A Road Network Selection Process Based on Data Enrichment and Structure Detection

Guillaume Touya

A Road Network Selection Process Based on Data Enrichment and Structure Detection

Guillaume TOUYA

COGIT Laboratory, Institut Géographique National, 2 av. Pasteur, 94160 St-Mandé, France

ABSTRACT

In the context of geographical database generalisation, this paper deals with a generic process for road network selection. It is based on the geographical context that is made explicit and on the characteristic structure preservation. It relies on literature work that is adapted and gathered. The first step is to detect significant structures and patterns of the road network such as roundabouts or highway interchanges. It allows to enrich the initial dataset with explicit geographic structures that were implicit in initial data. It helps both to explicit the geographical context and to preserve characteristic structures. Then, this enrichment is used as knowledge input for the following step that is the selection of roads in rural areas using graph theory techniques. After that, urban roads are selected by means of a block aggregation complex algorithm. Continuity between urban and rural areas is guaranteed by modelling continuity using strokes. Finally, the previously detected characteristic structures are typified to maintain their properties in the selected network. This automated process has been fully implemented on Clarity™ and tested on large datasets.

1. Introduction

Generalisation is a process that seeks to summarise geographical information from a geographic database in order to produce a less detailed database or map. When the goal is to derive a new database and not a map, it is called model generalisation and it is not constrained by cartographic symbols (Weibel and Dutton, 1999). Selection, that can be considered as an operation concerning the abstraction of the database (Mackaness, 2007), is a key step to model generalisation. It consists in choosing the relevant information in relation to the target map or database specifications. This paper presents a generic road network selection process based on data enrichment using spatial analysis.

Selection seeks to choose the essential and relevant elements of a geographic database but also to maintain the main characteristics of geographic information while reducing the level of detail. As roads are important features of maps and geographic databases, road network selection is a key topic of generalisation that has already been tackled in the past (Richardson and Thomson, 1996, Thomson and Richardson, 1999, Ruas, 1999, Jiang and Claramunt, 2004). Other papers deal with structure and pattern recognition in order to ease road selection (Marshall 2005, Heinzle et al., 2005, Heinzle and Anders, 2007).

Road network selection for model generalisation raises several problems. First, selecting roads requires to infer the usage of the roads that heavily depends on roads implicit geographical context: methods have to be found to make the context explicit for an automatic process. Then, a hierarchy of roads is necessary in order to select the most important ones but road saliency is hard to infer: it depends on geometry, attributes, topology and role in the network i.e. it depends on implicit geographical context. Moreover,

road networks in urban or rural areas are clearly different in road density and network shape and such differences have to be maintained by selection; road networks contain typical complex implicit road structures that a good selection should maintain. Finally, it is hardly manageable to precisely quantify the selection reduction level (in objects number? In meters? In relation to specifications?). Because of the diversity of data and generalisation problems, all previous work on the topic dealt with a specific part of these problems of road network selection. There has not been any attempt yet to develop a generic method that would gather this material in order to tackle all the problems globally. Thus, the objective of this work is to design an automatic and generic (for different data and resolution changes) process that would deal with the global problem of road network selection in model generalisation as well as developing solutions for the missing bricks of the process. The proposed method enriches data with the proper road structures and selects differently rural and urban networks thanks to the enriched data.

The next part of the paper describes the design constraints and the overall methodology for road network selection. The third part deals with data enrichment by structure detection in road networks and with the structures typification. The fourth part deals with the selection steps of our proposed process. Section five illustrates the process with results obtained with a real database. Finally, last section draws some conclusions and discusses further work.

2. Methodology

2.1. Hypotheses

Some hypotheses are made on the input vector data in order to use the proposed methodology. But the properties required for the data are quite generic and are met by most of vector geographic databases that contain roads. The last two requirements are optional as they are not necessary for the process but they improve its results. In the parts of the paper that describe how the optional data are used, alternative solutions are proposed. Here are the seven data requirements:

– linear road data is required rather than surfacic roads.
– the road network needs to be correctly topologically connected.
– only one attribute representing roughly road classes is required.
– no traffic direction data is required.
– urban areas are necessary to separate urban and rural roads, or buildings to create the urban areas with a method from Boffet (2000).
– optionally, facility points (airports, train stations, schools, malls...) can improve the process quality.
– optionally, buildings can help in some structure detection algorithms.

The test data of this work are a large dataset (150 x 100 km) extracted from BD TOPO®, the 1m resolution topographic vector database produced at IGN France. The road data correspond to our requirements and the optional requirements like facility points and buildings are also extracted from this database. The test area is characterised by a large heterogeneity of landscapes, which allows to anticipate the work validity on other areas of the same database.

2.2. Experiments

The methodology used to design the presented road network selection method was based on experiments: in order to have a clear view of the problems raised by road network selection, several tests and literature algorithm implementations were carried out on the test data. Then the methodology consisted in gathering literature resources that appeared to be useful, in adapting them to make them work together: according to the experiments results, some algorithms were mixed, chained, some were modified and some were dropped when alternative ones proved to be more effective. When no resource was available, new algorithms were developed. The last ones, as well as the adaptations, are more focused on in this paper.

2.3. Overall Process

This methodology led to the design of a complete road selection process composed of five important steps (Figure 1). First, the amount of implicit information contained in road networks requires a huge enrichment of initial data to provide what Brassel and Weibel (1988) called *structural knowledge*. The enrichment also provides some particular structures of the network that rather need to be typified than being simply selected to maintain properly the structures during selection. Then, the problems raised by urban networks appeared to be very specific: road density is very high and network shape and patterns are here the major factor of selection (to be able to still differentiate old European cities and modern US ones after selection). In rural areas, the problem is more about removing the less significant roads. As a consequence, two different processes have to be applied on rural and urban areas immediately followed by a process that guarantees continuity between the two types of area. Each step of the process will be described in detail in the following sections of the paper.
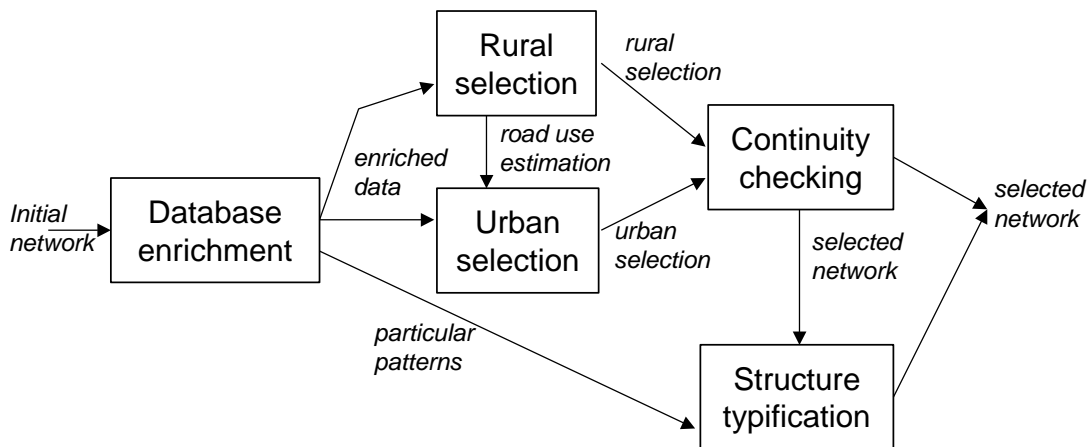

Figure 1. the overall process of road network selection. In italics, the different inputs and outputs of the process steps.

3. Data Enrichment

Brassel and Weibel (1988) recommended data enrichment as a prerequisite for automatic generalisation to formalise the *structural knowledge* contained in geographic information. This structural knowledge includes structure and patterns detection whose importance in generalisation has been claimed by Mackaness and Edwards (2002). Mackaness and Edwards (2002) define patterns as a property (shape, orientation...) within an object or between objects that is repeated with sufficient regularity. A structure can be defined as a particular distribution of objects representing an implicit geographic entity. It corresponds to what Ruas (1999) calls *meso objects*. We consider road meaningful patterns as structures and we will use only this concept from now on. Road networks perfectly illustrate the importance of structures as roads form complex man-made networks full of particular structures. This section presents the main developed enrichments and the typification of some detected structures.

3.1. Road Structures Detection

This section presents the road structures that we consider as important to detect for selection and the algorithms developed for their automatic detection in a vector database.

3.1.1. Crossroads Classification. Road crossroads can be considered as an atomic element of road network and it is very useful to better characterise them for selection according to Heinzle et al. (2005). For example, y-nodes correspond to slip roads and T-nodes represent crossroads between a minor and a major road. A simple taxonomy of crossroads has been designed based on Grosso (2004) and Sester (1995). It contains T-shaped nodes (T-nodes), Fork nodes, y-nodes (slip road end), Cross-shaped nodes (CRS-nodes), Star nodes and Undifferentiated nodes.
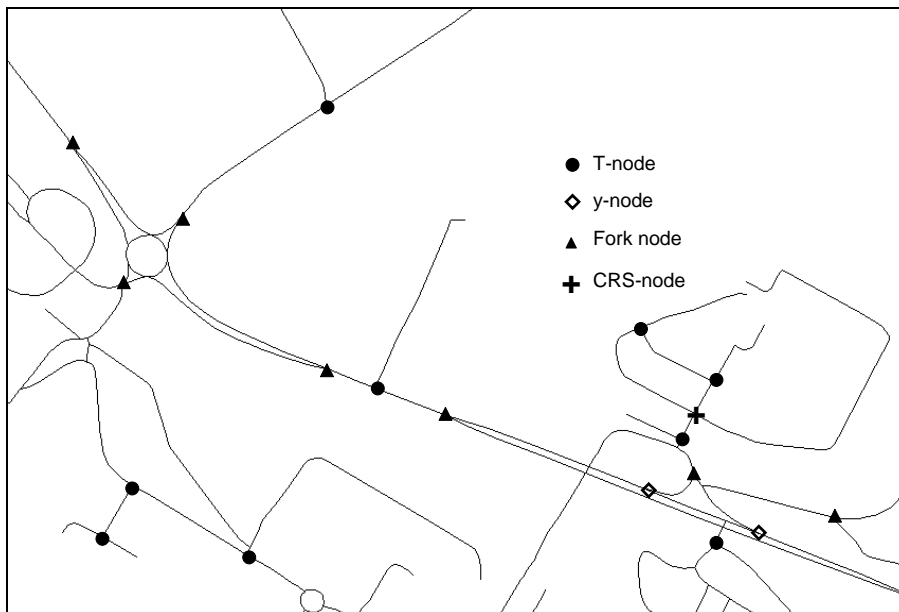

Figure 2. Results of the automatic detection of some simple crossroads.

The spatial analysis algorithms developed rely on a mixed use geometry and graph measures. The road network is seen as a graph thanks to GIS topology tools. T-nodes are topologically characterised by a degree 3 node and geometrically characterised by two nearly right angle α and a flat angle β between the arcs entering the node. With abstraction, it is characterised by a minor road leading to a major road. The tolerance thresholds used for α and β in the test case are 20° and 15°. As another example, *y*-nodes are characterised by a flat angle β (5° threshold), a slipping angle α (15° threshold) and an arc shape for the slip road. The Fork, CRS and Star nodes are detected and characterised with similar methods that are not developed here for briefness reasons. The Undifferentiated nodes are the remaining nodes.

Figure 2 shows the results of the automatic detection of T-nodes, y-nodes, fork-nodes and CRS nodes (no Star node in the picture and the Undifferentiated nodes are the unmarked ones) : all the nodes are iterated and if the degree and the angles correspond to one of the characteristic crossroads, the nodes are properly classified. This enrichment is particularly useful for the detection of more complex structures but also for the structure typification and rural selection processes.

3.1.2. Complex Junctions. Some widespread road patterns like roundabouts or branching crossroads can be seen as complex junction in relation to simple crossroads. With abstraction, they have the function of a crossroad and can be generalised as simple crossroads. In order to detect roundabouts, the faces of the mathematical graph are used rather than arcs and nodes (Sheeren *et al.*, 2004). To keep only the small round faces (that clearly correspond to roundabouts (Figure 2)), a measure of polygon compactness is used on all the small faces of the graph. Miller's measure of compactness (see definition in Figure 4b) is chosen with a threshold of 0.98 determined after experiments. The Miller's measure varies from 0 to 1 (for a circle).

Branching crossroads are a bit more complex to detect as Grosso (2004) distinguishes two types. The first type concerns the small triangular faces of the network; the second is related to the branching junctions attached to roundabouts (Figure 2, on the left). The detection algorithm consists in considering the faces with only three degree 3 nodes. Then, a surface distance is computed between the face and the triangle formed by the three nodes. If the distance is low, the face is considered as a branching crossroad. In the test data, the thresholds are 10000 m² for size and 0.45 for surface distance (that is between 0 et 1). Then when a potential face is branched to a roundabout, the surface distance threshold is bigger.

3.1.3. Road Strokes. The notion of strokes is based on one of the perceptual grouping principles enunciated by the Gestalt Psychologists (Wertheimer, 1938), namely the "good continuation" principle. Strokes are groups of roads gathered by continuous curvature (Figure 3a). The stroke creation algorithm described in Thomson and Richardson (1999) is used taking into account road class continuity: for each road section, the following sections continuity is analysed in relation to curvature and road class continuity. The best continuing section is added to the stroke if at least one provides a good continuity and the process iterates until all the sections of the network are part of a stroke, possibly being the only component of their stroke. In addition, the algorithm is adapted to manage previously detected complex junctions such as roundabouts or branching crossroads: the characterisation of such junctions allows to know the road segments leading to the junction.

When a stroke arrives at such junctions, continuity is only computed with the "outside" segments (Figure 3b) whereas the stroke would be stopped without the adaptation. As stroke length is a factor used in selection and as such junctions are very common in networks, strokes stopping at roundabouts would bias the selection and would correctly represent important routes as it intends to.
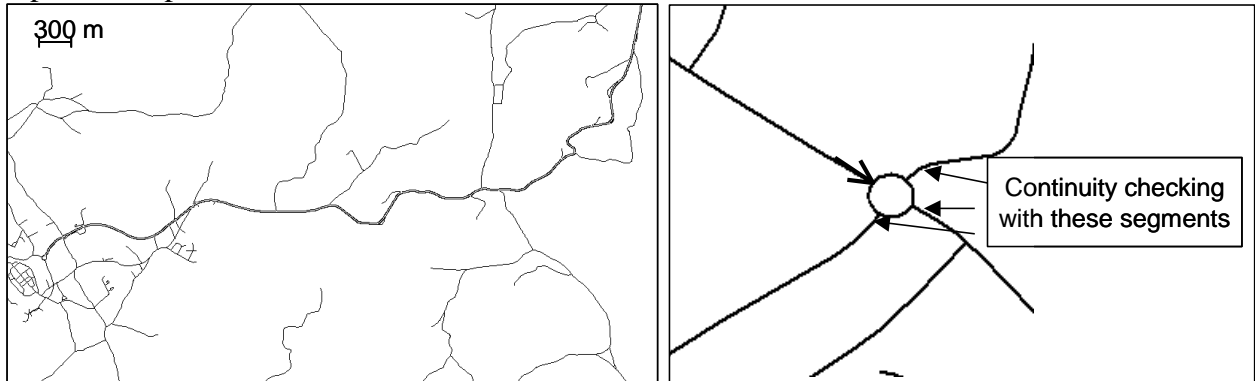

Figure 3. (a) A stroke in a road network: a group a roads following the "good continuation" principle. (b) strokes pass through roundabouts and branching crossroads.

3.1.4. Dual Carriageways. Considering the hypotheses (section 2.1), no traffic direction information is available in our data so the dual carriageways detection can not rely only on traffic direction as in Thom (2005). Thus, like in the roundabout detection method, the algorithm iterates through the network faces but this time, it seeks for very long and thin or narrow faces or small faces interlocked in thin faces. Indeed, looking for the parallel roads in data appeared to be ineffective. Road faces are classified by three measures, convexity (definition in Figure 4c), elongation (Figure 4a) and compactness (Figure 4b) that help to determine whether the face is thin (it belongs to a dual carriageway) or not. When the polygon is convex, the elongation measure is a correct assessment of the thinness of the face and then is used. But when the face is concave, this elongation measure can no longer be used because turning dual carriageways would not be detected. Thus, a compactness measure is used coupled with an area threshold as long and thin faces are not compact because the perimeter is very big compared to area. After long experiments, thresholds are set so that a face might belongs to a dual carriageway if:

- convexity > 0.8 and elongation > 5
- or convexity < 0.8 and compactness < 0.2 and size < 50000 m²
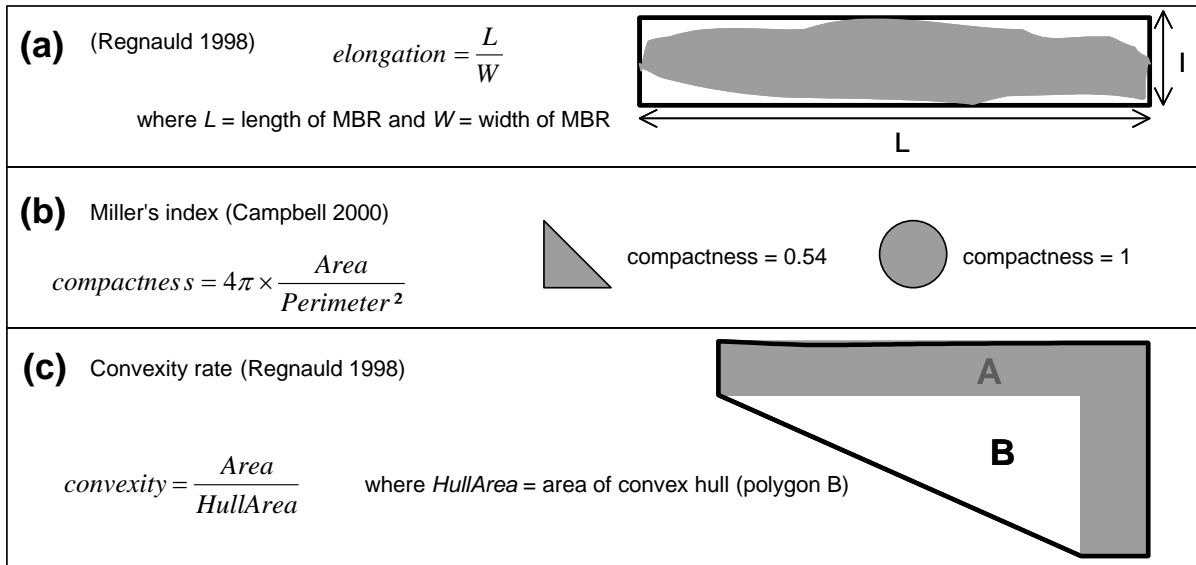- or size <10000 m² and has 2 part of dual carriageways for neighbours

Figure 4. The 3 measures used to assess the thinness of polygons: (a) elongation. (b) compactness by Miller's index. (c) convexity.

3.1.5. Highway Interchange. Highway interchanges are a significant example of the road structures that need a typification process to be properly generalised. Grosso (2004) noticed that highway interchanges are characterised by a high density in nodes that correspond to fork and y-nodes in our classification. The detection algorithm consists in a clustering by spread search in the road graph on the fork and y-nodes. Then, all roads inside the buffered convex hull of the cluster belong to the highway interchange (Figure 5). Over-detection is managed excluding the clusters whose extent contains a significant amount of buildings. The cluster search threshold was determined considering the usual length of slip roads and set to 400 m. If there is no building data available, measures can be used only on roads like the amount of dead ends, the presence of dual carriageways and the belonging of nodes to roundabouts.
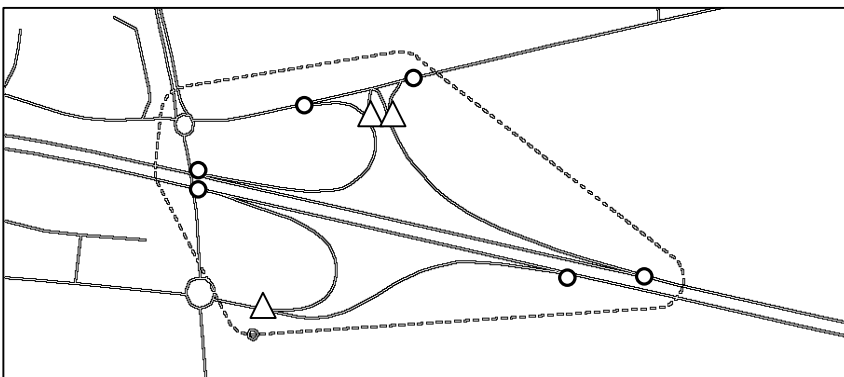

Figure 5. Highway interchange detected by clustering the fork (triangles) and y-nodes (circles).

3.1.6. Rest Areas. Rest areas are also important to detect. The properties of the structure are the presence of an entrance and an exit with roads between located on both sides of dual carriageways or other major road. The detection algorithm is composed of two main steps: the detection and grouping of entrances and exits, and the addition of in-between roads.

Entrances and exits are detected with *y*-nodes and their orientation (Figure 6a). The *y*-nodes belonging to an interchange are excluded and couples (entrance, exit) are formed considering the highway direction. Then, we switch once again to faces and a buffer is used on the good side (considering the direction) of the major road in which the small neighbouring faces are aggregated (Figure 6b). According to our experiments, a buffer size of 500 m and an area threshold for the faces to be included in the rest area of 50000 m² are effective thresholds.
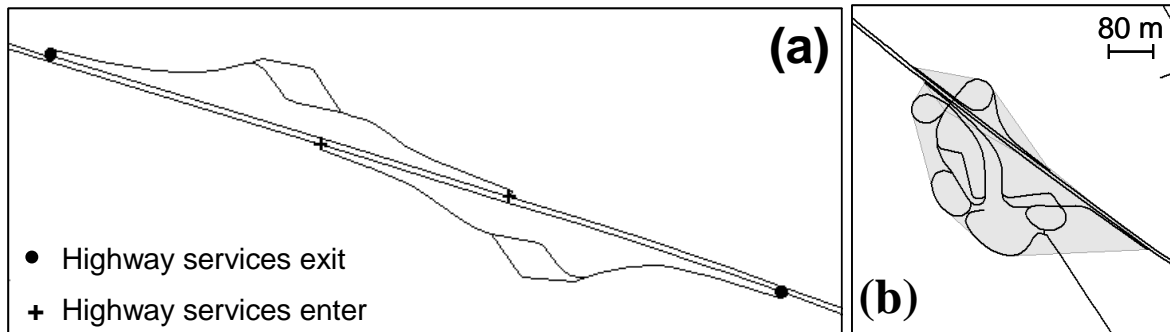

Figure 6. (a) Detection of enters and exits of rest areas. (b) detection of the extent of the rest area.

3.1.7. Dead Ends. Dead end roads are another road structure that play a key role in selection processes because they are either useless (leading to nowhere important) or very important (leading to a significant facility). Unlike simple dead ends, the detection of dead end groups is not obvious. It uses the notion of minimal graph cycles: the roads that do not belong to a minimal cycle or that belong to a cycle disconnected from the graph (case with a roundabout at the end of the dead end) are considered as dead ends and are then grouped by connectivity. In the test data, optional facility data is available so dead ends are enriched with access to the facility when exists (when the nearest access to the network of a facility point is a dead end).

3.2. Typification of particular structures

Pattern typification is considered as the final step of the selection process (Figure 1) but it is presented here as it is completely linked to the structures recognition. A typification operator has been defined by McMaster and Shea (1992) like this : 'A selective number and pattern of the symbols are depicted. Generally, this is accomplished by leaving out the smallest features, or those which add little to the general impression of the distribution (...) or using a representative pattern of the symbols...'. This step concerns a large part of the road structures presented in section 3.1. The road segments belonging to a structure needing typification are ignored by the different selection steps (rural, urban and continuity checking) in order to be finally typified. As a consequence, a parameter of the overall process is the choice of the important road patterns that have to be typified rather than to be simply selected or eliminated by the appropriate process. Two implemented methods to typify structures are presented in this section.
First, a simple typification of roundabouts and branching crossroads relies on their characterisation with an enriched data schema. The roles of the road segment in the complex junction are identified: the road segments leading to a roundabout or belonging to

it are distinguished. Then the last ones are collapsed into one point where the first ones are extended. Branching crossroads are typified into a T-node thanks to the enriched schema that identifies the minor and major roads of the abstracted T-shaped junction. A size parameter is added to decide whether the junction has to be typified or not. In the test case presented in the results section 5, if the diameter of a roundabout is under 100m, it is typified.

Dual carriageways can be typified with the collapse of the structure into its centreline. The centreline can be obtained with algorithm computing skeletons of surfaces. Thom's (2005) method using Delaunay triangulation was implemented.

4. The Road Network Selection Process

The previous section presents a prerequisite step for the selection process described in this section as all the structures and patterns recognised in the road network enable a correct and relevant selection. As exposed in Figure 1, two different selection processes were developed, one for rural areas, one for urban areas. Added to that, a continuity checking process at the interface of the areas is included to recombine the selections.

4.1. Road selection in rural areas

Rural selection is based on determining roads saliency by shortest paths between attraction points following the method of Richardson and Thomson (1996). The aim of the method is to estimate road use by cars, so shortest paths are computed in the network between nodes where cars are supposed to start (residential areas or airports) and nodes where cars are supposed to arrive (commercial or work areas). The attraction points result from the optional facility data weighted by importance and attached to a node of the network (Figure 7). If facility points are not available, random points can be used (Ruas and Morisset, 1997). Shortest paths are computed in the complete road graph including roads in urban areas to be really meaningful. Moreover, road use estimation is considered as an enrichment of road segments and it will be useful for urban selection process as input described in section 4.2.
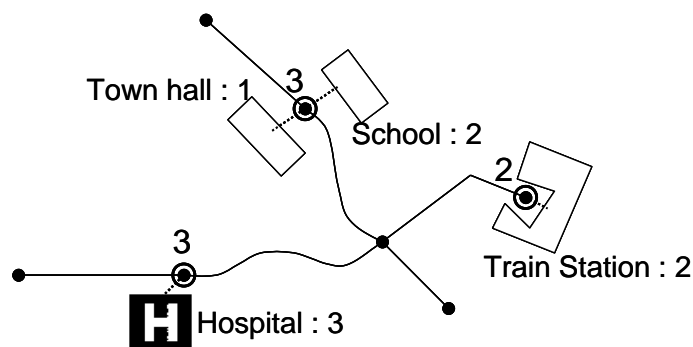

Figure 7. How weighted attraction points are attached to the road network.

Instead of using simply road length to compute shortest paths, a travel time estimation is used. The travel time is estimated with a weight function (Figure 8) taking into account attribute values or geometric characteristics : road length is multiplied by factors smaller than 1 for quickening characteristics and greater than 1 otherwise. Road class (main roads

are designed to go faster), road sinuosity (the geometric measure of sinuosity is taken from Mustière (2001)), town crossing, are used to compute factors by default. When such attribute data can be available (like in our test data), road width (or number of ways) or altitude difference can be added to the cost function. The factors were determined empirically in relation to the speed limitation they correspond to and validated by testing. For instance, major road class corresponds to a 1 factor as it is considered as the default speed. A minor road class corresponds to a 1.2 factor and high sinuosity corresponds to a 1.35 factor.

Length : 3525 m
Sinuosity : high
Road class : minor
Weight : 5790

Length : 3804 m
Sinuosity : low
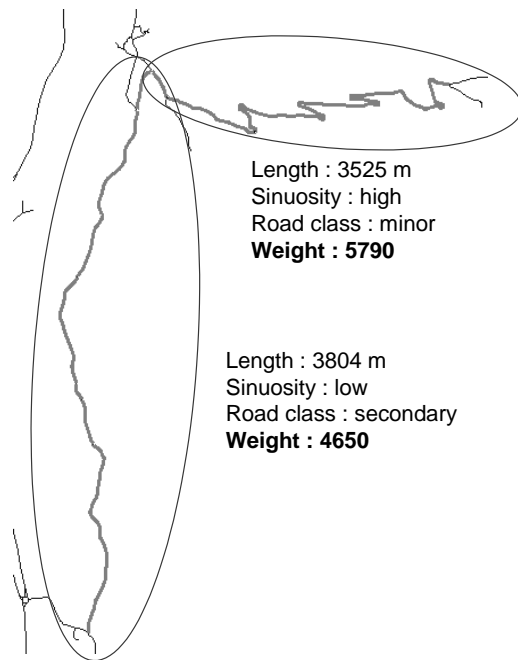Road class : secondary
Weight : 4650

Figure 8. Example of cost function estimating road travel time. The two roads have comparable lengths but the road on the right has a much bigger weight because of its sinuosity, its minor road class and its altitude difference (it is a mountain road).

Only the shortest paths between an attraction point and the ones in a radius of 50 km are computed, using the algorithm of Dijkstra (1959), because we consider that road trips over this distance will massively use highways that are already considered as important. Like in Richardson and Thomson (1996), after the computation of all the shortest paths, roads used by shortest paths have their road use estimation value incremented to assess their importance in relation to initial and final attraction point weights. In Figure 9a, there is a graph with 5 attraction points (AP) (a to e) weighted from 1 to 3 according to their importance. In Figure 9b the shortest path from AP "c" to AP "d" is computed and the 2 roads used by the shortest paths are assigned a value equal to final AP weight (here 3). In Figure 9c, all shortest paths from AP "c" to the others are computed and road values are incremented as soon as they are used by a shortest path. For example, the road between AP "c" and the centre simple node is used by several shortest paths and its value is incremented to 7 (3 from c to d, 2 from c to a and 2 from c to e). Figure 9d shows the values of all roads after all shortest paths have been computed.
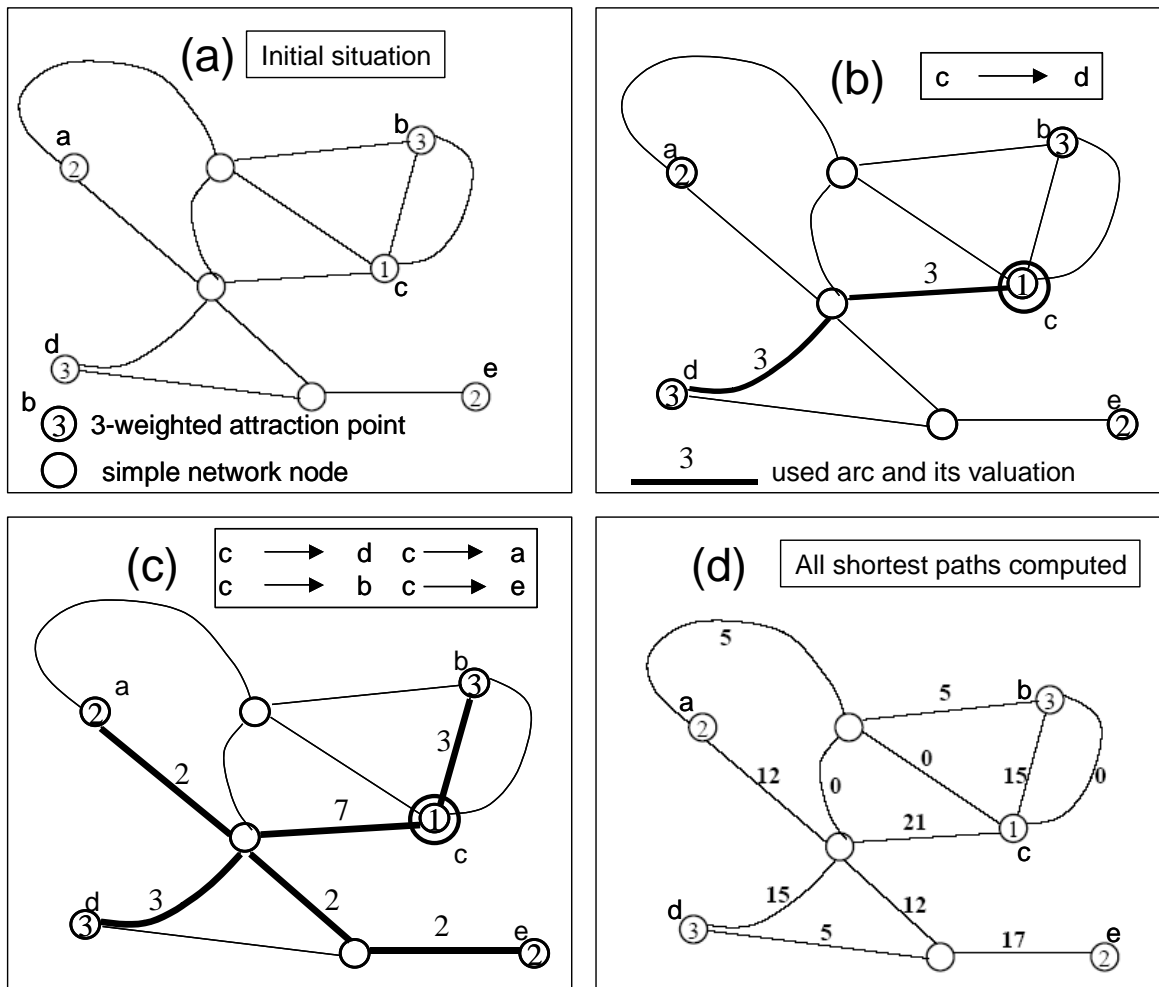
Figure 9. Valuing of roads in shortest path computation. See description of the process in the text.

Finally, rural selection is processed using different parameters that correspond to thresholds. A road is selected if :

− estimated road use is bigger than a given threshold (very used road segments are important).
− or stroke (it belongs) length is bigger than a given threshold (long strokes are important).
− or stroke (it belongs) crosses a number of T-nodes bigger than a given threshold (it is important because it is served by many minor roads).
− or it belongs to a dead end longer than a given threshold (long dead ends are significant landmarks in a map).
− or it belongs to a dead end with facility access longer than a given threshold lower than the threshold just above (the access to the facility has to be maintained).

The choice of the selection criteria is partly application-related. Other selection criteria could be chosen when it is necessary to meet particular specifications. For example, some specifications may require that roads with particular attribute values have to be always selected.

4.2. Street selection in urban areas

As street selection is a more complex problem than rural road selection, different approaches have been tried: (Ruas, 1999, Edwardes, 2000, Jiang and Claramunt, 2004). The approach presented in this paper is a kind of synthesis of these three ones based on the road block aggregation principle of Ruas (1999): as for roundabouts for instance, the faces of the graph are considered rather than the arcs. The face aggregation principle is quite similar to area aggregation methods (van Oosterom, 1995) : the smallest area is aggregated to the neighbour area that minimises a cost function. In the case of road faces, when two faces are aggregated, the road that separates them is eliminated and the remaining roads at the end of the process are selected. Ruas (1999) aimed at cartographic generalisation so buildings density was the key factor for aggregation. Hence, the principle was kept but factors for aggregation and dynamics were changed.

4.2.1. Urban network enriched model. In our process, data has to be enriched in urban areas (here called "towns" for the sake of simplicity (Figure 10)). Towns are characterised by a size ("big" or "small") and attribute statistics: strokes mean length in town, road use estimation mean for the streets and streets density. Two parameters are also assigned to the town: the maximum cost of aggregation (see section 4.2.3) and the maximum area of the faces. The parameters assigned can be different if the town is either big or small if we want big towns to be more simplified than small ones. Indeed, we have noticed that such requirements are common in generalised database specifications.
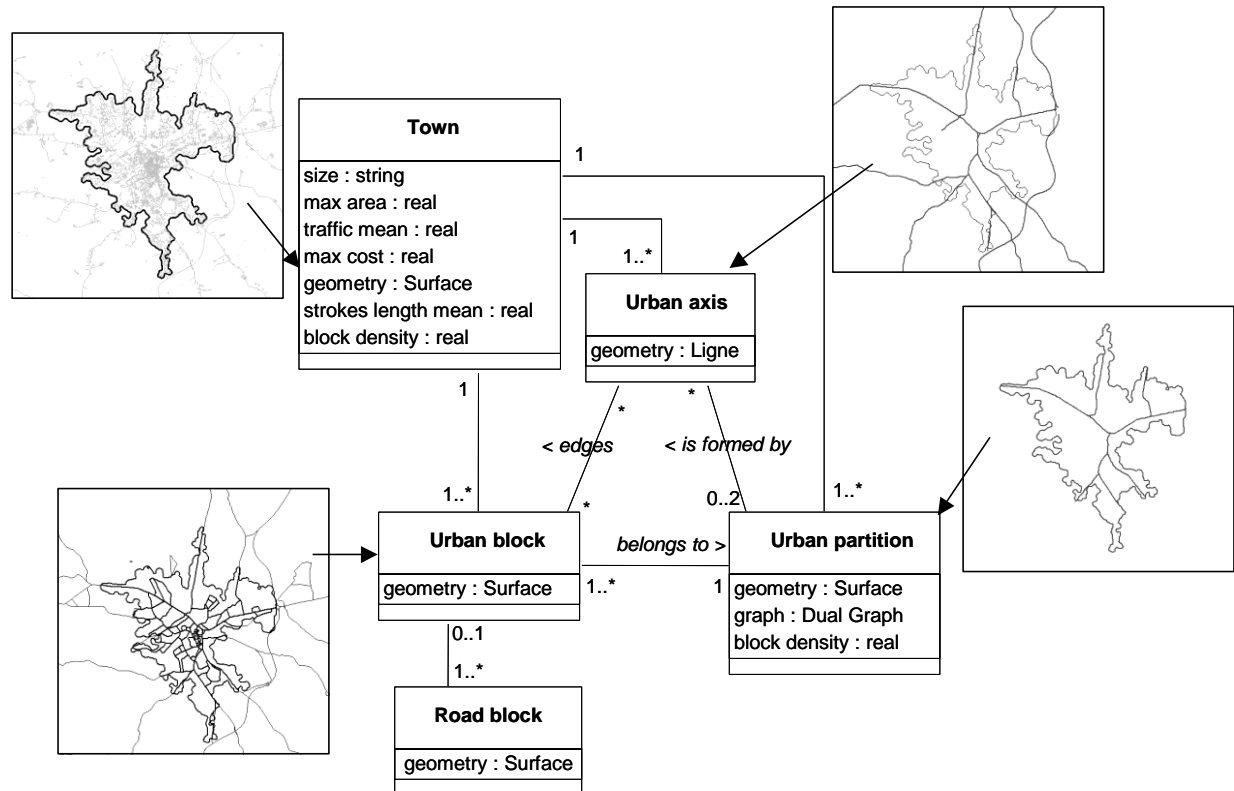

Figure 10. Enriched data model of an urban area for street selection, illustrated with data samples from the town of Salies-de-Béarn in the southwest of France.

Then, three new classes are added and populated. First, the road graph faces cut up by town limits are called "urban blocks". The urban blocks are the features that will be aggregated during the urban selection process.

Then main roads passing inside the town, determined with stroke length, road class and road use estimation (result of rural selection), are classified as urban axes. The thresholds used are more restrictive than the ones for rural selection, which selects only the very important streets. It is both a way to be sure that these streets will not be eliminated in the process and to structure the town with its main axes.

Finally, towns are partitioned by the urban axes to create "urban partitions". Like towns, partitions are characterised by a measure about the density of streets. It allows to take into account and maintain in the process road density differences like in town centre (many small roads) compared to suburbs (few longer roads).

4.2.2. Dynamics of aggregation. The aggregation algorithm is based on an aggregation cost. For each small block that needs to be aggregated, a cost is computed for each neighbour of the same partition (two blocks of different partitions can't be aggregated). The candidate neighbour with the lowest aggregation cost is chosen and aggregated if the cost is lower than a maximum cost. When there is no block small enough left or when all aggregation costs are above the maximum cost, the algorithm stops (Figure 11).
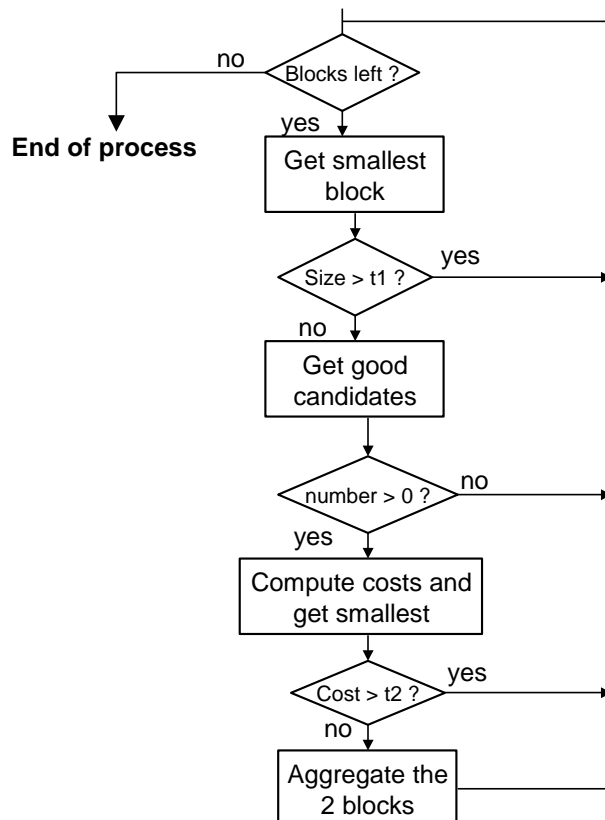

Figure 11. Diagram of the dynamics of the aggregation algorithm.

We defined the cost function formula below to compute the aggregation cost.

$$\cos t = (1 - compactness)^3 \times \sqrt{centralities\_mean} \times \frac{traffic\_estimation}{traffic\_estimation\_mean} \times$$

$$\frac{stroke\_length}{stroke\_length\_mean} \times \sqrt{\frac{partition\_density}{town\_density}} \times size\_factor$$

The main factor is compactness (Figure 4b) as compact blocks are a good final result of the algorithm. The more appropriate the aggregation is, the lower the cost is. All factors used are normalised in order to be multiplied. If a factor is higher than 1 (for example the road use on the separating street is higher than the road use mean for town streets), the cost increases to make the aggregation more difficult. Apart from compactness, the factors used in the cost function are:

- the graph centralities of separating streets as proposed Jiang and Claramunt (2004): centrality measures the importance of nodes in a graph and to measure the centrality of streets, they proposed to build a dual graph of the network where streets are the nodes connected by arcs to the nodes that represent intersecting streets. Thus, central separating streets are harder to eliminate by aggregation.
- the traffic estimation of separating streets with the same value as the one computed for rural selection (but that was computed on all roads). Thus, streets with a traffic estimation higher than the town mean are harder to eliminate by aggregation.
- the length of the strokes that contain the separating streets. Thus, streets that belong to strokes longer than the town mean are harder to eliminate by aggregation.
- the partition block density as aggregation is made easier in town partitions that were not initially very dense in order to maintain the differences between town centre and suburbs.
- the aggregated block size as aggregation is made easier when it generates small aggregated blocks.

Once the factors were defined, the formula was adapted after testing by weighting more or less some factors to give them more or less importance in the cost. For instance, we believe that compactness is the key factor and should have the main influence in the cost, that is why the factor is powered by three while the centrality factor is square rooted.

Figure 12 shows an example of aggregation : (a) block 1 is candidate for aggregation and has four neighbours. (b) Blocks 4 et 5 are eliminated because separated from 1 by important roads that belong to long and central strokes and cost is higher than maximum. (c) shows that the compactness of the aggregate 1-3 is quite low which increases the cost of aggregation. (d) shows that the aggregate 1-2 is very compact and as a consequence has a lower aggregation cost. Finally, block 1 is aggregated to block 2.

After the aggregation, dead end selection in urban areas is managed the same way as in rural areas as it cannot be managed by a block aggregation process. Moreover, the dead ends linked to the network by a street that was eliminated are also eliminated whatever its length is to maintain network continuity. The two parameters of the algorithm are the minimum size of a block and the maximum aggregation cost that is related to the cost function formula.
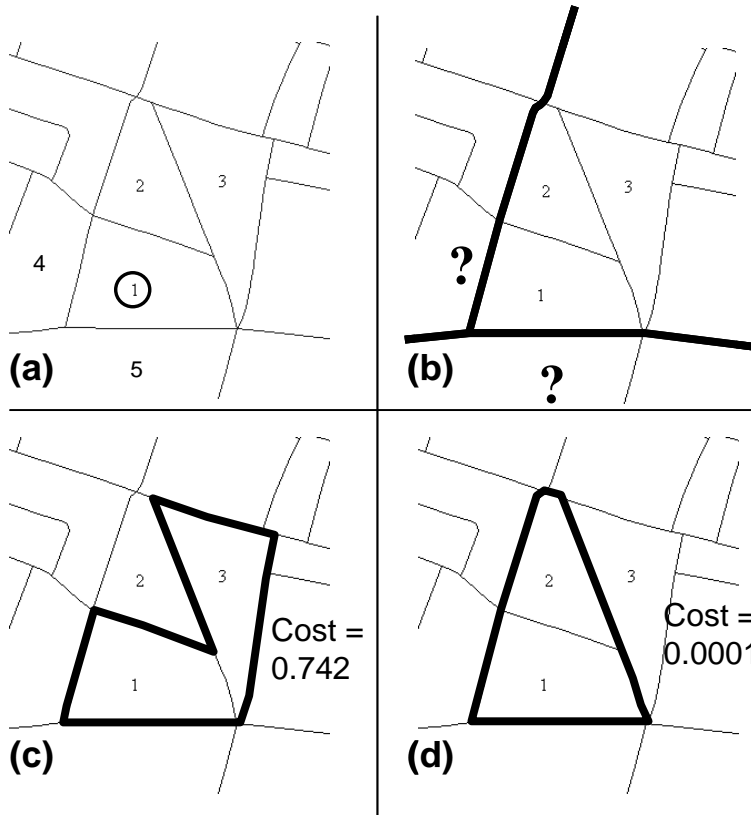
Figure 12. Which neighbour block can be aggregated to block 1?

4.3. Continuity checking

The presented method for road network selection uses two different kind of processes for rural and urban areas. Using them independently would cause discontinuities and inconsistencies. In order to avoid such problems, a continuity checking step is necessary and is decomposed in two steps : first, continuity needs to be controlled at urban/rural interface zones; then, the connectivity within the complete selected network has to be checked and corrected if necessary.

The idea is that a road partly selected in one of the processes should have its other part selected. If it has not been selected by the other process, it is restored during the continuity checking step. Strokes seem to be the ideal way to allow this (Figure 13). In this example, the road is only selected in urban selection and continuity checking allows to restore the part of the stroke that was not selected during rural selection providing a more logical selected network.

Furthermore, to avoid discontinuities that are internal to one of the areas (it is most likely to appear in rural areas), an additional continuity checking post-process is carried out: groups of connected road segments that are not connected to the rest of the network are detected and the roads used by the shortest way to connect them are added to selection. The algorithm used here is similar to the one developed by Chaudhry and Mackaness (2005) : if discontinuities are detected, the unselected strokes are tested by length ascending order until the addition of one of them re-connects the network. The re-connecting stroke is added to selection.
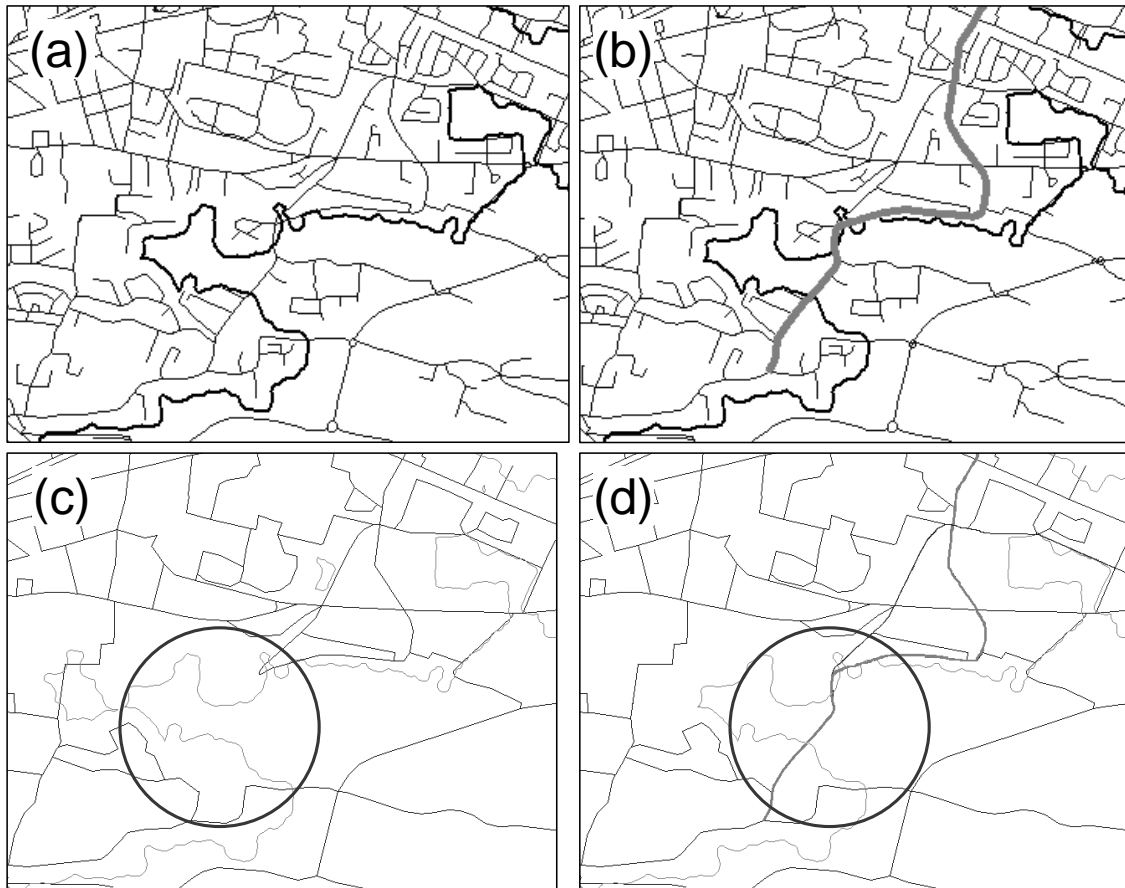
Figure 13. (a) An initial network at the limit (in bold) of a town. (b) A big stroke that crosses both urban and rural areas. (c) The use of only the two independent processes creates a problem of continuity highlighted by the circle. (d) The continuity problem is solved by the use of the stroke highlighted in (b).

5.   Results and discussion

As mentioned in section 2, the test data is BDTOPO® (1m resolution topographic database) from IGN France. The main test case is the generalisation of BDCARTO® (10m resolution database) from the test data. The process has been applied with selection parameters corresponding to BDCARTO® specifications. But the process has been tested with other parameters. The complete process has been implemented on Clarity™ GIS and the results presented here as well as the ones presented in the previous sections, are all obtained automatically.

Figure 14 shows a result of the global process presented in the section 2 with the test case of BDCARTO® on the town of Salies-de-Béarn and its surrounding area. This result is good in relation to the specifications and is totally equivalent to actual BDCARTO®.

Figure 15 focuses on street selections obtained on a larger city with two sets of parameters. The first case (b) consist in a light selection of the network while the second one is a stricter selection that approximately correspond to a 10m resolution database. In both, main streets still appear and the density difference between the city centre and the suburbs is visible.

Such results are shown before the pattern typification step so roundabouts and dual carriageways, visible in the pictures, would disappear in the final results.

Figure 14. A result of the global process on the town of Salies-de-Béarn and its surrounding area.

The difficulty to evaluate such a process is that the quality of a selection is hardly quantifiable because of most selection specifications fuzziness. Even when the result is compared to a reference dataset, it is hard to say if the differences are due to an imperfect process, uncertainties in the specifications translation or errors in the reference dataset (up-to-dateness for example). Nevertheless, the results were compared in our test case with BDCARTO®. Rural selection selects 95% of BDCARTO® roads (in terms of road length) and urban selection provides a road length up to 98% of BDCARTO® road length but not exactly the same road segments (75%, in length, of the roads selected are identical). Most differences are due to specification fuzziness and up-to-dateness differences.

In terms of efficiency (how quickly the result is obtained), it is a quite heavy process with a lot of enrichment and process time consuming algorithms like shortest path and centrality computing, but database generalisation is not supposed to be 'on-the-fly'.
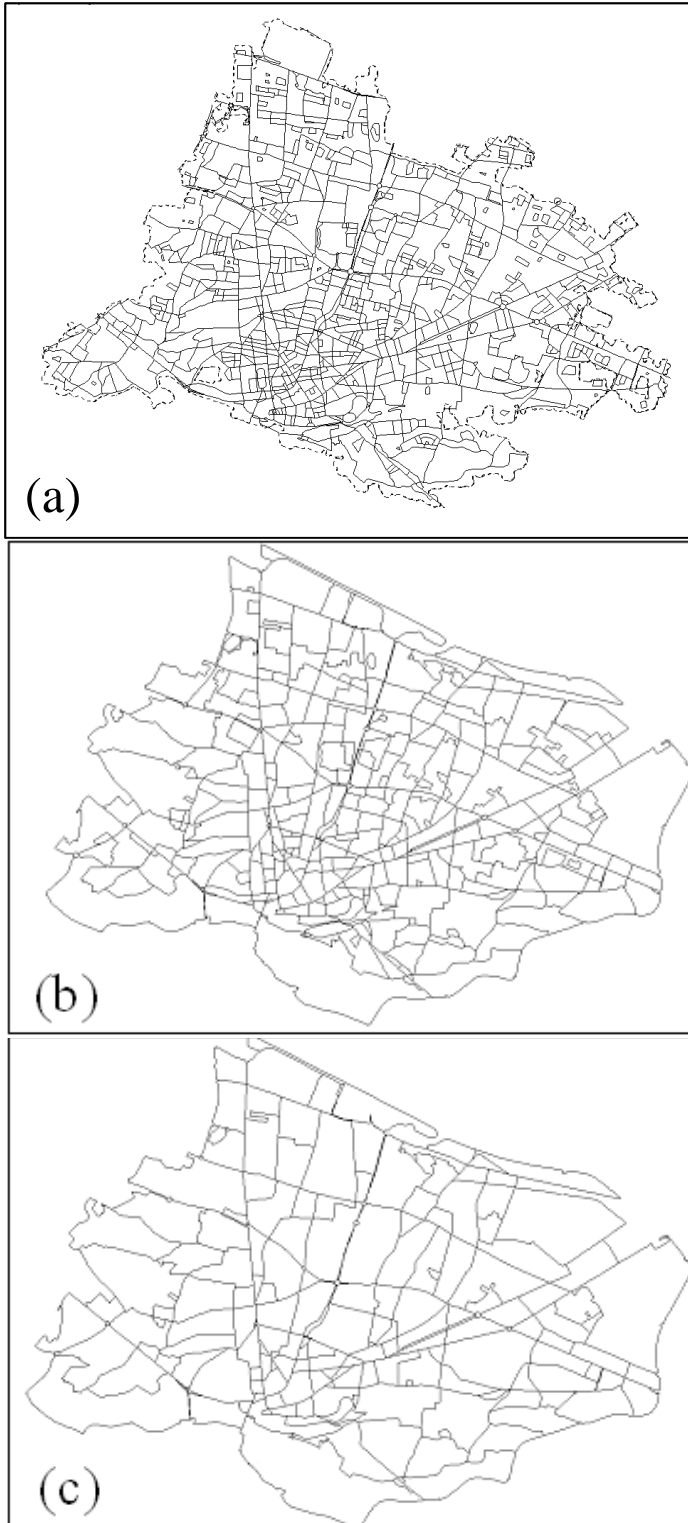
Figure 15. (a) The street network of the city of Pau (southwest of France) in BDTOPO®. (b) a light selection of this street network. (c) a stronger selection corresponding to 10 m resolution database.

The aim of this work was to design a generic process able to give results with different target specifications but also with different initial data. Nevertheless, the presented process is slightly dependant on the test data used to develop it. So, to be used with different datasets, the process would require little adjustments especially for shortest path computation where additional data is used for node and link weighting (point facilities data, road attributes like width or elevation).

A difficult part of the process design is the choice of the multiple parameters and thresholds. Indeed, the translation of the target database specification may not always be obvious and the process requires a significant number of parameters. We decided to develop sets of default parameters to make the choice easier. Tests were carried out on the sensibility of a major part of the parameters, particularly the thresholds for selection like 'selection if stroke length > 3000 km'. The tests showed that the process was not very sensible to the variation of such parameters. Then the parameters can be chosen by approximation without risking to change drastically the results.

6. Conclusion

To conclude, this paper presents a complete process to allow road network selection in model generalisation. It consists in (1) enriching the data by structures recognition, (2) rural selection based on assessing traffic by shortest paths, (3) street selection algorithm based on road block aggregation and (4) structures typification. The contribution is mainly the design of this generic process and the development of the urban selection method. There are also contributions in the structure detection algorithms. Tests carried out on large datasets show encouraging results for different kind of generalisation specifications.

To go further, pattern typification has to be improved by implementing some algorithms specific to each structure. Added to that, it would be interesting to integrate new particular road structures like circular roads or star-shaped structures (Heinzle and Anders 2007). Their detection could be used to enhance the selection algorithms. The idea exposed in Marshall (2005) that road patterns are like an unfolding fractal shows that many more structures can be detected in a road network to add knowledge and improve the selection.

As a further work, it would also be interesting to work on expressing the specifications of a selection to parameterise automatically the process, like in Hubert and Ruas (2003) that used generalised data samples to converge to the user needs. Finally, it could be interesting to insert this process into a global cartographic generalisation process that would deal with selection before the cartographic operations.

References

Boffet A 2000 Creating urban information for cartographic generalisation. In *Proceedings of 10th International Symposium on Spatial Data Handling*, China, Beijing.

Campbell J 2000 *Map Use and Analysis*. Dubuque : McGraw-Hill, 4th edition.

Chaudhry O, Mackaness W Rural and Urban Road Network Generalization Deriving 1:250,000 from OS MasterMap. In *Proceedings of 22nd International Cartographic Conference of ICA*, La Coruña, Spain.

Dijkstra E W 1959 A note on two problems in connexion with graphs. *Numerische Mathematik*, 1: 269-271.

Edwardes A J, Mackaness W A 2000, Intelligent Generalization of Urban Road Networks. In *Proceedings of GIS Research UK Conference'2000*, York, UK: 81-85.

Grosso E 2004 Etude des carrefours d'un réseau routier – Première approche : modélisation et implémentation d'outils de détection des carrefours. Master thesis. COGIT Laboratory.

Heinzle F, Anders K-H, Sester M 2005 Graph Based Approaches for Recognition of Patterns and Implicit Information in Road Networks. In *Proceedings of 22nd International Cartographic Conference of ICA*, La Coruña, Spain.

Heinzle F, Anders K-H 2007 Characterising Space via Pattern Recognition Techniques: Identifying Patterns in Road Networks. In Mackaness W, Ruas A, Sarjakoski T (eds) *The Generalisation of Geographic Information : Models and Applications*. Elsevier: 233-254.

Hubert F, Ruas A 2003 A method based on samples to capture user needs for generalisation. In *Proceedings of 5th Workshop on Progress in Automated Map Generalization*, Paris, France.

Jiang B, Claramunt C 2004 A Structural Approach to the Model Generalization of an Urban Street Network. *GeoInformatica*, 8: 157-171.

Kilpeläinen T 1997 Multiple representation and generalisation of geo-databases for topographic maps. Publications of the Finnish Geodetic Institute.

Mackaness W 2007 Understanding Geographic Space. In Mackaness W, Ruas A, Sarjakoski T (eds) *The Generalisation of Geographic Information : Models and Applications*. Elsevier: 1-10.

Mackaness W, Edwards G 2002 The Importance of Modelling Pattern and Structure in Automated Map Generalisation. In *Proceedings of Joint Workshop on Multi-scale Representations of Spatial Data*, Ottawa.

McMaster R B, Shea K S 1992 *Generalization in Digital Cartography*. Association of American Cartographers, Washington D.C.

Marshall S 2005 *Streets & Patterns*. Spon Press, Taylor & Francis Group, New York.

Mustiere S 2001 Apprentissage supervisé pour la généralisation cartographique. PhD Thesis, University Paris 6.

Regnauld N 1998 Généralisation du Bâti : structure spatiale de type graphe et représentation cartographique. PhD Thesis University Aix-Marseille 1.

Richardson D E, Thomson R C 1996 Integrating Thematic, Geometric and Topologic Information in the Generalization of Road Networks. *Cartographica* , 33: 75-83

Ruas A, Morisset B 1997 Simulation and Agent Modelling for road selection in generalisation. In *Proceedings of 18th International Cartographic Conference of ICA*. Sweden , Stockholm: 1376-1380.

Ruas A 1999 Modèle de généralisation de données géographiques à base de contraintes et d'autonomie. PhD Thesis, University Marne la Vallée.

Ruas A 2004 Le changement de niveau de détail dans la représentation de l'information géographique. Habilitation à diriger des recherches, Université de Marne-la-Vallée.

Sester M 1995 Lernen struktureller Modelle für die Bildanalyse. PhD Thesis, Institute for Photogrammetry, University of Stuttgart, Deutsche Geodätische Kommission, München, C (441).

Sheeren D, Mustiere S, Zucker J-D 2004 How to Integrate Heterogeneous Spatial Databases in a Constistent Way? In *Proceedings of Conference on Advanced Databases and Information Systems (ADBIS)*. Budapest, Benezur A, Demetrovics J and Gottlob G (eds) Springer: 364-378

Thom S 2005 A Strategy for Collapsing OS Integrated Transport Network Dual Carriageways. In *Proceedings of 8th ICA Workshop on Generalization and Multiple Representation*. La Coruña.

Thomson R C, Richardson D 1999 The "Good Continuation" principle of Perceptual Organization applied to the Generalization of Road Networks. In *Proceedings of 19th International Cartographic Conference*. ICA, Ottawa, Canada.

van Oosterom P 1995 The GAP-tree, an approach to 'on the fly' map generalization of an area partitioning. In J C Muller, J P Lagrange and R Weibel (eds) *GIS and Generalization: Methodology and Practise*. Taylor & Francis, London, UK: 120-132

Weibel, R., Dutton, G., 1999, Generalising Spatial Data and Dealing with Multiple Representations. In Longley P, Goodchild M F, Maguire D J and Rhind D W (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications, Second Edition.* Chichester: 125-155

Wertheimer, M. 1938. Laws of organization in perceptual forms. In Ellis W (Ed) *A Source Book of Gestalt Psychology*. 71-88