

## **Automatic Identification of Urban Settlement Boundaries for Multiple Representation Databases**

Omar Chaudhry \*, William A Mackaness,  
Institute of Geography, University of Edinburgh, Drummond St, Edinburgh EH8 9XP,  
UK, Fax: + 44 131 650 2524, [O.Chaudhry@sms.ed.ac.uk](mailto:O.Chaudhry@sms.ed.ac.uk) (O. Chaudhry)  
[William.mackaness@ed.ac.uk](mailto:William.mackaness@ed.ac.uk) (W. A. Mackaness)

\* Corresponding Author: Omar Chaudhry,  
Institute of Geography, University of Edinburgh, Drummond St, Edinburgh EH8 9XP  
Tel: +44 131 6502532, Fax: + 44 131 650 2524, [O.Chaudhry@sms.ed.ac.uk](mailto:O.Chaudhry@sms.ed.ac.uk)

**Abstract**

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. 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 generalisation. Prior to the cartographic portrayal of that information, model generalisation is required in order to derive higher order phenomena associated with the smaller scales. This paper presents a technique for automatically identifying settlement boundaries based on our understanding of what constitutes 'citiness'. From this, partonomic structures can be created that link the broad settlement with its constituent parts. The benefits of the resultant system include the automated populating of Multiple Representation Databases (MRDB), better spatial analysis and the creation of semantic reference systems capable of supporting intelligent query or zoom. The creation of such hierarchical partonomic structures provides a very useful framework within which generalisation can take place. The methodology and implementation are presented together with an evaluation of the results. Future developments are proposed.

**Keywords:** urban modelling; settlement boundary; density modelling; model generalisation; partonomic relationship

## 1. Introduction

The idea that we can abstract and portray geographic information at multiple scales in map form has existed for thousands of years (Turnbull, 1989). The process of abstraction allows humans to readily communicate various aspects and qualities of reality. Cartographers have long understood the link between scale (level of detail), the phenomena being represented and the task (theme). Various web based map services now support multi scale browsing of geographic information. As an example, if a user wished to view individual houses then Ordnance Survey's (OS) 1:25 000 product would be appropriate. A more generalised 'block view' is available at 1:50 000 scale and 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). We represent geographic information at different levels of detail in order to discern fundamentally different patterns, processes and properties (Mackness, 2007), in support of both conventional multiscale mapping (in qualitative map form), and for the purposes of spatial query, quantitative analysis and Exploratory Data Analysis (EDA).

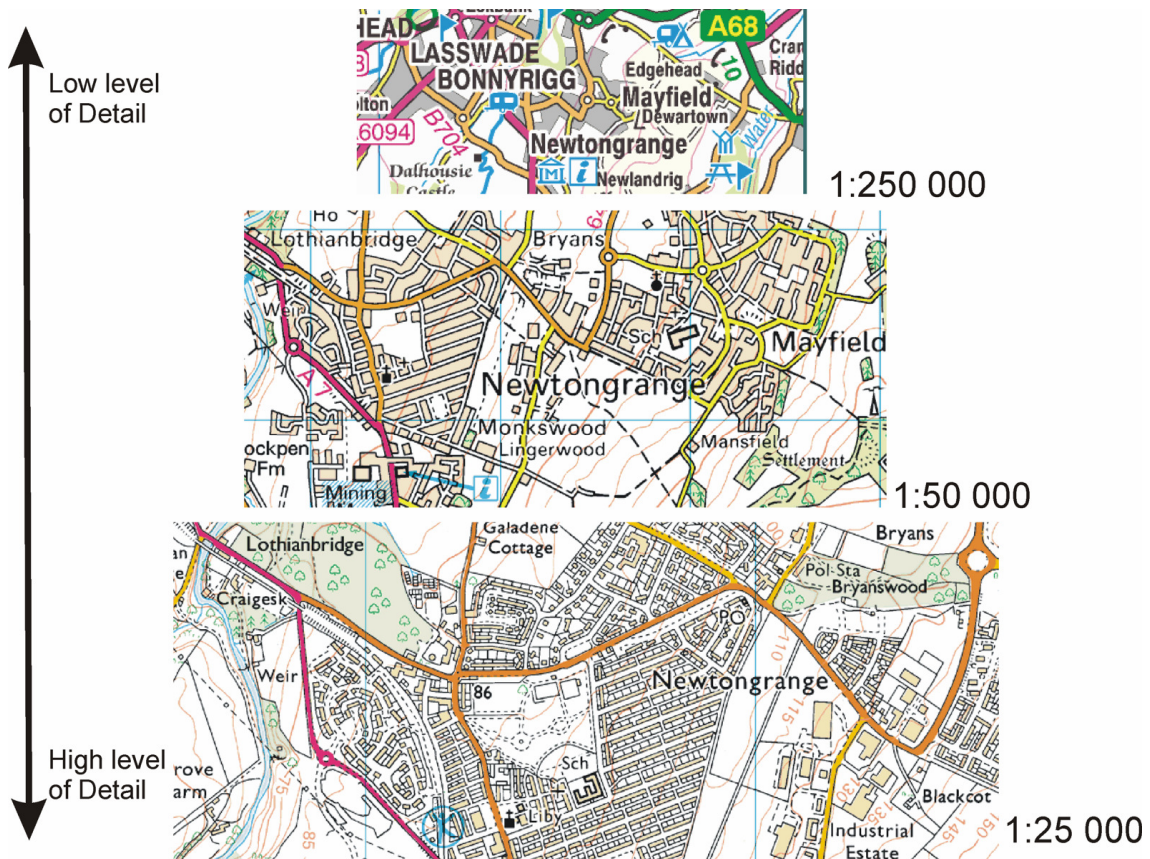


Figure 1: Different phenomena represented at different levels of detail (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved).

To fulfil this requirement National Mapping Agencies (NMAs) maintain spatial databases at different levels of detail - databases that store multiple representations of the same geographic phenomena. In the absence of links between these multiple representations, it is necessary to apply any changes across the scales. A more efficient alternative would be to reflect changes once, at the finest scale, and to use generalisation

techniques that automatically reflected the knock-on changes at lower levels of detail (Kilpeläinen and Sarjakoski, 1995), so that for example, recording of a new building development on green field sites would lead automatically to revision of the urban settlement boundary. Several researchers have proposed various ways of linking different representations of the same geographic phenomenon in such databases (Kilpeläinen and Sarjakoski, 1995; Timpf and Frank, 1995; van Oosterom, 1995; Devogele *et al.*, 1996; Jones *et al.*, 1996; Harrie and Hellström, 1999; Kilpeläinen, 2001) for the purposes of multi scale display and updating. The resultant database is called a Multiple Representation Database (MRDB). These MRDBs can either be created by connecting objects among existing databases or by generating smaller scale representations from a single large scale database via automatic generalisation (Sarjakoski, 2007). It is not always easy to link objects among existing databases because different cartographic operations (such as typification) complicate the matching process (Harrie, 2001). The second approach is more systematic (repeatable and reversible) and rigorous; one in which a series of target databases can be created via the process of model generalisation enabling description of the precise method by which the various scaled representations are derived.

### *1.1. Model and Cartographic Generalisation*

Populating MRDBs with objects associated with low levels of detail is achieved via *model* generalisation (Kilpeläinen, 1997). Thus model generalisation forms the basis from which these representations can be optimally visualised using *cartographic* generalisation (Kilpeläinen, 1997; Weibel and Dutton, 1999). In creating these 'higher order' or 'composite' objects we wish to preserve those salient or characteristic qualities appropriate to the intended scale, thus revealing a different set of interdependencies, topological qualities and patterns among higher order phenomena. This differs from cartographic generalisation which aims to improve the visual effectiveness and readability of a map (Brassel and Weibel, 1988), with the final output being dependent upon the synergistic operation of both sets of techniques (Kilpeläinen, 1997) (Figure 2).

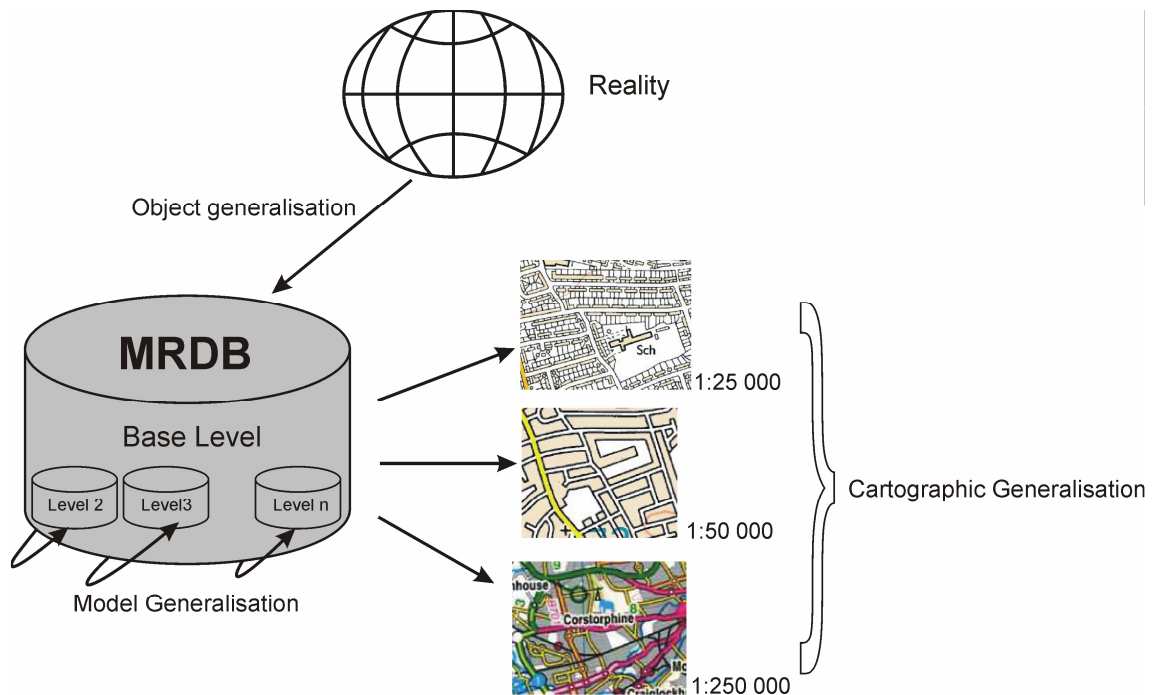


Figure 2: Model and Cartographic Generalisation (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved).

### 1.2. Modelling boundaries

Model generalisation is more than a process of subselecting the data. It is about the creation of higher order objects such as cities, forest regions, and mountain ranges from lower order objects in the source database (such as buildings, trees and groups of hills). Several techniques in the past have been proposed for creating such transformations based on classification hierarchies or taxonomies (Molenaar, 1998; Downs and Mackaness, 2002; Liu, 2002). Such techniques are appropriate over small changes in scale (van Smaalen, 2003) but over large scale changes new classes of objects need to be created from different classes of objects. In this research we are interested in automatically detecting boundaries of urban settlement from the building class. These boundaries can then act as ‘containers’ for all the other types of objects that are ‘part of’ any given urban settlement. There are many ways by which we might define urban settlements. They may be large and densely populated regions; they 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 and Yang, 2002).

There is no standard international definition of a city and many of the administrative definitions of a city are rather circular in their definition (Heikkila *et al.*, 2003; Angel *et al.*, 2005; Small *et al.*, 2005). Yet having systematic measurements for 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 and 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 (Mesev *et al.*, 1995; Heikkila *et al.*, 2003). Population density is commonly used (Bulger and Hunt, 1991) though research has shown there to be disparity between population growth and land consumption (Esnard and Yang, 2002). Similar to measures of population density, what is proposed in this paper is a definition of the urban extent based on the area and density of buildings (stored in a vector format) – the assumption being that the built environment reflects a set of social and economic activities that define ‘city’. Though this work has relevance in a range of applications, it is in anticipation of being used in a cartographic context, in the creation of mapping for notional display at 1:250 000 scale mapping (where it is not appropriate or possible to display fine scale phenomenon such as individual buildings). The building footprint is readily available in vector format from National Mapping Agencies and the examples presented here utilise the very detailed MasterMap (Topography layer) product of Ordnance Survey (the National Mapping Agency of Great Britain) and BDTopo data from the IGN France – the national mapping agency of France.

## **2. Methodology**

Initial ideas in the creation of a settlement boundary centred on the use of clustering techniques (section 2.1). This proved to be limiting but led to an approach in which objects typical of the city were used based on their density and size (section 2.2). Sections 2.3 illustrates how these ‘settlement boundaries’ were formed. From a partonomic perspective (Section 2.4), it was then possible to identify all the different objects (from different classes) that were ‘part of’ any given settlement. This information was used to enrich the source database (Figure 3); the enriched source was then used to create settlement objects (composite objects) from the aggregation of objects in the source database (section 2.5). The overall methodology is summarised in Figure 3.

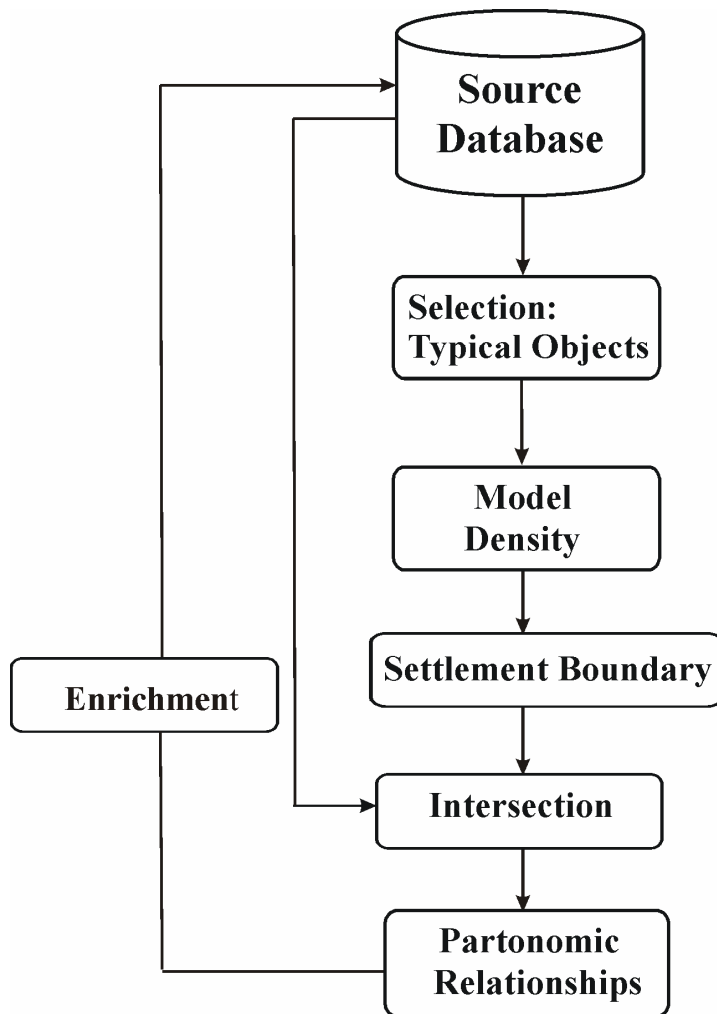


Figure 3: Overall Methodology.

### 2.1. Boundary detection

The degree to which we can precisely define the boundary of a geographic object varies enormously (Burrough and Frank, 1996; Campari, 1996). This observation is reflected in research on the modelling of fuzzy boundaries (Clementini and Felice, 1996; Cohn and Gotts, 1996). Nevertheless the boundary that separates the entity from its environment is one of the marks of its individuality (Casati *et al.*, 1998). Some boundaries are general in their form, and it is appropriate to represent those boundaries at a small scale (such as the extent of a wetland, or a mountainous region) whilst other boundaries can only meaningfully be conveyed at the large scale (such as a conveying the detail in a property boundary). A systematic treatment of boundaries has been attempted by Smith(1995) , who argued that boundaries can be divided into two basic types: bona-fide boundaries and fiat-boundaries (Smith and Varzi, 1997; Smith and 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 (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 in virtue of the different sorts of demarcations effected cognitively by human beings' (Smith 1995, p477). 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.

This paper focuses on derivation of the fiat boundaries of urban settlements and we use the concept of the 'good or necessary parts' (Cruse, 1986; Tversky, 1990; Gerstl and Pribbenow, 1995) of the target objects whose fiat boundary we want to create. For settlements we use 'buildings' as the 'good part'. We hypothesize that these objects are typical members of urban settlements reflecting the economic and social activities of the settlement (Figure 4a). From our empirical evaluation of existing cartographic settlement boundaries at 1:250 000 scale (OS Strategi Data) it became clear that settlement boundaries were created 'around regions with a high concentration of buildings' (as stated in the 1:250 000 data set specification (Ordnance Survey, 2005)). Our initial approach was to cluster the objects based on a distance threshold and create boundaries around each cluster (Figure 4b) (Chaudhry and Mackaness, 2005). This approach showed that small sparse building groups at the city periphery ('noise' or 'outliers' – Figure 4a) 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. It was therefore necessary to devise a method that could group buildings in a way that took into account their local concentration.

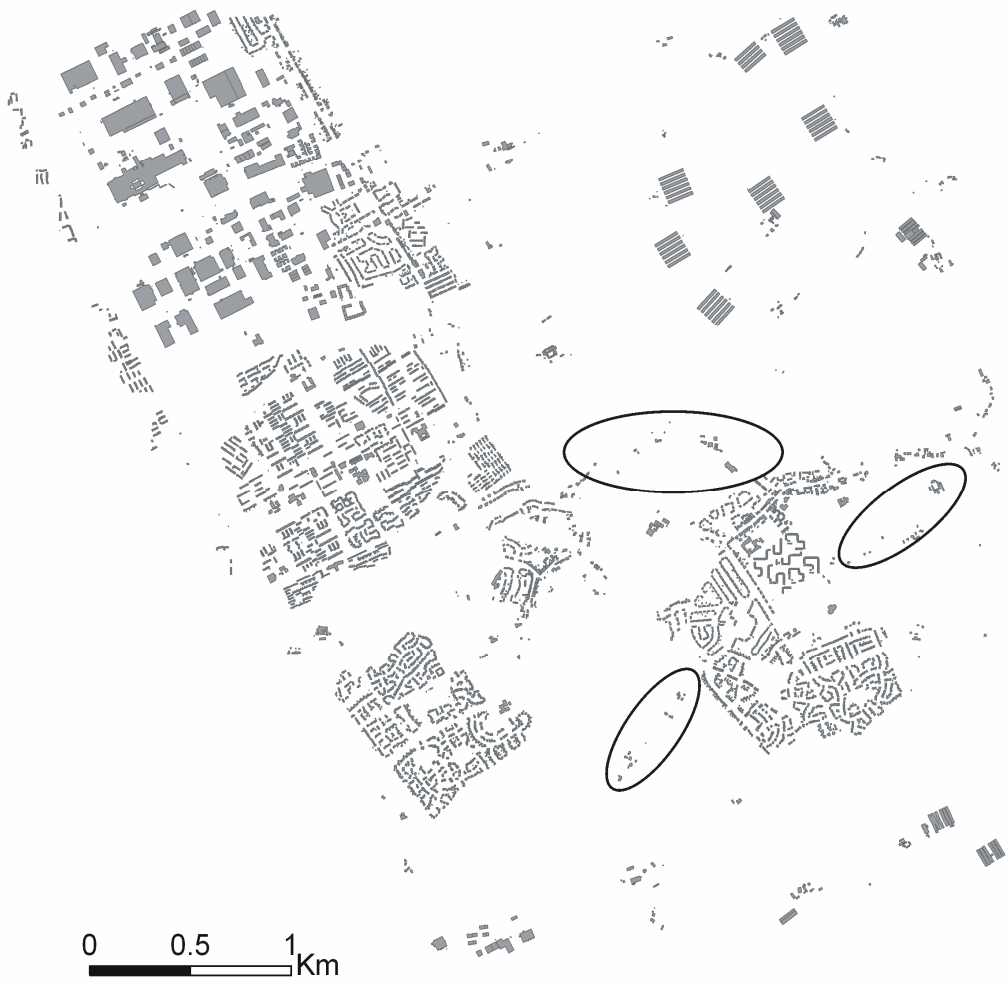




Figure 4: 4a) Example of the high variability in patterns of building distribution with low density regions ringed (taken from OS MasterMap Topography Layer). 4b) Boundary generated by a clustering algorithm (Chaudhry & Mackaness, 2005) with poor conceptual fit of the city extent (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved).

## 2.2. Calculation of Citiness

As a solution to this problem (Figure 4b) 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 1).

$$c_j = \frac{\sqrt{a_j} \sqrt{\sum_{i=1}^n a_i}}{\sum_{i=1}^n d_i^2} \quad (1)$$

In Equation 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  $c$  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, and in anticipation of notional display at a scale of 1:250 000, it was found to be sufficient to consider the fifty closest neighbouring buildings. Figure 5 shows the normalised output surface created based on the values of citiness calculated for each building shown in Figure 4a. These values of citiness were then used to create a settlement boundary. This boundary around the settlement was then used to identify all the objects that lay within the settlement (section 2.3).

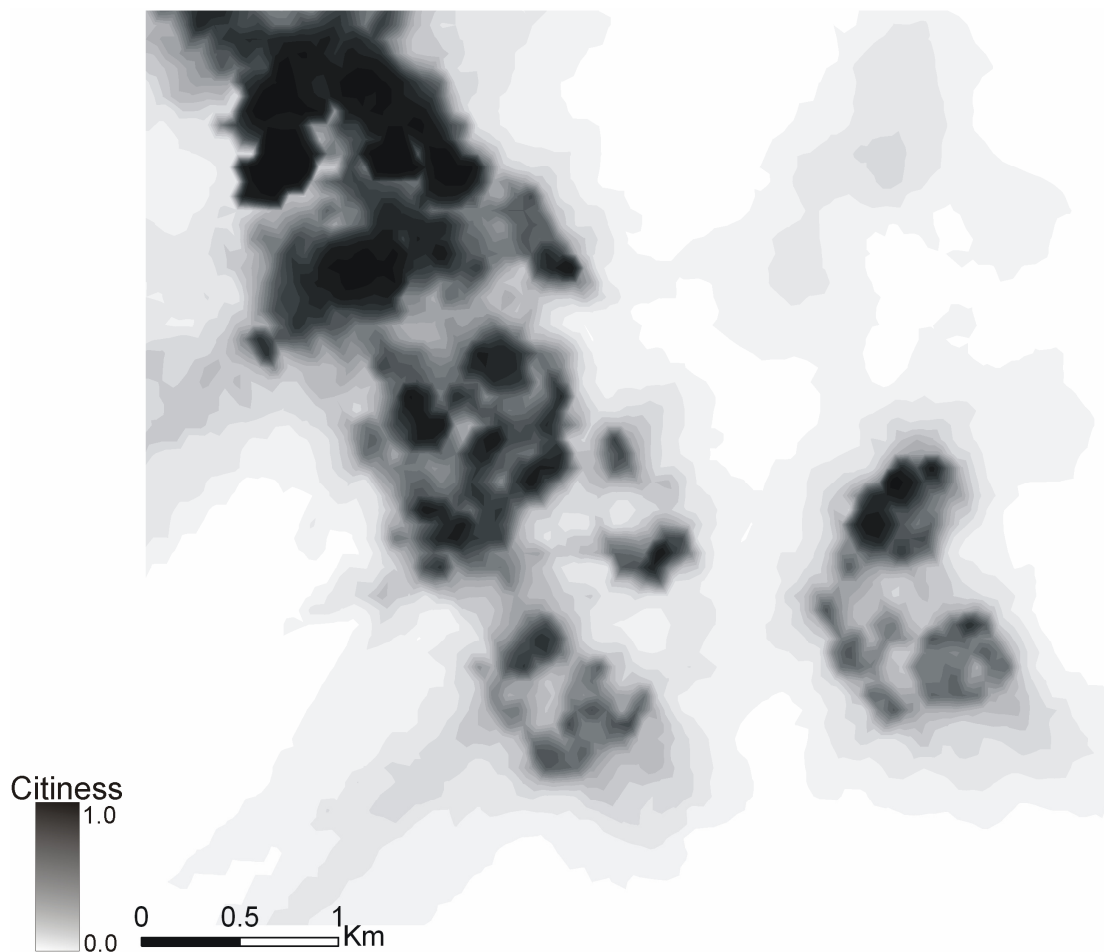


Figure 5: Interpolated surface based on citiness values. The three prominent settlements are ‘Livingston’, ‘Mid Calder’ and ‘East Calder’ in Scotland.

### 2.3. *Creation of the Boundary of the Settlement Container*

From the surface (Figure 5) we wish to identify a discrete region that is deemed to ‘contain’ the settlement. This is done by expanding and then aggregating the overlapping buildings. The areal extent of each building is expanded by the value  $e$  according to the value of  $c$ . The amount of expansion is calculated using equation 2.

$$e_a = k.c_a \quad \text{provided } e_a \leq k \quad (2)$$

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  $a$ . This idea of expansion is illustrated in Figure 6. In Figure 6a objects are grey scaled according to their value of  $c$ . These values are used to expand the objects according to Equation 2 resulting in Figure 6b. Where buildings overlap, they are aggregated into one boundary polygon (Figure 6c). 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. A rule was used derived from an Ordnance Survey map specification (1:250 000) namely that ‘for a settlement the area has to be equal to or greater than 0.01 sq km and the area of a open space (hole) within a settlement has to be equal to or greater than 0.5 sq km to be retained’ (Ordnance Survey, 2005) . Thus small holes (Figure 6c) or islands (where the area is below the area threshold are absorbed into their containing object (Figure 6d). 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.

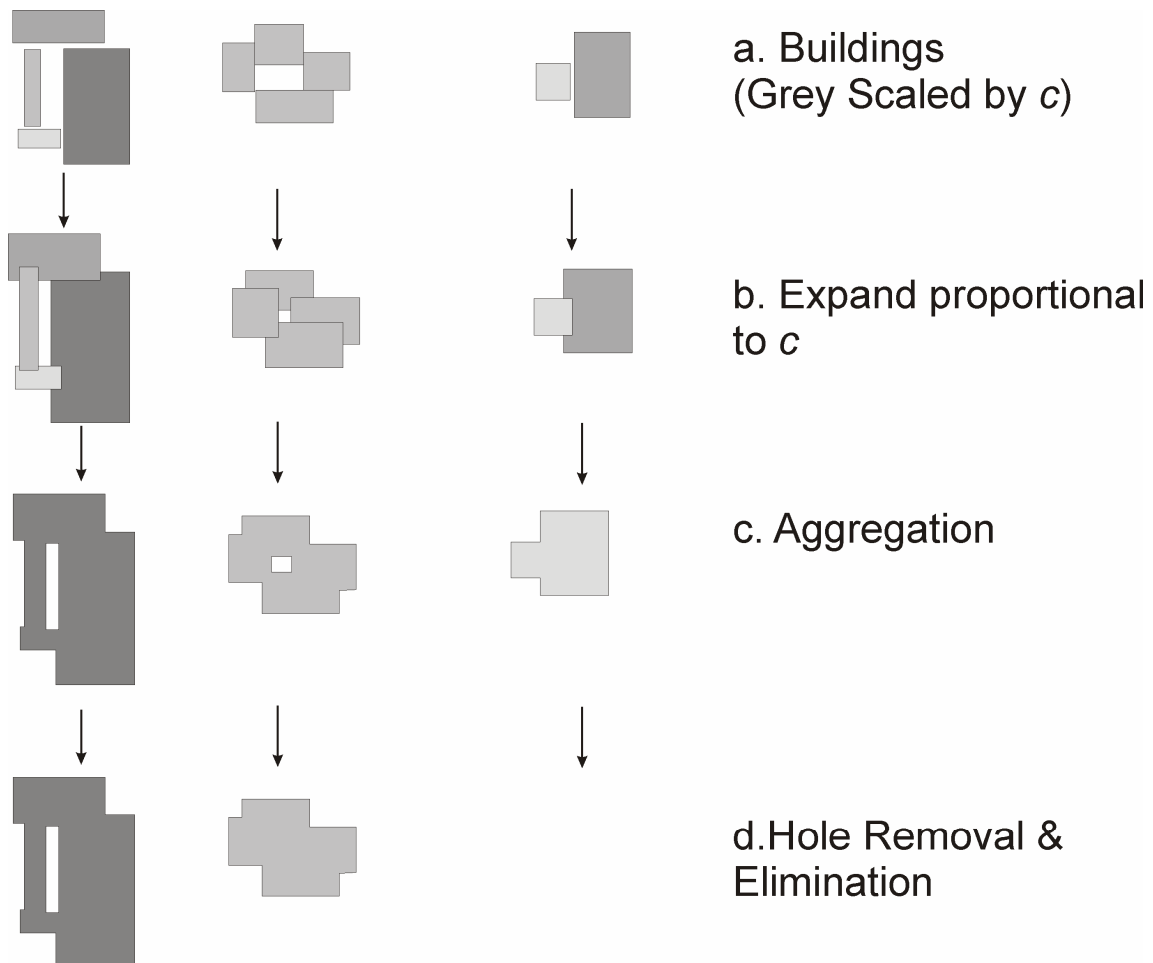


Figure 6: Process of expansion, aggregation and elimination for three examples.

The resultant boundary for the data in Figure 4 is illustrated in Figure 7. Note that the low density objects ringed in Figure 4b are no longer part of the resultant boundary. The

output boundaries generated by this approach were named by linking them with a dataset containing town names (available via the OS Strategi dataset). This dataset contains an annotation layer in which names are stored as point objects, together with an x, y coordinate. For each settlement boundary the naming algorithm found the nearest point objects that were within a specified distance (determined empirically and set to 50m) of the boundary objects. The boundary id, name and point coordinate are then stored in a name table. In this way more than one name can be associated with a particular boundary (Figure 7).



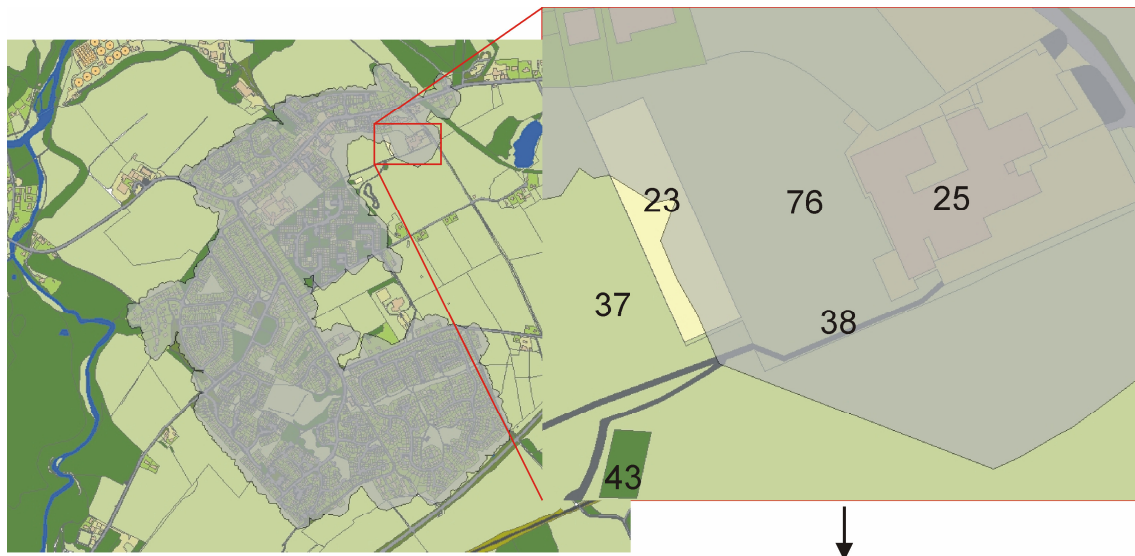
Figure 7: Resultant settlement boundaries generated by the algorithm superimposed on the input buildings (selected from OS MasterMap). (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved).

#### 2.4. *Creation of Partonomic Information*

The resultant boundary (Figure 7) acts as a ‘container’ and we can use this information to identify the partonomic relationships between different classes of objects, reflecting the idea that the city does not just contain buildings but that other geographic entities such as transportation networks, green spaces, and rivers are ‘part of’ and characterise any given city. Information relating to partonomic relationships is useful in respect of cartography, spatial analysis and model generalisation. For example a road that is part

of a city serves a different function to a road that is part of the rural setting. The former may service the daily commute, whilst the latter may serve to connect settlements. Knowledge of these different properties informs the cartographic process (in terms of emphasis, design constraints and symbolisation) and can support more intuitive forms of spatial query (Chaudhry and Mackaness, 2006b). For example a model that estimates driving travel time may wish to identify the major roads that connect cities (those that are not part of the urban settlement) separate from those that lie within. Using the settlement boundary, we can determine the partonomic relationships of the source objects in terms of the higher order objects in the target database. This is information that is not explicit in the source database but needs to be identified prior to database transformation via model generalisation (Molenaar, 1998; van Smaalen, 2003; Bobzien *et al.*, 2006). Such information enables the creation of aggregation hierarchies (Peng, 1997; Molenaar, 1998) which in turn support more intuitive data interaction (Frank and Timpf, 1994).

If the source objects are completely ‘contained by’ the resultant boundary, they are deemed to be part of the higher order object. If ‘disjoint’ they may be considered part of some other higher order object. If it is an ‘overlap’ relationship then the percentage of area intersection can be used to govern the degree of membership enabling an object to be part of more than one higher order entity. Figure 8 contains an inset showing different objects (ID = 23, 76, 25, 37, 38, 43) lying across the container boundary. The table shows the degree of area overlap.



ID	Inside	Disjoint	Overlaps	Overlap Area %	Part of
23	No	No	Yes	75%	Settlement_1
76	Yes	No	No	100%	Settlement_1
25	Yes	No	No	100%	Settlement_1
37	No	No	Yes	2%	Settlement_1
38	No	No	Yes	100%	Settlement_1
43	No	Yes	No	0%	General Land

Figure 8: Determining the membership of the partonomy.

## 2.5 Utility of Partonomies

The partonomic structures determined from this approach (section 2.4) make explicit information that is implicit in a map context. Once these relationships have been enriched in the source database (Figure 3), they can be utilised for the purpose of aggregation of source objects into composite objects. For example the settlement object demarcated in Figure 9b has been created from the aggregation of all the objects that were part of the settlement container boundary (Figure 8) and fulfil additional criteria such as adjacency, minimum degree of membership and object rank (Chaudhry and Mackaness, 2006a).

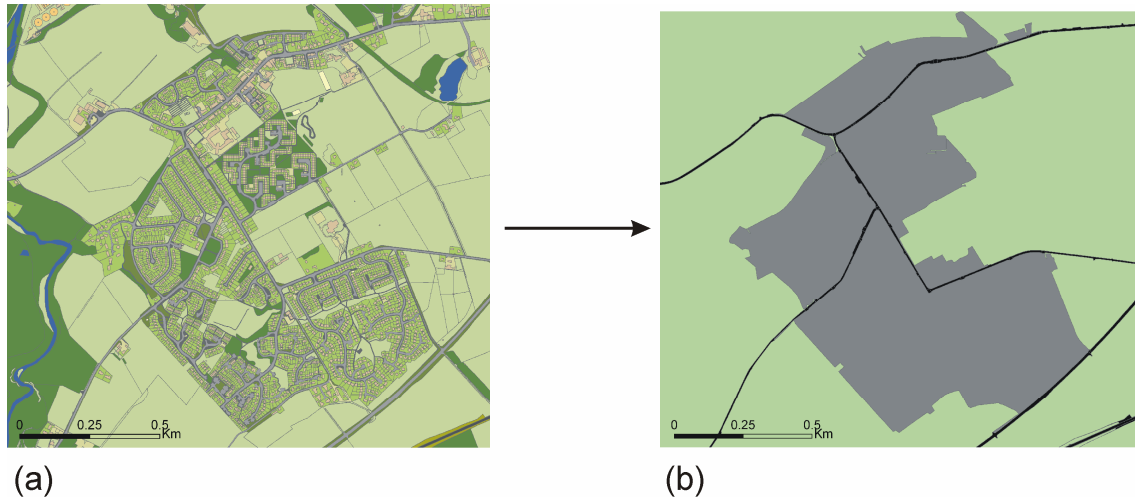


Figure 9: Aggregation of source database objects into composite objects based on partonomic relations (Figure 8) via model generalisation (Chaudhry & Mackaness, 2006b). (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved).

## 3. Implementation and Evaluation

The platform selected for the implementation of this methodology was Java, SQLJ and Oracle 10g. Oracle 10g supports the geometrical and topological functions defined by the Open Geospatial Consortium (OGC). Results of the approach on areas around Livingston in Scotland (Figure 4) have already been presented in Figure 7. Figure 10a and 11a show settlement boundaries for two other areas within Great Britain. Using the same parameter settings the approach was also tested on an IGN France building dataset (Figure 12a).

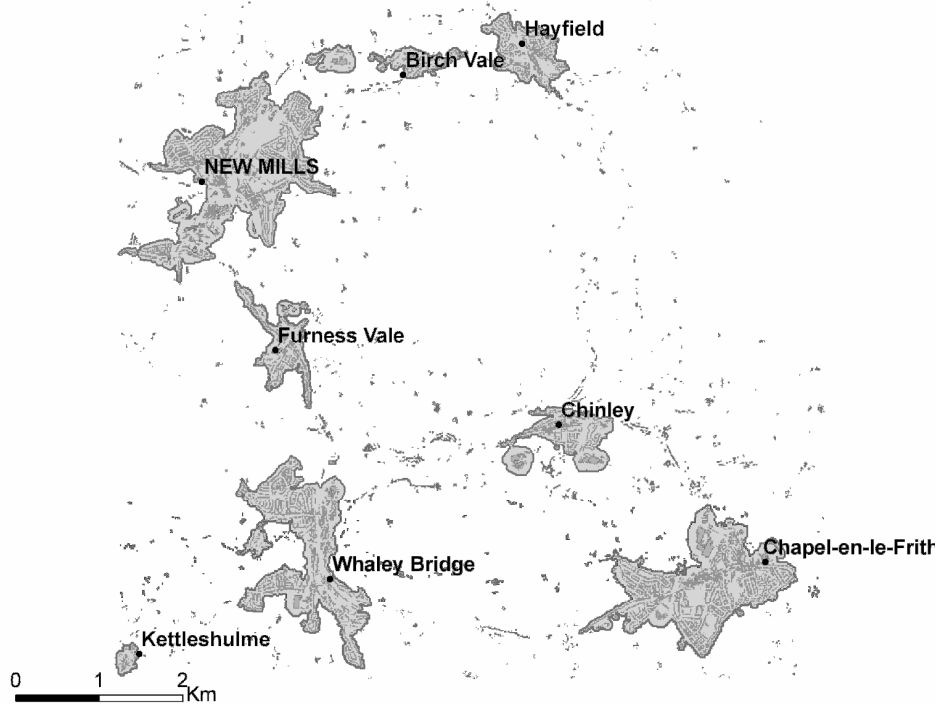


Figure 10a: Resultant boundaries generated by the algorithm with input buildings selected from OS MasterMap (Peak District, UK).

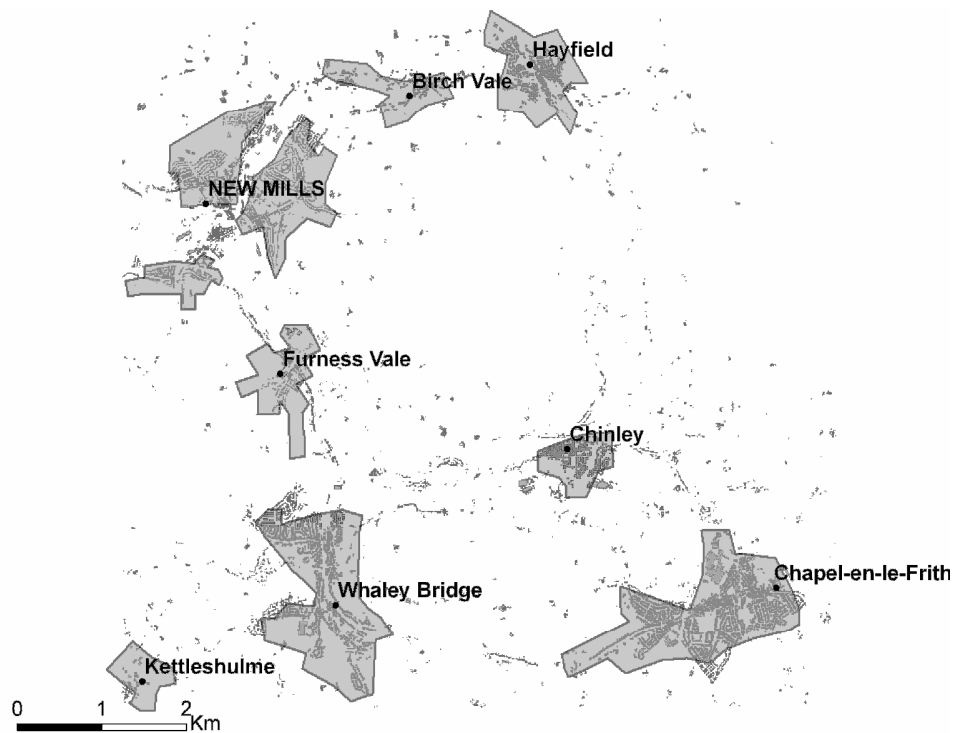


Figure 10b: Equivalent area showing Strategi data (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved).

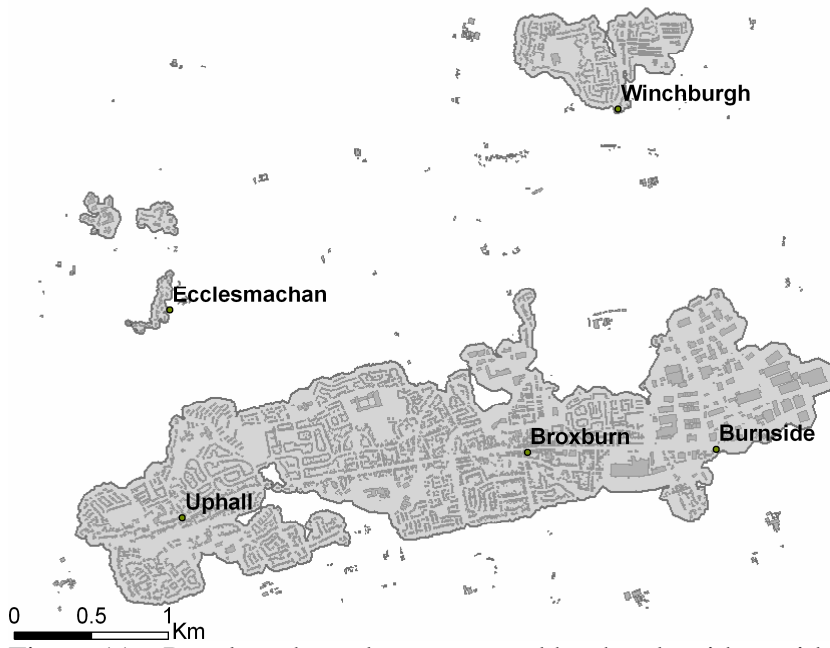


Figure 11a: Resultant boundary generated by the algorithm with input buildings selected from OS MasterMap.

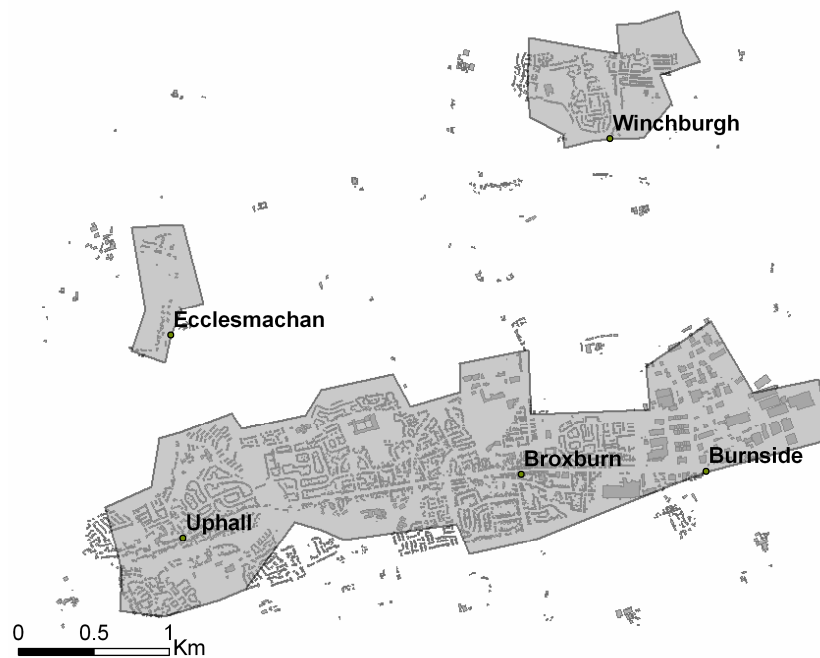


Figure 11b: OS Strategi settlement boundary with input buildings selected from OS MasterMap. (Mapping is Ordnance Survey ©Crown Copyright. All rights reserved).



Figure 12a: Resultant boundaries generated by the algorithm with input buildings selected from IGN BDTopo®, (Pamiers in France).



Figure 12b: IGN France settlement boundaries at 1:250 000 with input buildings selected from IGN BDTopo® dataset.

### *3.1. Evaluation of settlement boundaries*

Visual comparison was undertaken between various geographic regions. Figure 10a, 11a being two such examples using Great Britain data, with OS Strategi settlement layer

for comparison (Figure 10b and 11b). Figure 12a is part of an IGN France dataset, and the 1:250 000 scale dataset for comparison (Figure 12b). We acknowledge that visual comparison with paper maps is subjective (Weibel and Dutton, 1999; Mackaness and Ruas, 2007) but that this does give some indication of the success of the algorithm. The overall structure of boundaries is observed to be quite similar; there are various differences: 1) the settlement boundaries are not smooth nor simplified whereas the cartographic comparators are. The settlement boundary is a database object and is a result of model generalisation, not of cartographic generalisation; 2) the cartographic outputs from both OS and IGN France are poorly aligned and poorly reflect the distribution of buildings. This is probably due to the application of cartographic generalisation techniques (such as displacement) and secondly the representations are out of date, being less frequently updated than at the large scale.

The results were also evaluated by a cartographer and the generalisation team at OS. The cartographer highlighted the limitation of comparison with OS Strategi data set "...comparison with Strategi data is not a good measure, since Strategi is created from 1:250 000 data that is used to produce the printed paper map. The original data was captured from an existing paper copy and takes on board all changes to shape and position that have been applied by cartographers over the last 80 to 90 years. Features might have a reasonable relationship to each other but geographical position can be 500m (or more) out..". According to the cartographer the results created from our approach "...are far more accurate from a positional point of view.." and "From a 'Settlement' dataset viewpoint, where we use the data to attribute point data (e.g. Address) it has more going". They went on to conclude that the results "...could be useful as a revision guide for 1:250 000, but this would need testing in a production environment."

Comments from the generalisation team highlighted the potential use of the results as input to automatic cartographic generalisation: "...the results are good, but will be much more valuable in an automatic process than as input for a cartographer... the real value of your algorithm is in providing automatically a boundary for built-up areas, that can either be geometrically refined (simplified) for delivering a cartographic representation at a given scale, or used as it is for analysis purposes (including driving generalisation processes specially designed to work in built-up contexts)". It was also suggested that the boundary could be used for the automatic determination of quality indicators to "qualify the discrepancy between the data and the result (for example, any point within a built-up area is within x meters of a building)".

### *3.2. Analysis Based on Partonomic Information*

Whilst we have highlighted the value of partonomies in map generalisation it is important to stress that their utility extends beyond the visual. Partonomic information can also be used for spatial queries based on the higher order objects within the database. In Figure 13 we can distinguish between roads solely internal to the city, and those that are 'shared' by two separate settlements. It is these candidates that would form the basis for alternate routes between the two settlements. It is important to stress that such a query would not be possible prior to partonomic enrichment, since there was no object that represented the settlement.

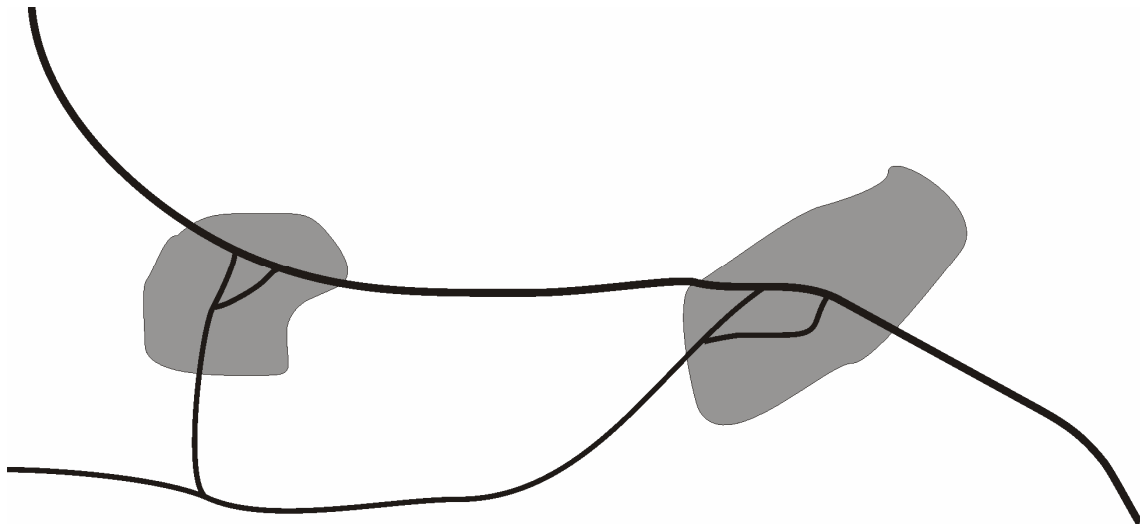


Figure 13: Part of a simple road network between two settlements.

It is also worth noting that these partonomic relations are not exclusive. There is no reason why an object cannot be part of multiple partonomies (Figure 14). Indeed it is likely that any given object would be part of more than one partonomy. For example a railway station is both part of the railway network and is also part of a city – reflecting the function (and relative importance) of the railway station as the means by which the populous connect into the network.

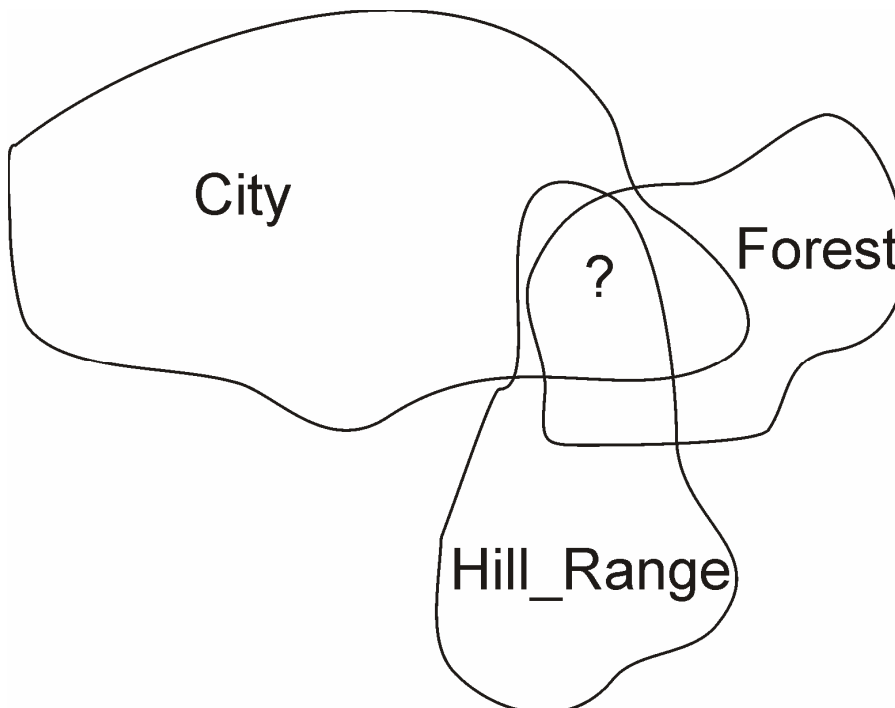


Figure 14: Using multiple partonomic information to support multi criteria spatial analysis.

Assuming we have defined the partonomies of city, forest and hill (Figure 14), we can answer the query ‘where can I find a wooded suburb with good views across the city?’. More generally and from a cartographic perspective, we would argue that thematic

mapping is all about conveying a subset of all the partonomic relationships that exist among geographic entities.

Whilst the methodology presented here has produced crisp boundaries, it is acknowledged that more generally it may be difficult to create discrete boundaries among continuous phenomena (Molenaar, 1996), that a boundary may be vague because of the fuzzy nature of the concept (Campari, 1996; Usery, 1996) and that both these factors may lead to a degree of uncertainty in the determination of partonomic relationships. Often geographic phenomena have a certain level of indeterminacy reflecting the need to define and understand continuous phenomena within a certain context of observation (Molenaar, 1993). A lot of research has been done on dealing with fuzziness in spatial objects, crisp and non crisp boundaries (Miyamoto, 1990; Molenaar, 1993; Campari, 1996; Molenaar, 1996; Goodchild *et al.*, 1998; Fisher, 2000; Duckham *et al.*, 2001; Duckham, 2002; Reinke and Hunter, 2002; Winter and Thomas, 2002). Here we have presented an approach for creating fiat boundaries of settlements based on the density of the buildings. We haven't modelled the fuzziness in the boundary and subsequently in their partonomic relationships but we believe that the 'citiness' values calculated for each buildings (and represented in the continuous surface of Figure 5) can be used as a basis for further work to model fuzziness in terms of upper and lower approximations (Pawlak, 1982; Worboys, 1998) and that this offers a more effective way of thematically conveying 'city' than just a simple polygon.

#### **4. Conclusion**

Generalisation is a process of transforming information from one form into another in order to reveal different qualities and characteristics of the phenomena being modelled. This paper has demonstrated a method by which settlement boundaries can be automatically created from their typical parts (buildings). This information enriches the database with partonomic information and provides a framework for both map generalisation and multi criteria spatial analysis. We have argued that the automatic creation of these partonomic relationships is a critical prerequisite to the automatic populating of databases in MRDBs. This paper has presented an algorithm for creating partonomic relationships between settlements and their constituent parts. One can envisage many other scale dependent partonomic structures that might define phenomena along a scale continuum; from airports, schools, train depots, to forested regions, mountainous regions or water catchments. Some can be defined mathematically/ cartometrically; others can be defined in terms of their function; most are composite. All must be defined in terms of our prototypical sense of what these things 'mean' (their 'good parts' (Tversky and Hemenway, 1984)). The challenge is in selecting appropriate criteria that efficiently defines their form according to our expectations and geographical understandings.

#### **Acknowledgements**

The authors are grateful for funding from Ordnance Survey and The University of Edinburgh. Special thanks to IGN France for providing data sets for evaluation; Ordnance Survey staff for their comments on settlement boundaries and to anonymous reviewers for their constructive comments.

## References

- Angel, S., Sheppard, S. and Civco, D. (2005). The Dynamics of Global Urban Expansion. . Transport and Urban Development Department, The World Bank
- Bobzien, M., Burghardt, D., Neun, M. and Weibel, R. (2006) Multi-Representation Databases With Explicitly Modelled Horizontal, Vertical and Update Relations. In AutoCarto, Vancouver-WA.
- Brassel, K.E. and Weibel, R. (1988) A review and conceptual framework of automated map generalization. *International Journal of Geographical Information Systems*, 2(3), 229-244.
- Bulger, D. and Hunt, H. (1991) The Plonski forest management decision support system project. In GIS '91 Symposium Proceedings, Vol. 1, pp. 101-109. Forestry Canada, Vancouver.
- Burrough, P.A. and Frank, A.U. (1996) *Geographic objects with indeterminate boundaries*. London, Taylor & Francis
- Campari, I. (1996). Uncertain Boundaries in Urban Space. In *Geographic Objects with Indeterminate Boundaries* (eds P.A. Burrough and A.U. Frank), pp. 57-69. London, Taylor & Francis.
- Casati, R., Smith, B. and Varzi, A. (1998). Ontological Tools for Geographic Representation. In *Formal Ontology in Information Systems* (ed N. Guarino), pp. 77-85. Amsterdam, IOS Press.
- 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.
- Chaudhry, O. and Mackaness, W. (2006a) Modelling Geographic Phenomena at Multiple Levels of Detail. In AutoCarto, Vancouver- WA.
- Chaudhry, O. and Mackaness, W. (2006b) Creation of Fiat Boundaries in Higher order Phenomenon. In 9th ICA Wrokshop on Generalisation and Multiple Representation, Vancouver - WA.
- Clementini, E. and Felice, P.D. (1996). An algebraic model for spatial objects with undetermined boundaries. In *Geographic Objects with Indeterminate Boundaries* (eds P.A. Burrough and A. Frank), pp. 155-169. London, Taylor & Francis.
- Cohn, A.G. and Gotts, N.M. (1996). The 'egg-yolk' representation of regions with indeterminate boundaries. In *Geographic Objects with Indeterminate Boundaries* (eds P.A. Burrough and A. Frank), pp. 171-187. London, Taylor & Francis.
- Crouch, C. (2006) *Territorial Reviews: Competitive Cities in the Global Economy*. Paris, OECD
- Cruse, D.A. (1986) *Lexical Semantics*. Cambridge, England, Cambridge Univ. Press
- Devogele, T., Trevisan, J. and 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 and E.M. Fendel), pp. 337-352. London , Taylor and Francis, Delft.
- Downs, T.C. and 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. (2002). Uncertainty and geographic information: computational and critical convergence. In *Representation in a digital geography*, Wiley: New York.
- Duckham, M., Mason, K., Stell, J. and Worboys, M. (2001) A formal approach to imperfection in geographic information. *Computers, Environment, and Urban Systems*, 25, 89-103.
- Esnard, A.-M. and Yang, Y. (2002) Descriptive and Comparative Studies of 1990 Urban Extent: Data for the New York. *Metropolitan Region URISA Journal*, 14(1), 57-62.
- Fisher, P.F. (2000). Fuzzy modelling. In *Geocomputing* (eds S. Openshaw and R. Abraham), pp. 161-186. London, Taylor & Francis
- Frank, A. and Timpf, S. (1994) Multiple representations for cartographic objects in a multi-scale tree: an intelligent graphical zoom. *Computers & Graphics*, 18(6), 823-829.
- Fujita, M., P. K. and Mori, T. (1999) On the evolution of hierarchical urban systems. *European Economic Review*, 43(2), 209-251.
- Gerstl, P. and 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.
- Goodchild, M., Monette, D., Fohl, P. and 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.
- Harrie, L. (2001) *An Optimisation Approach to Cartographic Generalisation*. Lund, Sweden, Lund University
- Harrie, L. and Hellström, A.K. (1999) A Prototype System for Propagating Updates between Cartographic Data Sets. *The cartographic Journal*, 36(2), 133-140.
- Heikkilä, E.J., Shen, T. and 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.
- Jones, C.B., Kidner, D., Luo, L., Bundy, G.L. and Ware, J.M. (1996) Database design for Multi-Scale Spatial Information System. *International Journal of Geographic Information Systems*, 10(8), 901-920.
- Kilpeläinen, T. (1997) *Multiple representation and generalisation of geo-databases for topographic maps*, Publications of the Finnish Geodetic Institute
- Kilpeläinen, T. (2001) Maintenance of Multiple Representation Databases for Topographic Data. *The cartographic Journal*, 37(2), 101-107.
- Kilpeläinen, T. and Sarjakoski, T. (1995). Incremental generalisation for multiple representations of geographical objects. In *GIS and Generalisation: Methodology and Practice* (eds J.C. Muller, J.P. Lagrange and R. Weibel), pp. 209-218.
- Liu, Y. (2002) Categorical Database Generalization in GIS. Dissertation number 88, ITC, the Netherlands
- Mackaness, W. (2007). Understanding Geographic Space. In *Generalisation of Geographic Information: Cartographic Modelling and Applications* (eds W. Mackaness, A. Ruas and L.T. Sarjakoski), pp. 1-10. Oxford, Elsevier.
- Mackaness, W. and Ruas, A. (2007). Evaluation in the Map Generalisation Process. In *Generalisation of Geographic Information: Cartographic Modelling and*

- Applications* (eds W. Mackaness, A. Ruas and L.T. Sarjakoski), pp. 89-112. Oxford, Elsevier.
- Mesev, T.V., Longley, P.A., Batty, M. and Xie, Y. (1995) Morphology from Imagery: Detecting and Measuring the Density of Urban Land Use. *Environment and Planning A*, 27(5), 759-780.
- Miyamoto, S. (1990) *Fuzzy sets in Information Retrieval and Cluster Analysis*. Dordrecht, Netherlands, Kluwer Academic Press
- Molenaar, M. (1993) Object hierarchies and uncertainty in GIS or why is standardisation so difficult. *GeoInformations-System*, 6(3), 22-28.
- Molenaar, M. (1996). A Syntactic Approach for Handling the Semantics of Fuzzy Spatial Objects. In *Geographic Objects with Indeterminate Boundaries* (eds P.A. Burrough and A. Frank), pp. 207-224. London, Taylor & Francis.
- Molenaar, M. (1998) *An Introduction to the Theory of Spatial Object Modelling for GIS*. London, Taylor and Francis
- Ordnance Survey (2005) 1:250,000 specification. *Ordnance Survey internal document*.
- Pawlak, Z. (1982) Rough Sets. *International Journal of Computer and Information Sciences*, 11(5), 341-356.
- Peng, W. (1997) Automated Generalization in GIS, Doctoral Dissertation, Wageningen Agricultural University and ITC
- Reinke, K.J. and Hunter, G.J. (2002). A Theory for Communicating Uncertainty in Spatial Databases. In *Spatial Data Quality* (eds W. Shi, P.F. Fisher and M.F. Goodchild), pp. 77-101. London, Taylor & Francis.
- Sarjakoski, L.T. (2007). Conceptual Models of Generalisation and Multiple Representation. In *Generalisation of Geographic Information: Cartographic Modelling and Applications* (eds W. Mackaness, A. Ruas and L.T. Sarjakoski), pp. 11-36. Oxford, Elsevier.
- Small, C., Pozzi, F. and 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 and D. Mark), pp. 475-484, Springer Verlag.
- Smith, B. and Varzi, A. (1997) The Formal Ontology of Boundaries. *Electronic Journal of Analytic Philosophy*, 5(5).
- Smith, B. and Varzi, A.C. (2000) Fiat And Bona Fide Boundaries. *Philosophy And Phenomenological Research*, 60, 401-420.
- Timpf, S. and Frank, A. (1995) A multiscale DAG for cartographic objects. In *Proceedings Auto-Carto*, Vol. 12, pp. 157-163.
- Turnbull, D. (1989) *Maps are territories Science is an atlas*, Deakin University
- Tversky, B. (1990). Where partonomies and taxonomies meet. In *Meanings and prototypes: Studies on Linguistic Categorization* (ed S.L. Tsohatzidis), pp. 334-344, London: Routledge.
- Tversky, B. and Hemenway, K. (1984) Objects, parts, and categories. *Experimental Psychology*, 113, 169-193.
- Usery, E.L. (1996). A Conceptual Framework and Fuzzy Set Implementation for Geographic Features. In *Geographic Objects with Indeterminate Boundaries* (eds P.A. Burrough and A. Frank), Vol. 3, pp. 71-85. London, Taylor & Francis.
- 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. Muller, J.P. Lagrange and R. Weibel), pp. 120-132. Bristol, Taylor & Francis.
- 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
- Weibel, R. and Dutton, G. (1999). Generalising Spatial Data and Dealing with Multiple Representations. In *Geographical Information Systems* (eds P. Longley, M.F. Goodchild, D.J. Maguire and D.W. Rhind), pp. 125-156. New York, John Wiley.
- Winter, S. and Thomas, B. (2002). Hierarchical Topological Reasoning with Vague Regions. In *Spatial Data Quality* (eds S. Wenzhong, P. Fisher and M.F. Goodchild), pp. 35-49. London, Taylor & Francis.
- Worboys, M. (1998) Imprecision in Finite Resolution Spatial Data. *GeoInformatica*, 2(3), 257-279.