicit in the source data. There is a requirement for spatial data at different levels of detail in order to discern different properties and patterns of geographic phenomena. Prior to cartographic portrayal, model generalisation is needed to transform a spatial database from one level of detail to another level of detail.

### 8.1 Summary of Thesis

This thesis has presented an automated model generalisation approach that aggregates objects of the source database into objects of the required classes. In this research the classes of the target data model were determined by analysis of 1:250,000 scale datasets. These included settlement, forest, hills, range, rail and important road (Motorway, A road, B road, minor Road, junction) classes. Except for rail and road classes all other were composite classes. These classes are combinations of different source (component) classes which are not essentially similar, i.e., they belong to different classification hierarchies. The objects of these composite classes in this research were created via aggregation of source database objects based on partonomic relationships.

The *partonomic* or functional relationships link objects of different classification hierarchies into a composite object. Such relationships were not explicit in the source database and required database enrichment. A set of methodologies were presented in this research, for each composite class, in order to determine the

partonomic relationships for source objects in terms of the objects of the required classes. Once the source database was enriched with these relationships, the required composite objects were created via aggregation operation using specific rules. These rules ensured topological consistency of the resultant database.

Objects of non composite classes such as rail and road classes, were selected and aggregated from a source database based on thematic similarity and adjacency principles. The database obtained from this approach was both thematically and geometrically partitioned, in which there are no holes or objects of unknown classes in the target database. Each object in the source database had a representation at the target level of detail.

The algorithms are implemented in Java. Oracle Spatial 10g provided the database management system (DBMS) for storing spatial object selected from the source dataset. It also provided spatial functions required in the research. The DBMS allows retrieval of spatial as well as thematic properties of the data to be queried simultaneously, as opposed to most current GISs that require separate queries for spatial and thematic data. Objects in the database are uniquely identified by an object identifier. Composite and component objects are related through object identifiers, enabling creation of a dataset with multiple representations.

## 8.2 Major Achievements

The major findings of this thesis are:

- *Partonomic relationships can provide the basis for aggregation over large changes in levels of detail*

An aggregation approach based on partonomic relationship is different from the many aggregation approaches currently proposed. Such aggregation methods are usually based on thematic similarity or on taxonomic relationships. Such relationships only allow aggregation of objects belonging to the same classification hierarchies. They are thus limited to small changes in the level of detail. Because of large changes in levels of detail, objects of dissimilar classes

may be next to each other and need to be aggregated in order to create the objects of the required classes. The aggregation approach presented in this research allows for the combination of dissimilar objects using their partonomic relationships. This research has explored the utility of *partonomic* or functional relationships for aggregation of source objects into target objects. Such relationships allow objects of different classes, classes belonging to different taxonomies, to be aggregated into objects of the required composite classes. The methodology presented in this research for determining partonomic relationships and using these for aggregation is unique. The methodology tested using Ordnance Survey datasets, but can be applied on other thematically and geometrically partitioned datasets with basic feature classification attributes.

- *Partonomic relationships provide the basis for creating links within MRDB*

Research in MRDBs is concerned with connecting different representations of the same geographic phenomena at different levels of details. This research has presented an approach based on generalisation of a single source database. Each source object has an explicit link, in terms of its partonomic relationship, with a higher order object. Each higher order or composite object has a geometric representation which has been created by aggregation of source objects. The resultant objects could be used to create representation at even higher abstraction levels using simple spatial functions. In this way we have multiple representations of the same geographic phenomena. The explicit linkage between the representations ensures consistency, automatic updating, increased efficiency and customised outputs.

- *Modelling partonomies affords more effective retrieval of spatial information*

The resultant objects can be automatically linked with place names using a gazetteer. A key benefit of a gazetteer is that it can give access to information based on spatial relevance not just thematic relevance. The partonomic relationships and the results identified in this research can be used to answer such spatial queries. For instance finding the shortest road path between two towns, selected from the gazetteer, has a certain scale associated with this information.

Such a query needs to know the extent of these two cities and in addition which roads are parts of these cities. This research fulfils requirements for such spatial queries. This approach can also be used for answering web searches that involve spatial regions and relationships. For instance “all restaurants in Edinburgh close to Arthur Seat” requires detection of extents of Edinburgh (a city) and Arthur seat (a hill) and database enrichment in terms of spatial relationship between the restaurant objects in the source database and these extents.

- *Modelling multiple partonomies enable creation of higher level geographic phenomena*

Partonomic relationships are non exclusive. Modelling multiple partonomic relationships for source database objects allows more detailed spatial analysis. Also these multiple partonomic relationships are useful for creating instances that are combinations of existing composite classes. For instance using these relationships we can identify and create a spatial object that is an instance of a ‘hilly forest’ class (combination of hill and forest classes). Such routines allow extension of target data models with new sets of classes.

- *Model generalisation as an essential Pre-requisite to Cartographic generalisation*

Over small changes in scale cartographic generalisation techniques can be directly applied on the source database in order to create a cartographic output. But over large changes in scale or levels of detail there is fundamental change in the concepts present at source and target scales. Thus transformation of the database becomes an essential prerequisite. Once the source database objects have been transformed into instances of required concepts, the cartographic generalisation can focus on making the output ‘aesthetically’ appropriate. The results obtained from the methodology can be used to as input to a cartographic generalisation process (carried out manually or automatically). This would involve application of cartographic operations such as exaggeration of small objects, symbolisation of objects, their enhancement and displacement of symbols or objects that are too close.

### 8.3 Future work

This thesis has contributed to the on going research of map generalisation. However, there are still areas that need to be investigated and some of the aspects treated in this research can be further refined. They are summarised as follows:

- **Expanding the number of classes in the target data model**

The target data model can be expanded further in terms of classes that are found at notational scale of 1:250,000. For each new composite class a methodology would be required in order to identify the partonomic relationships for source objects in terms of required objects of the new classes. Once the relationships have been identified the proposed aggregation approach would be used for the creation of geometries of target composite objects. Each new super or parent class aggregation can be based on taxonomic relationships.

- **Evaluating Techniques**

The systematic evaluation of the results obtained from this methodology is in need of additional research. Whilst a few techniques were presented illustrating the success of the approach these are limited for detail evaluation of model generalisation results. Future work needs to look into other evaluation techniques perhaps based on cognitive studies (Steiniger *et al.*, 2006). It is suggested that it would be more appropriate to evaluate the result against result from another automated model generalisation technique that generates objects of similar classes. Clearly such an evaluation would be much easier to perform if the proposed techniques are available as web generalisation services.

- **Developing Generalisation Services**

Future work would look into making the algorithms implemented for this research available as web generalisation services. Such development would open a new avenue for comparison and evaluation of the results and the proposed techniques against the other automated solutions. This would also demonstrate

the flexibility of the approach with different setting for the threshold and parameters used as set by the user of the service.

- **Detecting implicit patterns within Settlement class**

Implicit patterns within city or settlement class can be further investigated. This would involve extension of the hill and range container boundary methodology (section 5.4) using a settlement's density surface (section 5.2.2) instead of a DTM. This would reveal patterns within a settlement object such as city centre, districts, and suburban area. These would facilitate spatial analysis as well as generalisation strategies for other levels of detail (for example at 1:50,000 and 1:100,000).

- **Modelling fuzziness in the results**

Objects of composite classes, considered here, are derived using specific statements. Such statements have an inherent degree of indeterminacy or fuzziness. This is because of the fiat or fuzzy nature of the concept. The need is to model the degree of fuzziness as part of an object's attribute, geometry or fuzzy relationship between objects. This would facilitate both analysis and rule or constraint settings during the process of generalisation.

- **Extending network generalisation algorithm**

To ensure topological consistency of resultant network objects in the database. This would involve integration of a structural generalisation approach with the current methodology.

- **Ontology Driven Generalisation System (Chapter 8 Future work)**

Ontological specification of geospatial concepts will allow the sharing of knowledge across domains. This will also facilitate the process of generalisation as well. Firstly if the concepts modelled by a generalisation algorithm are expressed via ontologies, then it would allow the sharing of code. This is because the implicit assumptions made would be expressed more formally

(Edwards *et al.*, 2007). Secondly it would lead to development of ontology driven generalisation systems (Kulik *et al.*, 2005; Lüscher *et al.*, 2007; Regnaud, 2007). Such system would be able to derive spatial results from different databases based on user specified ontologies or specific task oriented ontologies. This would result in the creation of customised outputs and also would make the generalisation independent of the data schema since it will be able to reason based on ontologies in the creation of results (Regnaud, 2007). But most importantly ontological description would allow sharing of geographic information across different information systems (Harvey *et al.*, 1999; Mark *et al.*, 2004)

- **Developing dataset partitioning Techniques**

A seamless detailed database, such as OS MasterMap Topography layer, requires appropriate data partitioning prior to application of generalisation processes for a large region of interest because of the large quantities of data involved. The partitioning needs to be such that the boundaries of the selected regions do not affect the generalisation process. Since different generalisation techniques are required for different classes so different ways of partitioning would be needed. Future research will look into development of partitioning techniques based on network features (roads, hydrology, railways, junctions), terrain features (morphological units) as well as existing boundaries for urban areas, forests and coast lines. Thus partitioning will be based on geographic features not arbitrary grids.

“Generalisation is more than just mimicry of human cartographer, it is about modelling geographic space” (Mackness *et al.*, 2007 p.316). This process has its origin in manual cartography. However the increased use of GIS, integrating vast amounts of spatial data and advances in IT, it is now an integral part of techniques and technologies that extend from data capture to visualisation and interaction. The need is to develop generalisation techniques that can model the meaning, properties and relationships of geographic phenomena at different levels of detail. Such geographic modelling techniques would identify and emphasise salient patterns and



associations inherent among spatial datasets. These techniques are not only needed for the transformation of spatial databases from higher to lower levels of detail but are also highly significant in the creation of effective visual outputs, spatial analysis, data mining and *meaningful* interrogation of spatial data. It is hoped that this thesis has contributed to a deeper understanding of the relevance and application of model generalisation techniques to the interpretation of geographic information.

## References

- Agarwal, P. (2005) Ontological considerations in GIScience. *International Journal of Geographical Information Science*, 19(5), 501-536.
- Angel, S., Sheppard, S., & Civco, D. (2005). The Dynamics of Global Urban Expansion. Transport and Urban Development Department, The World Bank, Washington D.C. Url: <http://www.citiesalliance.org/doc/resources/upgrading/urban-expansion/worldbankreportsept2005.pdf> (24/10/07)
- Badard, T. & Richard, D. (2001) Using XML for the exchange of updating information between geographical information systems. *Computers, Environment and Urban Systems*, 25(1), 17-31.
- Barker, R.G. & Wright, H.F. (1954) *Midwest and its children: The psychological ecology of an American town* Row, Peterson and Company, Evanston, Illinois.
- Barrault, M. (1995) An automated System for Linear Feature Name Placement which Complies with Cartographic Quality Criteria. In *AutoCarto 12*, Vol. 4, pp. 321-330. ACSM/ASPRS, Charlotte, North Carolina.
- Bernauer, J. (1996) Analysis of part whole relations and subsumption in the medical domain. *Data & Knowledge Engineering*, 20, 405-415.
- Bertin, J. (1967) *Sémiologie Graphique* Gauthier-Villars, Paris.
- Bertin, J. (1983) *Semiology of Graphics: Diagrams, Networks, Maps* University of Wisconsin Press, Madison, WI.
- Bian, L. (2007) Object-Oriented Representation of Environmental Phenomena: Is Everything Best Represented as an Object? *Annals of the Association of American Geographers*, 97(2), 267-281.
- Bittner, T. & Geoffrey, E. (2001) Toward an ontology for geomatics. *Geomatica*, 55(4), 475-490.
- Bonsall, F.F. (1974) The Separation of Munros. *Scottish Mountaineering Club Journal*, 30, 254-256.
- Brassel, K.E. & Weibel, R. (1988) A review and conceptual framework of automated map generalization. *International Journal of Geographic Information Systems*, 2, 229-244.
- Bredon, G.E. (1993) *Topology and geometry* Springer-Verlag, New York.
- Bregt, A.K. & Bulens, J. (1996). Application-Oriented Generalisation of Area Objects. In *Methods for the Generalisation of Geo-Databases* (ed M. Molenaar), pp. 57-64. NCG, Delft.
- Bulger, D. & Hunt, H. (1991) The Plonski forest management decision support system project. In *GIS '91 Symposium Proceedings*, Vol. 1, pp. 101-109. Forestry Canada, Vancouver.
- Burghardt, D., Neun, M., & Weibel, R. (2005) Generalisation Services on the Web- A Classification and an Initial Prototype Implementation. In *AutoCarto 2005 Research Symposium, Proceedings of American Congress on Survey and Mapping*, Las Vegas, USA.

- Burrough, P.A. (1996). Natural Objects with Indeterminate Boundaries. In *Geographic Objects with Indeterminate Boundaries* (eds P.A. Burrough & A. Frank), pp. 3-28. Taylor & Francis, London.
- Burrough, P.A. & Frank, A.U. (1996) *Geographic objects with indeterminate boundaries* Taylor & Francis, London.
- Campari, I. (1996). Uncertain Boundaries in Urban Space. In *Geographic Objects with Indeterminate Boundaries* (eds P.A. Burrough & A.U. Frank), pp. 57-69. Taylor & Francis, London.
- Campbell, R.N. (1992) Munro's Tables, 1891-1991. *Scottish Mountaineering Club Journal*, 35, 21-27.
- Carnap, R. (1956) *Meaning and Necessity, a Study in Semantics and Modal Logic*, 2 edn. The University of Chicago Press, Chicago.
- Casati, R., Smith, B., & Varzi, A. (1998). Ontological Tools for Geographic Representation. In *Formal Ontology in Information Systems* (ed N. Guarino), pp. 77-85. IOS Press, Amsterdam.
- Chaudhry, O. & Mackaness, W.A. (2005a) Rural and Urban Road Network Generalization Deriving 1:250,000 From 1:1250. In International Cartographic Conference pp. 10p. CD-ROM: Theme 9: Cartographic Generalisation and Multiple Representation, Session 4, La Coruña- Spain.
- Chaudhry, O. & Mackaness, W.A. (2005b) Visualisation of Settlements Over Large Changes in Scale. In 8th ICA Workshop on Generalisation and Multiple Representation, La Coruña- Spain. Url: [http://ica.ign.fr/Acoruna/Papers/Chaudhry\\_Mackaness.pdf](http://ica.ign.fr/Acoruna/Papers/Chaudhry_Mackaness.pdf) (24/10/07)
- Chaudhry, O. & Mackaness, W.A. (2007) Utilising Partonomic Information in the Creation of Hierarchical Geographies. In 10th ICA Workshop on Generalisation and Multiple Representation, Moscow, Russia. Url: <http://ica.ign.fr/BDpubli/moscow2007/Chaudhry-ICAWorkshop.pdf> (24/10/07)
- Clementini, E. & Felice, P.D. (1996). An algebraic model for spatial objects with undetermined boundaries. In *Geographic Objects with Indeterminate Boundaries* (eds P.A. Burrough & A. Frank), pp. 155-169. Taylor & Francis, London.
- Cline, O. (2004) Develop your applications using SQLJ. Url: <http://www.ibm.com/developerworks/db2/library/techarticle/dm-0412cline/> (24/10/07)
- Cohen, G. (1979) Some Unsolved Problems in Munroology. *Scottish Mountaineering Club Journal*, 31, 423-424.
- Cohn, A.G. & Gotts, N.M. (1996). The 'egg-yolk' representation of regions with indeterminate boundaries. In *Geographic Objects with Indeterminate Boundaries* (eds P.A. Burrough & A. Frank), pp. 171-187. Taylor & Francis, London.
- Crouch, C. (2006) Territorial Reviews: Competitive Cities in the Global Economy. OECD Paris. Url: [http://www.oecd.org/document/2/0,3343,en\\_2649\\_34413\\_37801602\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/2/0,3343,en_2649_34413_37801602_1_1_1_1,00.html) (24/10/07)
- Cruse, D.A. (1986) *Lexical Semantics* Cambridge University Press, Cambridge.
- Dawson, A. (1995) *The Murdos* TACit Press.

- Devoegele, T., Trevisan, J., & Ranal, L. (1996) Building a Multi Scale Database with Scale Transition Relationships. In *Advances in GIS Research II, Proceedings of the 7th International Symposium on Spatial Data Handling* (eds M-J Kraak, M. Molenaar & E.M. Fendel), pp. 337-352. London, Taylor and Francis, Delft.
- Dijkstra, E.W. (1959) A Note on Two Problems in Connection With Graphs. *Numeriskche Mathematik*, 1, 269-271.
- Downs, T.C. & Mackaness, W.A. (2002) Automating the Generalisation of Geological Maps: The Need for an Integrated Approach. *The cartographic Journal*, 39(2), 137-152.
- Duckham, M., Mason, K., Stell, J., & Worboys, M. (2001) A formal approach to imperfection in geographic information. *Computers, Environment, and Urban Systems*, 25(1), 89-103.
- Duckham, M. & Sharp, J. (2005). Uncertainty and Geographic Information: Computational and critical convergence. In *Representing GIS* (eds P.F. Fisher & D. Unwin), pp. 113-124. John Wiley, New York.
- Edwards, A., Burghardt, D., & Neun, M. (2007). Experiments in Building an Open Generalisation System. In *Generalisation of Geographic Information: Cartographic Modelling and Applications* (eds W.A. Mackaness, A. Ruas & L.T. Sarjakoski), pp. 161-175. Elsevier, Oxford.
- Egenhofer, M., Clementini, E., & Felice, P.D. (1994a) Topological Relations between Regions with Holes. *International Journal of Geographical Information Systems*, 8(2), 129-144.
- Egenhofer, M.J., Clementini, E., & Felice, P.D. (1994b) Evaluating Inconsistencies Among Multiple Representations. In *Spatial Data Handling* (eds T. Waugh & R. Healey), Vol. 2, pp. 901-919. AGI, Edinburgh.
- Egenhofer, M.J. & Franzosa, R.D. (1991) Point set Topological Spatial Relations. *International Journal of Geographical Information Systems*, 5(2), 161-174.
- Egenhofer, M.J. & Herring, J.R. (1990). A Mathematical Framework for the Definition of Topological Relationships. In *Proceedings of the 4th International Symposium on Spatial Data Handling* (eds K. Brassel & K. Kishimoto), pp. 803 - 813, Zürich.
- Esnard, A.-M. & Yang, Y. (2002) Descriptive and Comparative Studies of 1990 Urban Extent: Data for the New York. *Metropolitan Region URISA Journal*, 14(1), 57-62.
- Evans, I.S. (1972). General geomorphometry, derivatives of altitude, and descriptive statistics. In *Spatial Analysis in Geomorphology* (ed R.J. Chorley), pp. 17-90. Methuen & Co Ltd.
- Fisher, P. & Wood, J. (1998) What is a Mountain? Or The Englishman who went up a Boolean geographical concept but realised it was fuzzy. *Geography* 83, 247-256.
- Fisher, P., Wood, J., & Cheng, T. (2004) Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. *Transactions of the Institute of British Geographers*, 29(1), 106-128.
- Fisher, P.F. (2000). Fuzzy modelling. In *Geocomputing* (eds S. Openshaw & R. Abraham), pp. 161-186. Taylor & Francis London.

- Foerster, T. & Stoter, J. (2006) Establishing an OGC Web Processing Service for generalisation processes. In 9th ICA Workshop on Generalisation and Multiple Representation, Vancouver- WA, USA. Url: [http://aci.ign.fr/Portland/paper/ICA2006-foerster\\_stoter.pdf](http://aci.ign.fr/Portland/paper/ICA2006-foerster_stoter.pdf) (24/10/07)
- Fonseca, F.T. (2001) Ontology-Driven Geographic Information Systems. Doctoral Dissertation, The University of Maine.
- Frank, A. & Timpf, S. (1994) Multiple representations for cartographic objects in a multi-scale tree: an intelligent graphical zoom. *Computers and Graphics Special Issue: Modelling and Visualization of Spatial Data in Geographic Information Systems*, 18(6), 823-829.
- Frank, A.U. (1997). Spatial ontology: a geographical information point of view. In *Spatial and Temporal Reasoning* (ed O. Stock), pp. 135-153. Kluwer, Dordrecht, Netherlands.
- Frank, A.U. & Egenhofer, M. (1988). Object Oriented Database Technology for GIS. Seminar Workbook, San Antonio Texas, USA. Url: [ftp://ftp.geoinfo.tuwien.ac.at/frank/afmerep95obj1\\_2.88.pdf](ftp://ftp.geoinfo.tuwien.ac.at/frank/afmerep95obj1_2.88.pdf) (24/10/07)
- Fujita, M., P, K., & Mori, T. (1999) On the evolution of hierarchical urban systems. *European Economic Review*, 43(2), 209-251.
- Gersting, J.L. (1992) *Mathematical Structures of Computer Science*, 3 edn. Computer Science Press, New York.
- Gerstl, P. & Pribbenow, S. (1995) Midwinters, end games and body parts: A classification of part-whole relations. *International Journal of Human-Computer Studies*, 43(5-6), 865-889.
- Gold, C.M. (1998) The Quad-Arc data structure. In 8th International Symposium on Spatial Data Handling, pp. 713-724, Vancouver, Canada.
- Gold, C.M., Nantel, J., & Yang, W. (1996) Outside-in: An alternative approach to forest map digitizing. *International Journal of Geographical Information Systems*, 10(3), 291-310.
- Goodchild, M., Monette, D., Fohl, P., & Gottsegen, J. (1998) Fuzzy Spatial Queries in Digital Spatial Data Libraries. In IEEE World Congress on Computational Intelligence, Vol. 1, pp. 205-210, Anchorage, AK, USA.
- Goodchild, M.F. (1992) Geographical Data Modelling. *Computers and Geoscience*, 18(4), 401-408.
- Grenon, P. & Smith, B. (2004) SNAP and SPAN: towards dynamic spatial ontology. *Spatial Cognition and Computation*, 4, 69-104.
- Gruber, T.R. (1993) A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2), 199-200.
- Gruber, T.R. (1995) Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human and Computer Studies*, 43, 907-928.
- Grünreich, D. (1985). Computer assisted generalisation. In *Papers CERCO-Cartography Course* Frankfurt a. M., Institut für Angewandte Geodäsie.
- Guarino, N. & Giaretta, P. (1995). Ontologies and Knowledge Bases: Towards a Terminological Clarification. In *Towards Very large Knowledge Bases* (ed N.J.I. Mars). IOS Press.
- Hampe, M. & Sester, M. (2002). Real-time generalisation of spatial data for mobile applications. In *Geowissenschaftliche Mitteilungen, Technische Universität Wien*, Vol. 60, pp. 147-155. Heft, Wein.

- Harrie, L. (2001) An Optimisation Approach to Cartographic Generalisation. Doctoral Thesis, Lund University, Lund, Sweden.
- Harrie, L. & Hellström, A.K. (1999) A Prototype System for Propagating Updates between Cartographic Data Sets. *The Cartographic Journal*, 36(2), 133-140.
- Harrie, L.E. & Weibel, R. (2007). Modelling the Overall Process of Generalisation. In *Generalisation of Geographic Information: Cartographic Modelling and Applications* (eds W.A. Mackaness, A. Ruas & L.T. Sarjakoski), pp. 67-88. Elsevier, Oxford.
- Harvey, F., Kuhn, W., Pundit, H., Bishr, Y., & Reiedemen, C. (1999) Semantic interoperability: a central issue for sharing geographic information. *Annals of Regional Science, Special issue on Geo-spatial Data sharing and Standardization*, 33, 213-232.
- Heikkila, E.J., Shen, T., & Yang, K. (2003) Fuzzy urban sets: theory and application to desakota regions in China *Environment and Planning B: Planning and Design*, 30(2), 239-254.
- Hughes, J.G. (1991) *Object Oriented Databases* Clarendon Press, Oxford.
- Iris, M., Litowltz, B., & Evens, M. (1988). Problems of The Part-Whole Relation. In *Relational Models Of The Lexicon* (ed W. Evers), pp. 262-288 Cambridge University Press.
- Jiang, B. & Claramunt, C. (2004) A structural approach to the model generalisation of an urban street network. *GeoInformatica*, 8(2), 157-171.
- João, E.M. (1998) *Causes and Consequences of Map Generalization* Taylor & Francis, London.
- Jones, C.B. (1991) Database Architecture for Multi-scale GIS. In *Auto-Carto 10: Technical Papers of the 1991 ACSM-ASPRS Annual Convention*, Vol. 6, pp. 1-14, Baltimore, Maryland.
- Kidner, D.B. & Jones, C.B. (1994) A Deductive Object-Oriented GIS for handling Multiple Representation. In *Advances in GIS Research, Proceedings Sixth International Symposium on Spatial Data Handling* (eds T.C. Waugh & R.G. Healey), Vol. 2, pp. 882-900, Edinburgh, Scotland.
- Kilpeläinen, T. (1992) Multiple representation and knowledge based generalization of topographical data. *Proceedings ISPRS Washington*, 954-964.
- Kilpeläinen, T. (1997) Multiple representation and generalisation of geo-databases for topographic maps. Doctoral dissertation (no 124), Publications of the Finnish Geodetic Institute, Kirkkonummi
- Kilpeläinen, T. (2001) Maintenance of Multiple Representation Databases for Topographic Data. *The Cartographic Journal*, 37(2), 101-107.
- Kilpeläinen, T. & Sarjakoski, T. (1995). Incremental generalisation for multiple representations of geographical objects. In *GIS and Generalisation: Methodology and Practice* (eds J.C. Müller, J.P. Lagrange & R. Weibel), pp. 209-218. Taylor & Francis, London.
- Kothuri, R.K., Ravada, S., & Abugov, D. (2002) Quadtree and R-tree Indexes in Oracle Spatial: A Comparison using GIS Data In *ACM SIGMOD Madison*, Wisconsin, USA.
- Kuhn, W. (2001) Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science*, 15(7), 613-633.
- Kulik, L., Duckham, M., & Egenhofer, M.J. (2005) Ontology-Driven Map Generalization. *Journal of Visual Languages and Computing*, 16(3), 245-267.

- Leitner, H. (2004). The Politics of Scale and Networks of Spatial Connectivity: Transnational Interurban networks and the Rescaling of Political Governance in Europe. In *Scale and Geographic Inquiry: Nature Society and Method* (eds E. Sheppard & B. McMaster), pp. 236-255. Blackwell Publishing, Malden, MA.
- Li, Z. (2007) Digital Map Generalization at the Age of Enlightenment: a Review of the First Forty Years. *The Cartographic Journal*, 44(1), 80-93.
- Linturi, R. & Simula, T. (2005). Virtual Helsinki: Enabling the Citizen; Linking the Physical and Virtual. In *Digital Cities III, Information Technologies for Social Capital: Cross-Cultural Perspectives, Third International Digital Cities Workshop* (eds P. Besselaar & S. Koizumi), pp. 113-140. Berlin, Springer, Amsterdam, the Netherlands.
- Liu, C.L. (1983) *Elements of Discrete Mathematics* McGraw Hill, New York.
- Liu, Y. (2002) Categorical Database Generalization in GIS. Doctoral Dissertation (no 88), ITC, the Netherlands
- Liu, Y., Molenaar, M., Ai, T., & Liu, Y. (2003) Categorical Database Generalisation Aided by Data Model. In Proceedings of the 21st International Cartographic Conference, pp. 2308-2318, Durban, South Africa. Url: [http://www.itc.nl/library/Papers\\_2003/art\\_proc/liu.pdf](http://www.itc.nl/library/Papers_2003/art_proc/liu.pdf) (24/10/07)
- Lüscher, P., Burghardt, D., & Weibel, R. (2007) Ontology-driven Enrichment of Spatial Databases. In 10th ICA Workshop on Generalisation and Multiple Representation, Moscow, Russia. <http://ica.ign.fr/BDpubli/moscow2007/Luescher-ICAWorkshop.pdf>
- Mackaness, W., A & Mackechnie, G., A (1999) Automating the Detection and Simplification of Junctions in Road Networks. *GeoInformatica*, 3(2), 185-200.
- Mackaness, W.A. (2006) Automated Cartography in a Bush of Ghosts. *Cartography and Geographic Information Science*, 33(4), 245-256.
- Mackaness, W.A. (2007). Understanding Geographic Space. In *Generalisation of Geographic Information: Cartographic Modelling and Applications* (eds W. Mackaness, A. Ruas & L.T. Sarjakoski), pp. 1-10. Elsevier, Oxford.
- Mackaness, W.A. & Beard, M.K. (1993) Use of Graph Theory to Support Map Generalization. *Cartography and Geographic Information Systems*, 20, 210 - 221.
- Mackaness, W.A. & Chaudhry, O. (2007a). Map Generalization. In *International Encyclopedia of Human Geography (in press)* (eds N. Thrift & R. Kitchin). Elsevier.
- Mackaness, W.A. & Chaudhry, O. (2007b). Symbolization and Generalization In *Encyclopedia of Geographical Information Science (in press)* (eds S. Shekhar & H. Xiong). Springer, New York.
- Mackaness, W.A. & Edwards, G. (2002) The importance of Modelling Pattern and Structure in Automated Map Generalization. In Joint ISPRS/ICA workshop on multi-scale representations of spatial data, Ottawa, Canada. Url: <http://www.ikg.uni-hannover.de/isprs/workshop/macedwards.pdf> (24/10/07)
- Mackaness, W.A. & Ruas, A. (2007). Evaluation in the Map Generalisation Process. In *Generalisation of Geographic Information: Cartographic Modelling and Applications* (eds W.A. Mackaness, A. Ruas & L.T. Sarjakoski), pp. 89-111. Elsevier, Oxford.

- Mackaness, W.A., Ruas, A., & Sarjakoski, L.T. (2007). Observations and Research Challenges in Map Generalisation and Multiple Representation. In *Generalisation of Geographic Information: Cartographic Modelling and Applications* (eds W.A. Mackaness, A. Ruas & L.T. Sarjakoski), pp. 315-323. Elsevier, Oxford.
- Mark, D.M., Egenhoer, M., Hirtle, S., & Smith, B. (2004). Ontological Foundations for Geographic Information Science. In *Research Challenges in Geographic Information Science* (eds R.B. McMaster & L. Usery), pp. 335-350. John Wiley and Sons, New York.
- Mark, D.M. & Smith, B. (2004). A science of topography: bridging the qualitative-quantitative divide. In *Geographic Information Science and Mountain Geomorphology* (eds M.P. Bishop & J. Shroder), pp. 75-100. Springer-Praxis, Chichester.
- Mark, D.M., Smith, B., & Tversky, B. (1999). Ontology and Geographic Objects: An Empirical Study of Cognitive Categorization. In *Spatial Information Theory: A Theoretical Basis for GIS, Berlin: Springer-Verlag, Lecture Notes in Computer Science* (eds C. Freksa & D.M. Mark), Vol. 1661, pp. 283-298.
- Martinez, C.J.A. & Molenaar, M. (1995) Aggregation Hierarchies for Multiple Scale Representations of Hydrographic Networks in GIS. In 17th International Cartographic Conference, Vol. 1, pp. 358-362, Barcelona, Spain.
- Martinez, J.A. (1994) Hydrographic Information Abstraction for Erosion Modeling at Regional Level: A Database Perspective in a GIS Environment. MSc, Wageningen Agricultural University, Wageningen, the Netherlands.
- McMaster, R.B. & Shea, K.S. (1992) *Generalization in Digital Cartography* Association of American Geographers, Washington D.C.
- Mesev, T.V., Longley, P.A., Batty, M., & Xie, Y. (1995) Morphology from Imagery: Detecting and Measuring the Density of Urban Land Use. *Environment and Planning A*, 27(5), 759-780.
- Milton, S.K. & Kazmierczak, E. (2003) An ontology of data modelling languages: a study using a common-sense realistic ontology. *Journal of Database Management*, 2, 19-38.
- Minsky, M. (1975). A framework for representing knowledge. In *The psychology of computer vision* (ed P. Winston), pp. 211 - 277. McGraw Hill Inc, New York.
- Miyamoto, S. (1990) *Fuzzy sets in Information Retrieval and Cluster Analysis* Kluwer Academic Press, Dordrecht, Netherlands.
- Moise, E.E. (1977) *Geometric Topology in Dimension 2 and 3* Springer-Verlag, New York.
- Molenaar, M. (1993) Object hierarchies and uncertainty in GIS or why is standardisation so difficult. *GeoInformations-System*, 6(3), 22-28.
- Molenaar, M. (1996a) The extensional uncertainty of spatial objects. In Proceedings of the 7th International Conference on Spatial Data Handling (eds M.-J. Kraak & M. Molenaar), pp. 571-584, Delft.
- Molenaar, M. (1996b). A Syntactic Approach for Handling the Semantics of Fuzzy Spatial Objects. In *Geographic Objects with Indeterminate Boundaries* (eds P.A. Burrough & A. Frank), pp. 207-224. Taylor & Francis, London.
- Molenaar, M. (1998) *An Introduction to the Theory of Spatial Object Modelling for GIS* Taylor & Francis, London.



- Molenaar, M. (2004) Spatial Data Modelling, Dual Partitions and The Specification of Semantics. In 20th ISPRS Congress, Istanbul, Turkey. Url: <http://www.isprs.org/istanbul2004/comm4/papers/311.pdf> (24/10/07)
- Monmoier, M.S. (1991) Ehtics and map design: Six strategies for confronting the traditional one-map solution. *Cartographic Perspectives*, 10, 3-8.
- Monmonier, M.S. (1984) Geographic information and cartography. *Progress in Human Geography*, 8, 381-391.
- Müller, J.C. (1989) Theoretical considerations for automated map generalisation. *ITC Journal*, 3(4), 200-204.
- Müller, J.C. (1991). Generalization of spatial databases. In *Geographical Information Systems: Principles and Applications* (eds D. Maguire, M. Goodchild & D. Rhind), Vol. 1, pp. 457-475. Longman, Harlow, Essex.
- Müller, J.C., Lagrange, J.P., & Weibel, R. (1995). GIS and Generalization: Methodology and Practice. In *GISDATA 1* (eds I. Masser & F. Salge). Taylor & Francis, London.
- Müller, J.C. & Wang, Z. (1992) Area-Patch generalisation: a competitive approach. *The Cartographic Journal*, 29(1), 137-144.
- Mustière, S. & Moulin, B. (2002) What is Spatial Context in Cartographic Generalisation? In Symposium on Geospatial Theory, Processing and Application, Ottawa, Canada. Url: <http://www.isprs.org/commission4/proceedings02/pdfpapers/377.pdf> (24/10/07)
- Mustière, S. & van Smaalen, J. (2007). Database Requirements for Generalisation and Multiple Representations. In *Generalisation of Geographic Information: Cartographic Modelling and Applications* (eds W.A. Mackaness, A. Ruas & L.T. Sarjakoski), pp. 113-136. Elsevier, Oxford.
- Neun, M. & Burghardt, D. (2005) Web Services for an Open Generalisation Research Platform. In 8th ICA WORKSHOP on Generalisation and Multiple Representation, A Coruña, Spain.
- Neun, M., Weibel, R., & Burghardt, D. (2004) Data enrichment for adaptive generalisation. In 7th ICA Workshop on generalisation and multiple representation, Leicester, United Kingdom.
- Nyerges, T.L. (1991). Representing geographical meaning. In *Map generalization: Making rules for knowledge representation* (eds B.P. Buttenfield & R.B. McMaster), pp. 59-85. Longman Scientific & Technical, Essex.
- Oracle (2005) Spatial User's Guide and Reference 10g Release 1 (10.1). Url: [http://www.oracle-10g-buch.de/oracle\\_10g\\_documentation/appdev.101/b10826/toc.htm](http://www.oracle-10g-buch.de/oracle_10g_documentation/appdev.101/b10826/toc.htm) (24/10/07)
- Ordnance Survey (2005) 1:250,000 specification. *Ordnance Survey internal document*.
- Ordnance Survey (2007a). OS Land-Form PROFILE Plus user guide. Url <http://www.ordnancesurvey.co.uk/oswebsite/products/landformprofileplus/pdf/LandFormPROFILE%20Plus.pdf> (24/06/07)
- Ordnance Survey (2007b). OS MasterMap real-world object catalogue. Url: <http://www.ordnancesurvey.co.uk/products/osmastermap/information/technical/docs/realworldobjectcatalogue.pdf> (24/10/07)

- Ordnance Survey (2007c). OS MasterMap Topography Layer: User guide and Technical specification. Url: <http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/userguide/docs/OSMMTopoLayerUserGuide.pdf> (24/10/07)
- Ordnance Survey (2007d). OS MasterMap user guide. Url: <http://www.ordnancesurvey.co.uk/products/osmastermap/userguides/docs/userguidepart1.pdf> (24/10/07)
- Ordnance Survey (2007e). OS Strategi User guide and technical specification. Url: <http://www.ordnancesurvey.co.uk/oswebsite/products/strategi/pdf/StrategiUserGuide.pdf> (24/10/07)
- Ormsby, D. & Mackaness, W.A. (1999) The Development of Phenomenological Generalisation Within an Object Oriented Paradigm. *Cartography and Geographical Information Systems*, 26(1), 70-80.
- Paiva, J.A. (1998) Topological Equivalence And Similarity In Multi-Representation Geographic Databases. Doctoral Dissertation, University of Maine.
- Peng, W. (1997) Automated Generalization in GIS. Doctoral Dissertation, Wageningen Agricultural University and ITC, the Netherlands.
- Peng, W. & Molenaar, M. (1995) An object oriented approach to automated generalisation. In *GeoInformatics 95*, pp. 295-304, Hong Kong.
- Petzold, I., Gröger, G., & Plüemer, L. (2003) Fast Screen Map Labeling – Data-Structures And Algorithms. In 21st International Cartographic Conference (ICC), Cartographic Renaissance, pp. 288-299 CD-ROM, Durban, South Africa. Url: [http://www.geo.unizh.ch/~petzold/publications/petzold\\_icc03.pdf](http://www.geo.unizh.ch/~petzold/publications/petzold_icc03.pdf) (24/10/07)
- Peucker, T.K. & Douglas, D.H. (1974) Detection of surface specific points by local parallel processing of discrete terrain elevation data. *Computer Graphics and Image Processing*, 4, 375-387.
- Peuquet, D.J. (1984) Conceptual framework and comparison of spatial data models. *Cartographica*, 21(4), 66 - 113.
- Press, F. & Siever, R. (1982) *Earth*, 3 edn. W.H. Freeman & Co, San Francisco, CA.
- Purchase, D. (1997) The Classification of Mountains. *Scottish Mountaineering Club Journal* 36, 252-271.
- Regnault, N. (1996) Recognition of Building Cluster for Generalization. In *Proceedings of the 7th International Symposium on Spatial Data Handling*, Vol. 1, pp. 185-198.
- Regnault, N. (2007) Evolving from automating existing map production systems to producing maps on demand automatically. In 10th ICA Workshop on Generalisation and Multiple Representation, Moscow, Russia. <http://ica.ign.fr/BDpubli/moscow2007/Regnault-ICAWorkshop.pdf>
- Regnault, N. & Mackaness, W.A. (2006) Creating a hydrographic network from its cartographic representation: a case study using Ordnance Survey MasterMap data. *International Journal of Geographical Information Science*, 20(6), 611-631.
- Regnault, N. & McMaster, R.B. (2007). A Synoptic View of Generalisation Operators. In *Generalisation of Geographic Information: Cartographic Modelling and Applications* (eds W.A. Mackaness, A. Ruas & L.T. Sarjakoski), pp. 37-66. Elsevier, Oxford.

- Reinke, K.J. & Hunter, G.J. (2002). A Theory for communicating Uncertainty in Spatial Databases. In *Spatial Data Quality* (eds W. Shi, P.F. Fisher & M.F. Goodchild), pp. 77-101. Taylor & Francis, London.
- Reitsma, F. & Bittner, T. (2003). Scale in object and process ontologies. In *Proceedings of Conference on Spatial Information Theory: Foundations of Geographic Information Science, COSIT'03, Lecture Notes in Computer Science* (eds W. Kuhn, M.F. Worboys & S. Timpf), Vol. 2825, pp. 13-30. Springer, Berlin.
- Revell, P. (2005) Seeing the Wood from the Trees: Generalising OS MasterMap® Tree Coverage Polygons to Woodland at 1:50 000 Scale. In 8th ICA WORKSHOP on Generalisation and Multiple Representation,, A Coruña, Spain.
- Richardson, D.E. (1993) Automatic Spatial and Thematic Generalisation using a Context Transformation Model. Doctoral Dissertation, Wageningen Agricultural University, the Netherlands.
- Robinson, G.J. (1995). A hierarchical top-down bottom-up approach to topographic map generalization. In *GIS and Generalization: Methodology and Practice* (eds J.C. Müller, L.J. P & R. Weibel), pp. 235-245. Taylor & Francis, London.
- Robinson, V.B. (1988) Some implications of fuzzy set theory applied to geographic databases. *Computers, Environment and Urban Systems*, 12, 89-97.
- Robinson, V.B. (2003) A perspective on the fundamentals of fuzzy sets and their use in geographic information systems. *Transactions in GIS*, 7, 3-30.
- Robinson, V.B. (2007). Fuzzy Classification and Mapping. In *The handbook of Geographic Information Science* (eds J.P. Wilson & A.S. Fotheringham), pp. 259-272. Blackwell Publishing, Oxford.
- Robinson, V.B., Miller, R., & Klesh, L. (1988) Issues in the Use of Expert Systems to Manage Uncertainty in Geographic Information Systems. In *Proceedings of the First International Symposium on GIS*, Vol. 2, pp. 89-100.
- Rotman, J.J. (1988) *An Introduction to Algebraic Topology* Springer-Verlag, New York
- Ruas, A. (1998) A Method for Building Displacement in Automated Map Generalisation. *International Journal of Geographic Information Science*, 12(8), 789-803.
- Ruas, A. & Lagrange, J.P. (1995). Data and knowledge modelling for generalization. In *GIS and Generalization: Methodology and Practice* (eds J.C. Müller, J.P. Lagrange & R. Weibel), pp. 73-90. Taylor & Francis, London.
- Ruas, A. & Mackaness, W.A. (1997) Strategies for Urban Map Generalization. In *Proceedings of the 18th International Cartographic Conference* (ed L. Ottorson), Vol. 3, pp. 1387-1394, Stockholm, Sweden, 23-27 June.
- Ruas, A. & Plazanet, C. (1997). Strategies for automated generalisation. In *Advances in GIS Research II (Proceedings of 7th International Symposium on Spatial Data Handling)* (eds M.J. Kraak & M. Molenaar), pp. 319-336. Taylor & Francis, London.
- Sarjakoski, L.T. (2007). Conceptual Models of Generalisation and Multiple Representation. In *Generalisation of Geographic Information: Cartographic Modelling and Applications* (eds W.A. Mackaness, A. Ruas & L.T. Sarjakoski), pp. 11-36. Elsevier, Oxford.

- Schylberg, L. (1993) Computational Methods for Generalization of Cartographic Data in a Raster Environment. Doctoral Thesis, Photogrammetric Reports No 60 - Royal Institute of Geodesy and Photogrammetry, Stockholm, Sweden.
- Sester, M., Anders, K.-H., & Walter, V. (1998) Linking objects of different spatial data sets by integration and aggregation. *GeoInformatica*, 2(4), 335-358.
- Sheeren, D., Mustière, S., & Zucker, J.D. (2004). Consistency Assessment between Multiple Representations of Geographic Databases: A Specification Based Approach. In *Developments in Spatial Data Handling, 11th International Symposium on Spatial Data Handling* (ed P. Fisher), pp. 617-628. Springer, Berlin.
- Sheppard, E. & McMaster, R.B. (2004). Scale and Geographic Inquiry. In *Scale and Geographic Inquiry: Nature Society and Method* (eds R.B. McMaster & E. Sheppard), pp. 1-22. Blackwell Publishing, Malden, MA.
- Simons, P. (1987) *Parts: A Study in Ontology* Oxford University Press, Oxford.
- Small, C., Pozzi, F., & Elvidge, C.D. (2005) Spatial analysis of global urban extent from DMSP-OLS night lights. *Remote Sensing of Environment*, 96, 277-291.
- Smith, B. (1995). On Drawing Lines on a Map. In *Spatial Information Theory. Proceedings of COSIT '95 Berlin/Heidelberg/Vienna/New York/London/Tokyo* (eds A.U. Frank, W. Kuhn & D. Mark), pp. 475-484. Springer, Verlag.
- Smith, B. & Mark, D. (2001) Geographical categories: an ontological investigation. *International Journal Geographical Information Science*, 15(7), 591-612.
- Smith, B. & Mark, D.M. (1998). Ontology and geographic kinds. In *Proceedings of the 8th International Symposium on Spatial Data Handling (SDH'98)* (eds T.K. Poiker & N. Chrisman), pp. 308-320, Vancouver: International Geographical Union.
- Smith, B. & Mark, D.M. (2003) Do Mountains exist? Towards an Ontology of Landforms. *Environment and Planning B :Planning and Design*, 30(3), 411-427.
- Smith, B. & Rosse, C. (2004) The Role of Foundational Relations in the Alignment of Biomedical Ontologies. In *Medinfo 2004*. IOS Press, Amsterdam. Url: <http://ontology.buffalo.edu/medo/isa.pdf> (24/10/07)
- Smith, B. & Varzi, A. (1997) The Formal Ontology of Boundaries. *Electronic Journal of Analytic Philosophy*, 5(5).
- Smith, B. & Varzi, A.C. (2000) Fiat And Bona Fide Boundaries. *Philosophy And Phenomenological Research*, 60, 401-420.
- Smith, J.M. & Smith, D.C.P. (1977) Database abstractions: aggregation and generalisation. *ACM Transactions on Database Systems*, 2, 105-133.
- Steiniger, S., Burghardt, D., & Weibel, R. (2006) Recognition of Island Structures for Map Generalisation. In *ACM-GIS Arlington, Virginia, USA*.
- Steiniger, S. & Weibel, R. (2007) Relations among Map Objects in Cartographic Generalisation. *Cartography and Geographic Information Science*, 34(3), 175-197.
- Summerfield, A.M. (1991) *Global Geomorphology* Longman, London.
- Tang, L. (1992) Automatic Extraction of Specific Geomorphological Elements from Contours. In *Proceedings of the 5th International Symposium on Spatial Data Handling* (eds E. Corwin, D. Cowen & P. Bresnahan), Vol. 2, pp. 554-566. IGU Commission of GIS, Charleston.

- Taylor, P.J. (2004). Is there a Europe of cities? World cities and the limitations of Geographical Scale Analyses. In *Scale and Geographic Inquiry: Nature, Society, and Method* (eds E. Sheppard & B. McMaster), pp. 213-235. Blackwell Publishing, Malden, MA.
- Thompson, P.J. (1989) *Data with semantics: Data models and data management* Van Nostrand Reinhold, New York.
- Thomson, R.C. & Richardson, D.E. (1995) A Graph Theory Approach to Road network Generalization. In Proceedings of 17th International Cartographic Conference, pp. 1871-1880, Barcelona, Spain.
- Töpfer, F. & Pillewizer, W. (1966) The principles of selection: a means of cartographic generalization. *The Cartographic Journal*, 3(1), 10 - 16.
- Tryfona, N. & Egenhofer, M. (1997) Consistency among Parts and Aggregates: A Computational Model. *Transactions in GIS*, 1(3), 189-206.
- Turnbull, D. (1989) *Maps are territories Science is an atlas* Deakin University, Geelong.
- Tversky, B. (1990). Where partonomies and taxonomies meet. In *Meanings and prototypes: Studies on Linguistic Categorization* (ed S.L. Tsohatzidis), pp. 334-344. Routledge, London.
- Tversky, B. & Hemenway, K. (1984) Objects, parts, and categories. *Experimental Psychology*, 113, 169-193.
- UGIS (1998) Cognition of geographic space, revised research agenda - white papers.
- Uitermark, H. (2001) Ontology-Based Geographic Data Set Integration. Doctoral Dissertation, Universiteit Twente, the Netherlands.
- Usery, E.L. (1996). A Conceptual Framework and Fuzzy Set Implementation for Geographic Features. In *Geographic Objects with Indeterminate Boundaries* (eds P.A. Burrough & A. Frank), Vol. 3, pp. 71-85. Taylor & Francis, London.
- van Oosterom, P. (1995). The GAP-tree, an approach to 'on-the-fly' map generalization of an area partitioning. In *GIS and Generalization: Methodology and Practice* (eds J.C. Müller, L.J. P & R. Weibel), pp. 120-132. Taylor & Francis, London.
- van Oosterom, P., Quak, C.W., & Tijssen, T.P.M. (2002) Testing current DBMS products with real spatial data. In Proceedings Urban Data Management Symposium, Prague, Czech Republic. Url: [http://www.gdmc.nl/publications/2002/Testing\\_current\\_DBMS.pdf](http://www.gdmc.nl/publications/2002/Testing_current_DBMS.pdf) (24/10/07)
- van Smaalen, J. (1996) A Hierarchic Rule Model for Geographic Information Abstraction. In Proceedings of the 7th International Symposium on Spatial Data Handling (eds M.-J. Kraak, M. Molenaar & E.M. Fendel), pp. 215-226, Delft, the Netherlands.
- van Smaalen, J. (2003) Automated Aggregation of Geographic Objects: A New Approach to the Conceptual Generalisation of Geographic Databases. Doctoral Dissertation, Wageningen University, the Netherlands.
- Varzi, A. (2007). Spatial Reasoning and Ontology: Parts, Wholes, and Locations. In *The Logic of Space* (eds M. Aiello, I.P. Hartmann & J. van Benthem). Springer-Verlag, Berlin.

- Varzi, A.C. (2003) Mereology. In Stanford Encyclopedia of Philosophy (ed E.N. Zalta). Stanford: CSLI (on line publication). Url: <http://plato.stanford.edu/entries/mereology/> (24/10/07)
- Weibel, R. (1986). Automated Cartographic generalisation. In *A Selected Bibliography on Spatial Data Handling: Data Structures, Generalisation and Three dimensional Mapping* (eds R. Sieber & K. Brassel), Vol. 6, pp. 20-35.
- Weibel, R. (1992) Models and Experiments for Adaptive Computer Assisted Terrain Generalization. *Cartography and Geographic Information Systems*, 19(3), 133-153.
- Weibel, R. (1995a) Map Generalization in the context of digital systems. *Cartography and Geographical Information Systems*, 22(4), 259-263.
- Weibel, R. (1995b). Three essential building blocks for automated generalization In *GIS and Generalization: Methodology and Practice* (eds J.C. Müller, J.P. Lagrange & R. Weibel), pp. 56-70. Taylor & Francis, London.
- Weibel, R. & Dutton, G. (1999). Generalising Spatial Data and Dealing with Multiple Representations. In *Geographical Information Systems* (eds P. Longley, M.F. Goodchild, D.J. Maguire & D.W. Rhind), pp. 125-156. John Wiley, New York.
- Willard, S. (1970) *General Topology* Addison-Wesley, Reading, Mass.
- Winston, M., Chaffin, R., & Herrmann, D. (1987) A Taxonomy of Part-Whole Relations. *Cognitive Science*, 11, 417-444.
- Winter, S. & Thomas, B. (2002). Hierarchical Topological Reasoning with Vague Regions. In *Spatial Data Quality* (eds S. Wenzhong, P. Fisher & M.F. Goodchild), pp. 35-49. Taylor & Francis, London.
- Wood, J. (1996a) The Geomorphological Characterisation of Digital Elevation Models. Doctoral Dissertation, University of Leicester, UK.
- Wood, J. (1996b). Scale-based characterisation of digital elevation models. In *Innovations in GIS 3* (ed D. Parker), pp. 163-176. Taylor & Francis, London.
- Wood, J. (1998) Modelling the Continuity of Surface Form Using Digital Elevation Models. In 8th International Symposium on Spatial Data Handling (eds T. Poiker & N. Chrisman), Vol. 1, pp. 725-736, Vancouver, Canada.
- Worboys, M.F. (1995) *A Computing Perspective* Taylor and Francis, London.
- Worboys, M.F., Hearnshaw, H.M., & Maguire, D.J. (1990). Object-oriented data and query modelling for Geographical Information Systems. In *Proceedings of the 4th International Symposium on Spatial Data Handling*, Vol. 2, pp. 679 - 688, Zürich, Switzerland.
- Yaolin, L., Molenaar, M., & Kraak, M.J. (2002) Semantic similarity evaluation model in categorical database generalization. In Proceeding ISPRS Commission IV Symposium on Geospatial Theory, Processing and Applications, Ottawa, Canada. Url: <http://www.isprs.org/commission4/proceedings02/pdppapers/413.pdf> (24/10/07)
- Zacks, J.M. & Tversky, B. (2001) Event Structure in Perception and Conception. *Psychological Bulletin*, 127(1), 3-21.

## Preface

The following appendices (Appendix I –V) contains copies of published papers prepared during this research.

**Appendix I** is a chapter accepted for publication in an up-coming book “*Encyclopedia of Geographical Information Science*”. **Appendix II** is a chapter accepted for publication in the up-coming book “*Encyclopedia of Human Geography*”. These two chapters introduce and explain the concepts of map generalisation, discussed in Chapter 2 and 3. These appendices also discuss important frameworks proposed in research used to model the process of map generalisation process. These two papers also describe model and cartographic generalisation and their operations and affects. They discuss different application areas of map generalisation in addition to cartography. As a second author, my contributions to these chapters have been in the sections discussing: model and cartographic generalisation approaches and their operations, the frameworks of generalisation and development of figures for both papers.

**Appendix III** is the paper accepted for publication in the Journal “*Computer Environment and Urban Systems*”. This paper presents the settlement container boundary detection approach presented in Chapter 5 (section 5.2). This paper illustrates application of the methodology on a few regions other than those presented in Chapter 5. These include a region from France using IGN France BDTopo data using the same parameter settings as were used for Ordnance Survey data. The paper also presents the evaluation of these boundaries in terms of visual comparison with manually generalised solutions and comments from cartographic experts.

**Appendix IV** is the paper accepted for publication in the Journal “*The Cartographic Journal*”. This paper presents the forest container boundary detection approach presented in Chapter 5 (section 5.3). This paper presents implementation of the proposed approach, using an open source platform. The paper also presents a

boundary simplification approach to the creation of appropriate cartographic representations. The algorithms are available via the web generalisation service at <http://www.geo.unizh.ch/~neun/webgen/>. The output boundaries are visually evaluated using manually generalised dataset and review by cartographic experts. As a third author, my contributions included development of methodology, assisting implementation, creation of output files and editing of revisions according to reviewers comments.

**Appendix V** is the paper accepted for publication in the Journal “*Transactions in GIS*”. This paper presents the summit container boundary detection approach presented in Chapter 5 (section 5.4). The paper illustrates the application of the methodology on regions other than those presented in Chapter 5 using the same parameter settings as were used for the regions shown in Chapter 5. The paper also presents the utility of the proposed approach in the creation of ‘parent-child’ relationships between summits. These relationships group summits on a landform into a hierarchy showing which summits are ‘sub-peaks’ of others. In this way ranges and hills can be classified.



## **Appendix I**

Mackness, W.A. and Chaudhry, O. (in press) Symbolization and Generalisation In *Encyclopedia of Geographical Information Science*, (eds S. Shekhar and H. Xiong), New York, Springer.

## Appendix II

Mackness, W.A. and Chaudhry, O. (in press) Map Generalisation, In *International Encyclopedia of Human Geography*, (eds N. Thrift and R. Kitchin), Elsevier.

**Appendix III**

Chaudhry, O. & Mackaness, W. (in press) Automatic Identification of Urban Settlement Boundaries for Multiple Representation Databases. *Computer Environment and Urban Systems*.

**Appendix IV**

Mackness, W., Perikleous, S. and Chaudhry, O. (in press) Representing Forested Regions at Small Scales: Automatic Derivation from the Very Large Scale. *The Cartographic Journal*

## Appendix V

Chaudhry, O. and Mackaness, W.A. (accepted) Creating Mountains out of Mole Hills: Automatic Identification of Hills and Ranges Using Morphometric Analysis, *Transactions in GIS*

.

## Curriculum Vitae

Omar Z Chaudhry  
Date of Birth: 12-Nov-1979  
Email: [omairzubair@gmail.com](mailto:omairzubair@gmail.com)

### Education

2004 -2007

PhD, Geographic Information Systems, University of Edinburgh UK, Dissertation  
"Modelling Geographic Phenomena at Multiple Levels of Detail: A Model  
Generalisation Approach based on Aggregation",  
Supervisors: Dr. William A Mackaness, Nicholas R.J. Hulton

2003 -2004

Master of Science, Geographic Information Science, University of Edinburgh, UK,  
Dissertation "Modelling large changes in the level of detail in GIS Databases",  
Supervisor: Dr. William A Mackaness

1998-2002

Bachelor of Engineering, Computer Software, National University of Sciences and  
Technology Pakistan, Dissertation: "Voice Mailing System: Automatic Spoken  
Email Delivery on digital telephone", CGPA=3.58

1996-1998

Higher Secondary Education, Pre Engineering, F.G Sir Syed College Pakistan,  
Overall Grade: A

### Awards and Grants

Ordnance Survey/ Geographical Information Science Research Group Prize for "Best  
Postgraduate GIS Paper" at RGS-IBG Annual International Conference September  
2006

GISRUK 2007 Half Student Bursary

Ordnance Survey, UK, Research grant, 2004-2007

School of GeoScience, The University of Edinburgh, Half Teaching Studentship,  
2004-2007

## Publications

**Chaudhry, O.** and Mackaness, W.A. (in press) Automatic Identification of Urban Settlement Boundaries for Multiple Representation Databases, *Computer Environment and Urban Systems*.

**Chaudhry, O.** and Mackaness, W.A. (accepted) Creating Mountains out of Mole Hills: Automatic Identification of Hills and Ranges Using Morphometric Analysis, *Transactions in GIS*.

Mackaness, W.A. and **Chaudhry, O.** (in press) Symbolization and Generalisation In *Encyclopedia of Geographical Information Science*, (eds S. Shekhar and H. Xiong), New York, Springer.

Mackaness, W.A. and **Chaudhry, O.** (in press) Map Generalisation, In *International Encyclopedia of Human Geography*, (eds N. Thrift and R. Kitchin), Elsevier.

Mackaness, W.A., Perikleous, S. and **Chaudhry, O.** (in press) Representing Forested Regions at Small Scales: Automatic Derivation from the Very Large Scale, *The Cartographic Journal*.

## Conference Papers (peer reviewed)

**Chaudhry, O.** and Mackaness, W. (2007) Utilising Partonomic Information in the Creation of Hierarchical Geographies. In 10th ICA Workshop on Generalisation and Multiple Representation, Moscow, Russia, 2-3 August.

**Chaudhry, O.** and Mackaness, W. (2007) Automatic identification of Hills and Ranges using Morphometric Analysis. In GISRUK, National University of Ireland Maynooth, 11-13 April.

Mackaness, W., Perikleous, S. and **Chaudhry, O.** (2007) Representing Forested Regions at Small Scales: Automatic Derivation from the Very Large Scale. In GISRUK, University of Ireland Maynooth, 11-13 April.

**Chaudhry, O.** and Mackaness, W. (2006) Density Modelling in Support of Automatic Recognition of Geographical Phenomena in Large Scale Topographic Databases. In GISRUK, University of Nottingham, 5-7 April.

**Chaudhry, O.** and Mackaness, W. (2006) Creation of Fiat Boundaries in Higher order Phenomenon. In 9th ICA Wrokshop on Generalisation and Multiple Representation, Vancouver – WA, 25 June.

Mackaness, W. and **Chaudhry, O.** (2005) Exploring representational issues in the visualisation of geographical phenomenon over large changes in scale. In GISRUK, Glasgow, 5 - 6 April.

**Chaudhry, O.** and Mackaness, W. (2005) Visualisation of Settlements Over Large Changes in Scale. In 8th ICA Workshop on Generalisation and Multiple Representation, La Coruña- Spain, 7-8 July.

#### **Conference Proceedings (abstract reviewed)**

**Chaudhry, O.** and Mackaness, W. (2007) Automated scale dependent views of hills and ranges via morphometric analysis. In International Cartographic Conference, Moscow, Russia, 6-10 August.

**Chaudhry, O.** and Mackaness, W. (2006) Modelling Geographic Phenomena at Multiple Levels of Detail. In AutoCarto, Vancouver- WA. 26-28 June.

**Chaudhry, O.** and Mackaness, W. (2005) Rural and Urban Road Network Generalisation Deriving 1:250,000 From 1:1250. In International Cartographic Conference pp. 10p. CD-ROM: Theme 9: Cartographic Generalisation and Multiple Representation, Session 4, La Coruña- Spain. 10-16 July

#### **Posters**

**Chaudhry, O.,** Mackaness, W. and Hulton, N. (2006) Gravitational Modelling for Automatic Identification of Boundaries of Geographic Phenomena. In AutoCarto, Vancouver – WA, 26-28 June.

**Chaudhry, O.** and Mackaness, W. (2006) Making Mountains out of Mole Hills (Poster). In Terra Futures, Ordnance Survey, Southampton, UK. 18 Sep.

**Chaudhry, O.** and Mackaness, W. (2005) Spatial Data Modelling: Derivation of 1:250,000 Topographic Database directly from OS MasterMap. In Terra Futures, Ordnance Survey, Southampton - UK. 18 Sep.