HAL open science

Multi-Criteria Geographic Analysis for Automated Cartographic Generalization

Guillaume Touya

▶ **To cite this version:**

Guillaume Touya. Multi-Criteria Geographic Analysis for Automated Cartographic Generalization. The Cartographic Journal, 2022, 59 (1), pp.18-34. 10.1080/00087041.2020.1858608 . hal-03321351

HAL Id: hal-03321351

https://hal.science/hal-03321351

Submitted on 17 Aug 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MULTI-CRITERIA GEOGRAPHIC ANALYSIS FOR AUTOMATED CARTOGRAPHIC GENERALISATION

A PREPRINT

Guillaume Touya
LASTIG, Univ Gustave Eiffel, IGN, ENSG
Saint-Mandé, F-94160
guillaume.touya@ign.fr

August 17, 2021

ABSTRACT

Cartographic generalisation is a process similar to text summarization that transforms a map when scale is reduced. Cartographic generalisation simplifies the map content while preserving as much as possible the initial characteristics and spatial relations of the map. The automation of this process requires a deep understanding of the context of each map feature, which involves different criteria such as the shape of the feature, the semantic nature of the feature, or the spatial patterns of its neighbouring features. This is why multiple criteria decision techniques can be relevant make multiple criteria decision during the process. This paper proposes two use cases of cartographic generalisation sub-tasks where multiple criteria decision techniques improve past techniques: the classification of urban building blocks and the ordering of the buildings to keep in the map as a priority. In both use cases, the proposed methods are experimented on large real cartographic datasets, and evaluated in comparison to alternative techniques.

***K*eywords** Multiple criteria analysis · Cartography · Cartographic generalisation · ELECTRE · PROMETHEE · spatial analysis

1 Introduction

Topographic maps are created from geographic databases in which topographic features such as buildings, roads, or rivers are captured with details. At large scales (*e.g.* 1:5,000, or 1:10,000), all these map features can be displayed with legible symbols in a map, but when scale decreases, legible symbols use more ground space to be displayed, and all the features cannot be displayed anymore. So, to make maps at small scales, cartographers select the prominent features, simplify, abstract them to keep a legible map. This process, close to text summarization, is called map generalisation, and just like text summarization, its automation is a complex issue that has been tackled by researchers in cartography for decades [1]. Thanks to the advances of research, automated map generalisation is being used more and more by administration and companies that produce large series of topographic maps, such as national or regional mapping agencies [2].

Automated map generalisation generally requires an enrichment step that makes explicit in the database the implicit spatial relations (*e.g.* the building is in the corner of a block, so an important landmark) or complex features (*e.g.* aligned buildings) that help a human cartographer to grasp the geography of a place [3, 4]. In this complex enrichment process that mimics how people see and understand the geographic space, there is a need for methods to help decision making while considering geometrical properties, semantics, and spatial relations altogether, at several scales. Multiple criteria decision techniques have already been successfully used in geographic information science (*e.g.* [5, 6]) and we believe that map generalisation could benefit from such techniques during this enrichment step, before the transformation algorithm are applied. This paper explores the application of multiple criteria decision techniques from the literature to map generalisation.

The second section of this paper describes related work on multiple criteria analysis for automated cartography and map generalisation. The third section presents the first use case on the classification of urban blocks. The second use case on the ordering of buildings to keep in the map is presented in Section 4. Finally, the conclusion discusses the contributions of the paper and describes future research.

2 Multi-Criteria Spatial Analysis for Automated Cartography

2.1 Related Work

Due to the intrinsic multi-criteria nature of automated cartography, the early methods that were developed basically combine geometry, spatial relations and semantics with simple rules or extremely complex cost functions. For instance, the function that the cost of eliminating a road from the map multiplies values of area and compactness of the adjacent faces of the graph, of the estimated traffic on the road, and centrality of the road in the graph [7]. But such criteria quickly become difficult to combine and more advanced techniques were searched. Thus, there are several examples where multiple criteria analysis was used to combine geometry, spatial relations and semantics:

- spatial relation based positioning with PROMETHEE [5] (*e.g.* "the accident happened in the first turn after a given crossroad, and in front of a building");
- automated text placement in a map that depends on text importance, and the placement of the other texts around [8];
- matching features that are differently represented in two different geographical datasets but correspond to the same real world feature, with the theory of belief functions [9];
- inferring the the level of detail of OpenStreetMap data from geometry, spatial relations, semantics and meta-data with ELECTRE [6, 10].

In map generalisation, such methods were previously used to know when to improve the heuristics that guide the automated agent-based processes [11].

In addition to this past success in similar problem, the multiple criteria decision techniques are really appealing for automated cartography and map generalisation for several reasons:

- the criteria are generally hardly comparable;
- there are clear bad decisions, but never only one obvious best solution;
- the number of criteria is quite limited most of the time;
- expert cartographers rarely agree on what exactly is a good generalisation.

In particular, multiple criteria decision techniques can be intuitively preferred over machine learning for a simple reason that comes out of the previous list: it is really complex to create training datasets (they don't exist and examples are not easy to annotate even for an expert cartographer, and extremely complex for non experts). And if we focus on the very popular deep learning techniques, the purely graphical/geometrical criteria could be properly learned by such techniques, as proved by recent research on building simplification [12], but when graphical criteria have to be mixed with semantics or spatial relations with the geographical neigbourhood, current deep learning techniques do not seem particularly adapted.

The remainder of the paper describes two use cases where multiple criteria decision techniques (ELECTRE and PROMETHEE) were applied to improve sub-tasks of a generalisation process.

2.2 ELECTRE methods

The ELECTRE methods are multiple criteria decision techniques with a partial aggregation of the criteria [13]. ELECTRE methods have been successfully used in many applications [13, 14, 15, 16]. Rather than aggregating the values of criteria to compare two vectors of values, the ELECTRE methods compute outranking relations (i.e. preference, indifference, or incomparability) based on the concordance of the criteria. While seminal ELECTRE method was able to infer the best decision among a set of possible decisions, further adaptations were proposed to complete ranking problems with ELECTRE III [17], and to classification problems with ELECTRE TRI [18, 19].

For our block classification problem described in Section 2.2, we used the ELECTRE TRI method. In ELECTRE TRI, the outranking relation is not computed between two actions or decisions that we want to rank, but between a decision

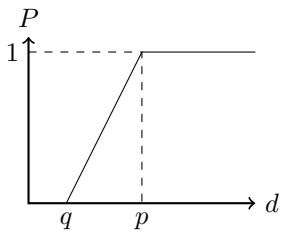

Figure 1: Example of preference function, type 5 [20], with parameters q, and p.

to classify (an urban block in our use case) and the boundaries of each class. For instance, to classify a building block, the block is modelled as a vector of values measured for each criterion (*e.g.* 0.35 for density), and each class is bounded by vectors of values for each criterion (*e.g.* from 0.2 to 0.4 for a density criterion).If the vector of the block is preferred to the lower boundary vector of the *suburban blocks* class, and the upper boundary vector of the *suburban blocks* class is preferred to the block, then the block is classified in the *suburban blocks* class.

In ELECTRE TRI, different parameters are chosen for each criterion:

- a *weight*: it conveys the importance of the criterion compared to the others;
- a *preference threshold*: if the difference between two values is above this threshold, this criterion votes for a preference outranking relation;
- an *indifference threshold*: if the difference between two values is below this threshold, this criterion votes for an indifference outranking relation;
- a *veto threshold*: if the difference between two values is above this threshold, this criterion votes for a preference outranking relation with a veto, i.e. the inverse outranking relation is not possible whatever the vote of the other criteria is.

We decided to use ELECTRE TRI because it was successfully used in other spatial problems such as knowledge revision in automated map generalisation [11], and OpenStreetMap level of detail inference [6, 10]. Furthermore, the block classification problem fully meets the criteria for a good use of ELECTRE methods [13]:

- the decision involves between 3 to 5 criteria;
- some of the criteria are evaluated with an ordinal scale instead of a continuous interval;
- the criteria are heterogeneous, as there are criteria based on geometric, semantic and topological properties (see Section 3.2);
- a low value for one criterion may not always be compensated by a high value in another criterion (this is where a veto mechanism is useful);
- small differences in values are not significant for one criterion but an accumulation of small preferences may become a significant difference.

2.3 PROMETHEE Methods

PROMETHEE are also outranking multiple criteria decision methods, based on pairwise comparisons of alternative decisions, where the decision-maker preferences are modelled to solve the incomparabilities between criteria. Preference is modelled by preference functions, that define the preference, for one criterion, of alternative a over alternative b given the deviation between values of a and b for this criterion. Figure 1 shows an example of preference function in PROMETHEE.

PROMETHEE has already been applied to geographic information science, to help the migration of thematic data from one road dataset to another with a different level of detail [5].

3 Classifying Urban Blocks from Suburban to Town Centre

The first use case presented in this paper is the classification of urban blocks in order to abstract the inner ones as built-up blocks in the generalised maps. The first subsection describes the use case with more details, and then the chosen method, ELECTRE TRI, is described. Then, the criteria are described and experiments are presented.

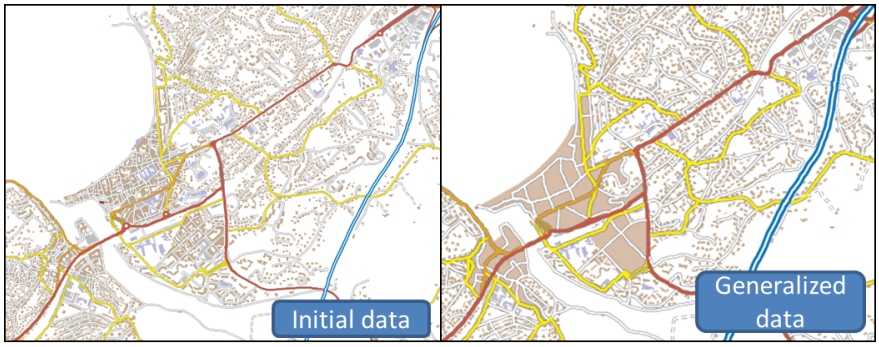

Figure 2: A topographic map at the 1:25k generalised to 1:40k scale with block abstraction in the town centre [22].

3.1 Abstracting Inner Building Blocks in Map Generalisation

Figure 2 shows the results of an automatic generalisation of an urban area at the 1:40k scale. At such scales, the inner city blocks are eroded by large road symbols and it is not possible to display the buildings independently, so a common practice in cartography is to abstract these blocks as built-up areas, displayed in beige in the figure [3], or as groups of buildings [21]. So, the initial problem is to classify the blocks as inner dense blocks or not, to decide which ones should be abstracted, and which ones should keep a display of the individual buildings. Then, at smaller scales, the whole conurbation is abstracted as a built-up area. So if we target multiple progressive generalisations at different scales [22], we need a classification that goes deeper than just inner block or not. In the context of map generalisation, such classifications were previously proposed: inner city/urban/suburban/industrial or commercial/rural/without building [23]; town centre/urban/housing estate/suburban/activity area/unitary/empty [24]. In both cases, the classification are implemented with rules derived from machine learning techniques.

Apart from the map generalisation context, other classifications of urban blocks were proposed [25, 26, 27], but they are too detailed for our needs. As a consequence, we decided to use a variation of the classifications from [23] and [24], in which we dropped the industrial or commercial category:

- empty (or almost empty) blocks
- rural blocks
- suburban blocks
- urban blocks
- inner city blocks

In order to classify the blocks, measures that describes these blocks are useful, and measures of density, proximity between buildings [3, 24, 23, 21], or of the spatial distribution [28, 24, 29] can be really useful.

To automate this classification from vector spatial data, we propose to use the ELECTRE TRI method described in Section 2.2.

3.2 Criteria Description

This section describes the five criteria we defined to classify the blocks with the ELECTRE TRI method: block density, block size, building size distribution, characteristic buildings, and proximity to town centre.

3.2.1 Block Density

This criterion is the most obvious one as inner city blocks are supposed to have a much higher density in buildings than rural blocks. Equation 1 measures density by dividing the area of the block by the sum of building footprint areas. As density values are all inferior to 1 (building footprints do not overlap road centre-lines), we empirically defined a 1.4 factor to better spread the criterion value between 0 and 1, for high density values (Equation 2). We use a minimum function to make sure that the value never exceeds 1.

$$density = \frac{\sum area_{building}}{area_{block}}) \tag{1}$$

$$value = \{\, d\ ensity, \quad if density < 0.5min(1.0, 1.4 \cdot density), \quad otherwise. \tag{2}$$

3.2.2 Block Citiness

In order to classify the citiness [30] of the blocks, their density is not sufficient, and the density of their neighbouring blocks is also important. This is why we add a criterion that measures the mean density of the adjacent blocks. The density of the blocks is computed with Equation 2.

3.2.3 Block Size

We then introduce a block size criterion because inner and urban blocks are supposed to be smaller than suburban and rural blocks where there is room for gardens, parks and agricultural fields. Equation 3 shows how we measure this block size criterion. First, we want exclude the very small blocks that are mostly artifacts of the road network such as roundabouts, and we assign a criterion value of 0 to all blocks with an area smaller than $750m^2$. Then, as the distribution of block sizes is often skewed with a small number of large and very large blocks in the outskirts of the cities, the value of the block size criterion is normalized with a maximum area above which the value is set to 0. This maximum area threshold is empirically to three times the mean area of blocks zone where the classification is performed (Equation 4).

$$value = \{\, 0\ , \quad if area_{block} \leq 750m^2 max(0.0, 1.0 - \frac{area_{block}}{max_a}), \quad otherwise. \tag{3}$$

$$max_a = 3 \cdot \frac{\sum_{block=1}^{n} area_{block}}{n} \tag{4}$$

3.2.4 Building Size Distribution

The inner parts of a city are often composed of apartment or office blocks, which means that the size of the buildings is rather large compared to the single houses that are mainly found in the suburban parts of the cities. This is why we introduce an additional criterion that analyses the distribution of the building areas inside the block. We compare the distribution of the areas in the block with the distribution of areas in the whole city. Equation 5 shows how the criterion value is computed according to the number of buildings inside the block, where $area_x$ is the smallest area of the x% percentile of the building areas in the city, max_a is computed with Equation 4, and function f is defined in Equation 6.

$$value = \{\, 0\ , \quad if n_b = 0 \frac{area_{building}}{2 \times m_a}, \quad if n_b = 1 and area_{building} \leq area_{90} \frac{max_a - area_{building}}{2 \times (m_a - max_a)} + 1, \quad if n_b = 1 and area_{building} > \tag{5}$$

In the case with large buildings in the block, we use both the mean and the median because we want to favour blocks where there is a majority of very large buildings, even if some are very small, and blocks where buildings are all large, but not necessarily very large.

$$f(x) = \{\, 0\ .9, \quad if x > area_{98} 0.8, \quad if x > area_{95} 0.7, \quad if x > area_{90} 0.6, \quad if x > area_{75} 0.5, \quad otherwise. \tag{6}$$

3.2.5 Characteristic Buildings

This criterion is based on the idea that blocks that contain a church and/or a city hall in France are more likely located in the city centre. These two characteristics buildings are related to European city patterns, but can be changed in places where there are different characteristic buildings in the city centre. Equation 7 describes how the criterion value is computed, where n is the number of buildings in the block and n_c the number of characteristic buildings in the block.

$$value = max(0.5, 1 - 0.2 \cdot (n - n_c)) \tag{7}$$

3.2.6 Proximity to Town Centre

Even if it might be constrained by local topography (*e.g.* rivers, relief), the town centre is often located around the gravity centre of the town as urban growth is often radial [31]. This is why we introduce a final criterion based on the

Criterion	Weight	Indifference	Preference	Veto
density	2.5	0.05	0.2	0.4
size	1.0	0.2	0.5	0.8
building size	1.0	0.1	0.25	0.6
char. buildings	0.7	0.05	0.2	0.1
town centre	0.5	0.1	0.2	0.95
citiness	1.2	0.05	0.2	0.4

Table 1: This table summarizes the parameters and weights used in our experiments on the classification of town centre blocks.

proximity to the geometrical town centre. Equation 8 describes how the criterion value is computed, where maxDist is $\frac{2}{3}$ of the maximum distance for all blocks. We do not use the maximum distance between the centroid of the town polygon and the centroids of the block polygons to normalise the value because some large protrusions in urban growth were causing underestimations of this criterion for some of the towns we tested. This criterion should not use a veto in most cases as a large distance may just indicate that the town centre is far from the centroid because of the local topography.

$$value = max(0.0, (1.0 - \frac{distance}{maxDist})^2) \tag{8}$$

3.3 Experiments

The proposed method was experimented on towns of different sizes from France (Limoges (4097 blocks), Clermont-Ferrand (7503 blocks), Lourdes (367 blocks), Orthez (558 blocks), Salies-de-Béarn (104 blocks)), Germany (Würzburg; 2201 blocks), and the USA (Madison, 3513 but many of them are artifacts from the multiple lanes of the roads). The data comes from the French national mapping agency (IGN), and from OpenStreetMap for Germany and USA. Table 1 shows the parameters we empirically defined in the experiments of block classification, with an important weight on the density criterion. The boundaries of the classes were kept pretty simple with the $[0, 0.2]$ interval for the *empty* class, the $[0.2, 0.4]$ interval for the *rural* class, the $[0.4, 0.6]$ interval for the *suburban* class, the $[0.6, 0.8]$ interval for the *urban* class, and the $[0.8, 1]$ interval for the *innner city* class.

Figure 3 shows the results on the smallest town in our experiment, Salies-de-Béarn. In this case, the results are really good, even if the suburban area could have extended a little bit in the south of the inner city.

In the larger cities of Limoges (Figure 4) and Madison (Figure 5), we still consider the results good for our purpose because most of the relative differences between blocks are well inferred, and we can see the different areas where housing types and densities vary. But a small there is a small bias in classification, particularly in the Limoges case, with a shift towards classes with less density: the blocks classified as urban tend to be inner city blocks, the blocks classified as suburban are rather urban blocks, and the blocks classified as rural are often suburban blocks even though some really are rural.

This shift could be explained by the mitigating effect of two criteria: characteristic buildings and town centre, that have even values for most of the blocks. We ran robustness tests by removing some of the criteria, presented in the following subsection, to verify this hypothesis. We also suspect an influence of our simple way to set the boundary values of the classes.

3.4 Evaluation

3.4.1 Evaluation and Robustness

To evaluate the results of the ELECTRE TRI classification, we visually classified a third of the blocks (115) of the town of Lourdes, and we then ran the ELECTRE classification on the same blocks to compare. This comparison clearly confirmed our first visual evaluation of the method with this shift towards less dense classes, and a globally good relative classification. The results of the comparison are presented in Table 6, where classification differences are measured considering the ordering of the classes: class *empty* is given a value of 0, and class *inner city* is given a value of 4. For most of the compared blocks, the difference is -1 because of this classification shift, and 87% of the blocks have a class difference comprised between -1 and 1, showing the consistency of the classification with our expert classification. The blocks with a difference value of 2 or 3 are in fact empty blocks created by roundabouts or branching crossroads. The expert classified them as *empty*, but ELECTRE, due to the citiness criterion, classified them in classes similar to their

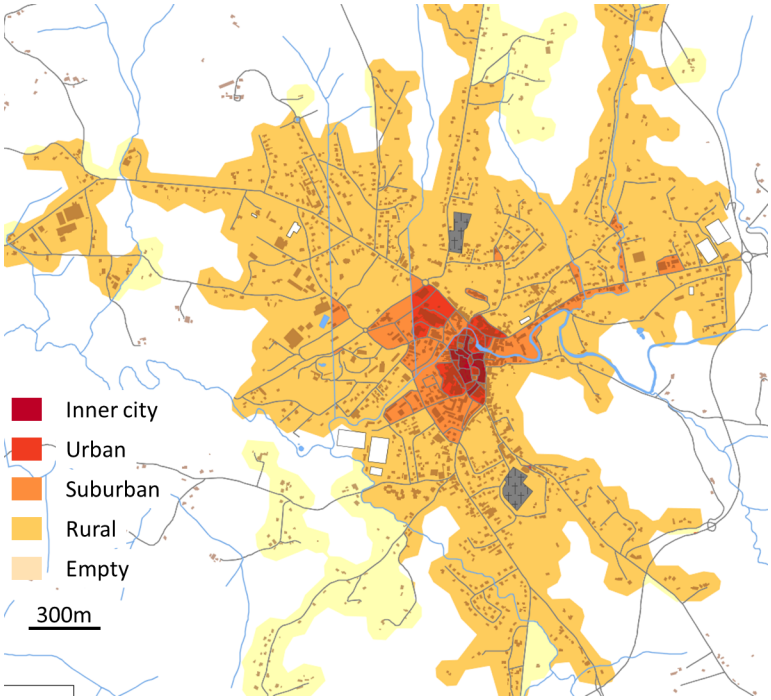

Figure 3: Results of the ELECTRE TRI block classification on the small town of Salies-de-Béarn in France.

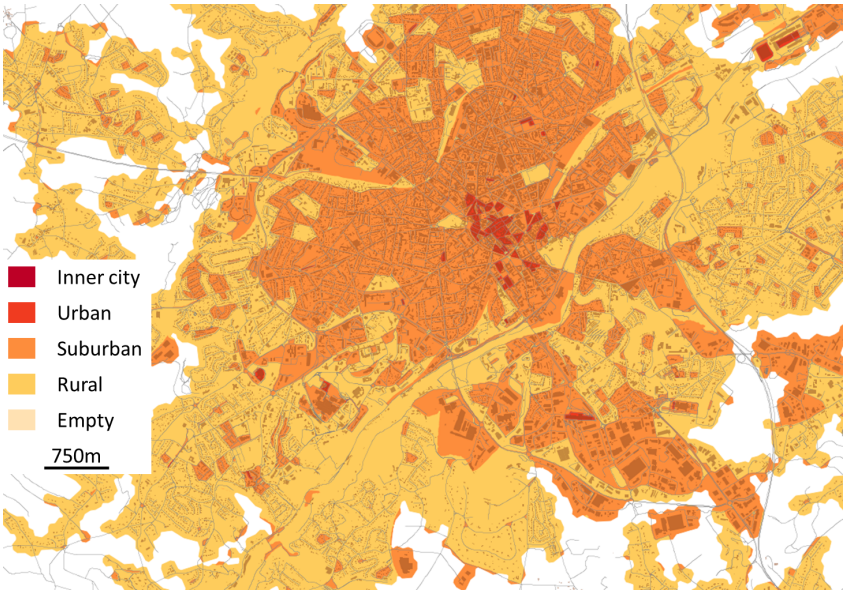

Figure 4: Results of the ELECTRE TRI block classification on the city of Limoges in France.

adjacent blocks. It gives a smoother view of the city classification, and it is not a problem for our map generalisation purpose where these blocks are easily sorted out, so we do not consider this artifact of the classification as a bad result.

In order to test the robustness of the classification method, and also to verify our hypothesis about the influence of some criteria in the classification shift we observed, we ran the classification multiple times with a varying set of criteria (but with similar parameters each time). When removing only one criteria, around 90% of the blocks remained classified the same way as with the six criteria. But all the criteria, do have a small influence on the results. For instance, in Figure 7, removing the town centre criterion changed the classification of some blocks at the border of the town, and removing the characteristic buildings criterion changed the classification of some blocks containing churches (there

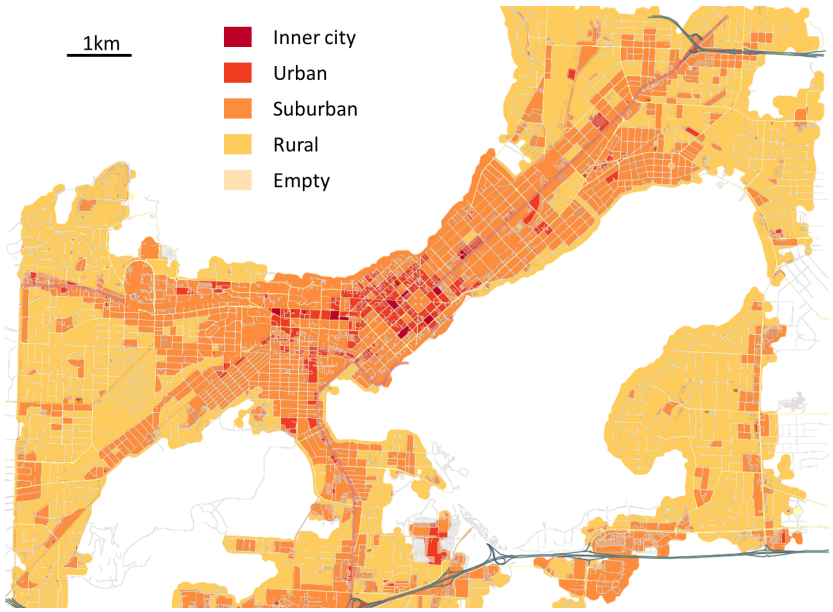

Figure 5: Results of the ELECTRE TRI block classification on the city of Madison, WI, USA, with OpenStreetMap data.

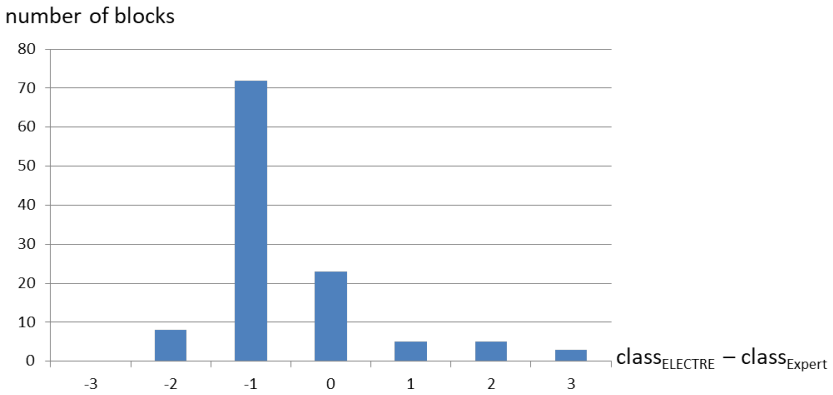

Figure 6: Comparison of ELECTRE classification with a classification from an expert cartographers on 115 blocks from the town of Lourdes, France.

are many churches in Lourdes). The figure also shows that the hypothesis on the influence of these two criteria on the classification shift, is clearly not verified as the urban block are still mostly classified as suburban and the suburban blocks are still mostly classified as rural blocks.

The other hypothesis was the influence of the boundary values of the ordered classes in the ELECTRE TRI process, which are set by default to 0/0.2/0.4/0.6/0.8/1 in our experiments. Figure 8 shows two classifications of the blocks in Lourdes with varying boundary values, and the classification clearly changes, with a left result that is quite close to the expert classification. This method has a large number of parameters to set empirically, and the best way to find optimal values would be to explore the parameters space with a specific method, as proposed in recent research on spatial data processing [32, 33]. This parameter space exploration are based on methods designed to explore the parameters of simulation models [34].

3.4.2 Comparative Evaluation

We compared the classification results from ELECTRE to a classicial and basic convolutional neural network (CNN) quite similar to the seminal LeNET model [35]. The model was trained with images containing only the buildings and the roads of block (no neighbouring block), extracted from three of the cities of the experimental datasets. We

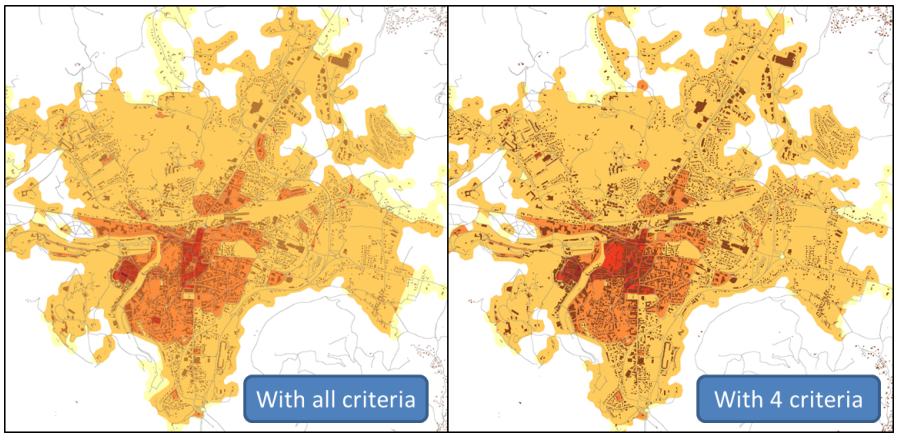

Figure 7: Two classifications of the blocks of the town of Lourdes, France. With all criteria at the top, and without characteristic building and town centre criteria at the bottom.

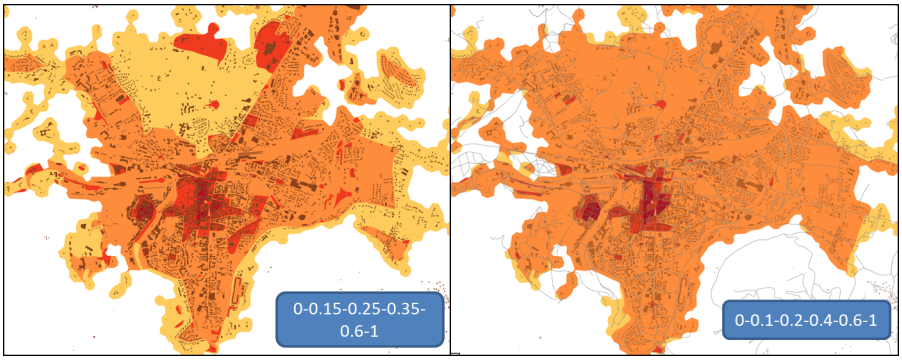

Figure 8: Two classifications of the blocks of the town of Lourdes, France, with alternative values for class boundaries: 0/0.15/0.25/0.35/0.6/1 on the left, and 0/0.1/0.2/0.4/0.6/1 on the right.

visually classified the images ourselves, and trained the model on a random sample of 75% of the blocks, the remaining 25% being used for evaluation. The images are 128x128 pixels generated with the vector data, roads and buildings being displayed in black. The total accuracy obtained on the evaluation dataset is slightly below 0.89, but the confusion matrix displayed in Table 2 shows that the CNN classifier is much better for classes without (class 0), or few (class 1 and 2) buildings than for the dense block classes that we are interested in.

Figure 9 shows example of block images that were misclassified by the CNN. These six blocks were correctly classified by the ELECTRE classifier.

We also reproduced the classification of blocks from [23] that uses a Support Vector Machine classifier (SVM) and descriptors on building size, shape and distribution, quite similar to the ones used in the ELECTRE TRI method. The classifier was trained on the same blocks as the CNN classifier. The results are presented on Figure 10. As this method

ground truth \ prediction	0	1	2	3	4
0	0.975	0.022	0.003	0.0	0.0
1	0.051	0.775	0.175	0.0	0.0
2	0.0	0.061	0.852	0.080	0.008
3	0.0	0.0	0.172	0.638	0.190
4	0.0	0.0	0.047	0.238	0.714

Table 2: This table shows the confusion matrix of the CNN applied on an evaluation dataset of blocks. The inner city blocks correspond to class 4.

Figure 9: the image on the left is classified as a class 2 (*i.e. suburban*) block by the CNN, and the 5 images on the right are classified as class 3 (*i.e. urban*) blocks; all should be classified as class 4 blocks, *i.e.* inner city blocks.

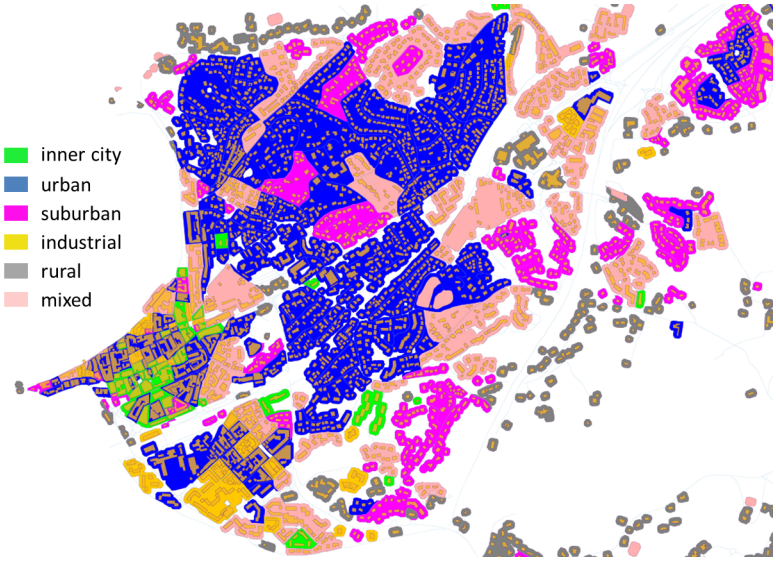

Figure 10: Results of the SVM-based classification of blocks.

classifies all buildings individually, and not blocks, the block classification is derived from the classification of the majority of buildings; if there is no majority among the classes of the buildings, the block is classified as *mixed*. This method gives interesting results but has two disadvantages: first, the different areas of the city are less smoothly defined with several urban blocks surrounded by inner city blocks; then, it is difficult to know how to process the mixed blocks.

We compared all three methods on the same cities (Figure 11). The differences are clearly closer to zero with the CNN than with the SVM, confirming the individual evaluation of each method. The classes of the CNN method tend to be a little too small compared to the ELECTRE method, which classes are also a little too small compared to the expert evaluation. It is interesting to notice that the empty blocks created by crossroads and roundabout are correctly classified by the CNN method, which explains the positive differences in Figure 11. Otherwise, the SVM method appears to be a bit random sometimes, with many dense blocks classified as rural or suburban, and conversely rural and suburban blocks classified as urban or inner city. This results can partly be explained by the mixed and industrial blocks that we had to transform into the 5-class classification, for the sake of comparison (we used the most represented class among the buildings of the blocks, excluding the building classified as industrial).

4 Ordering Buildings to Keep in the Map

4.1 Building Elimination in Map Generalisation

The cartographic generalisation of urban areas has been the main focus of cartographers for years: as the density of information to map is higher, the simplification and abstraction decision are more drastic. The usual way to generalise urban areas is to use blocks delineated by the road network to partition the data: each block is generalised as a whole [3]. At the 1:25k and smaller scales, the road symbols erode the room available in blocks to display the individual buildings: displacements, aggregations or eliminations are necessary [36, 37, 24]. When buildings are aligned along a road, a specific operation that preserves the alignment, called typification, can be used [38, 39, 40, 41]. In this use case, we focus on the elimination operation, which can be rephrased as a ranking problem where the building ranked first is the first one to be eliminated. This is clearly a multiple criteria problem as the existing method consider the shape, the size, the position in the block, or the local neighbourhood of each building when making the elimination decision [36, 37].

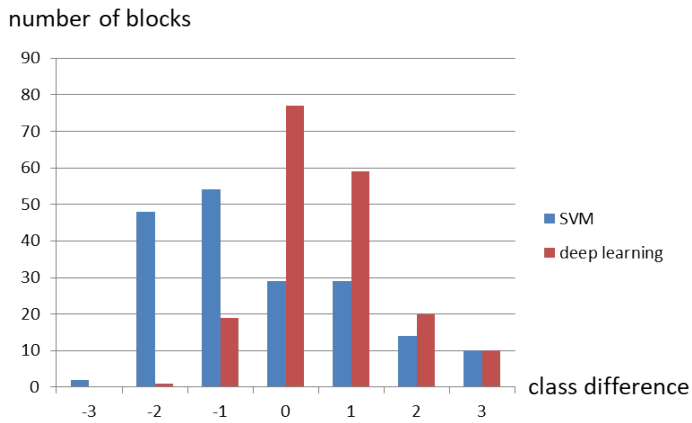

Figure 11: Differences of classification between ELECTRE and SVM, and ELECTRE and CNNs.

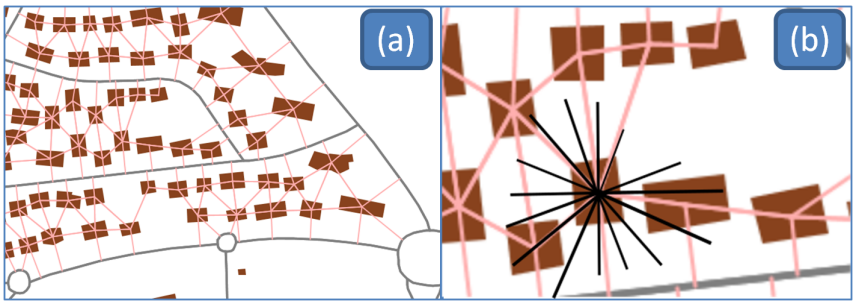

Figure 12: (a) Three blocks triangulated to compute proximity between buildings and roads. (b) congestion computed based on the proximity triangulation.

As building elimination in an urban block during map generalisation can be considered as a complete ranking problem with multiple criteria, two methods from the literature were tested: ELECTRE III, a variation of the ELECTRE methods for ranking problems [13], and PROMETHEE II, a complete ranking variation of the PROMETHEE methods [20].

4.2 Criteria Description

This section describes the criteria that are used for both ELECTRE and PROMETHEE methods, to rank the buildings in a block, the first one being the one to eliminate first, and the last one being the one to eliminate last.

4.2.1 Building Congestion

The building congestion criterion is based on an analysis of the proximity of each building with other buildings and surrounding roads. The proximity is computed with a Delaunay triangulation of the buildings centroids and roads vertices, as proposed by [36] (Figure 12a). Than, the weight of each edge is computed with the smallest distance between the symbols of both connected features, rather than the length of the edges.

Then, the congestion of a building is computed by analyzing angle portions and the length of the triangulation edges inside each angle portions (Figure 12b). The value of the criterion is then obtained by summing the congestion values for each angle portion (Equation 9).

$$value = \sum_{i=1}^{16} congestion(\frac{i \cdot \pi}{16} - \frac{\pi}{16}, \frac{i \cdot \pi}{16}) \tag{9}$$

4.2.2 Congestion Direction

In order to differentiate the buildings that have a strong congestion on some directions but none on another direction, which can be moved to that direction, to buildings that have a slight congestion on every direction, we introduced a complementary criterion that counts the angle portions where there is congestion. So the congestion direction value is an integer comprised in [0,16] where 0 means no congestion, and 16 means that all directions are congested.

4.2.3 Corner Building

Buildings located in the corner of a block, i.e. close to a road intersection are often important landmarks in map reading. This is important to keep them in the map during generalisation and they should be put at the end of the elimination list. To identify corner buildings, we use an algorithm from [42, p. 379], which first finds road corners, and then searches if there is a building in the corner (using the proximity triangulation).

$$value = \{ 1 \ , \quad ifthebuildingisinacorneroftheblock 0, \quad otherwise. \tag{10}$$

4.2.4 Building Size

The size criterion is an important one as small buildings should be eliminated before large buildings. Usually, map specifications contain thresholds of minimum area for a building to be kept in the map [43], so we used such a threshold to compute the value of the criterion (Equation 11).

$$value = \{ 1 \ , \quad if area_{building} < area_{min} - \frac{1}{2 \cdot area_{min}} \cdot area_{building} + \frac{2 \cdot area_{min} - 1}{2 \cdot area_{min}} \quad if area_{building} < 2 \cdot area_{min} - \frac{1}{4 \cdot area_{min}} \tag{11}$$

4.2.5 Nature of the Building Neighbours

This last criterion is similar to the one on corner buildings as it favours the elimination of buildings surrounded by other buildings compared to buildings surrounded by roads. The criterion is based on the principle that buildings along roads are important landmarks in the map and should not be eliminated if not necessary. Equation 12 shows how the value for this criterion is computed based on the proximity triangulation.

$$value = \{ 0 \ , \quad if |N| = 0 \frac{k}{|N|}, \quad otherwise. \tag{12}$$

where N is the set of neighbours in the proximity triangulation,

and k is the number of edges connected to a building in the triangulation.

4.3 Experiments

Criterion	Weight	Indifference	Preference	Veto
congestion	1.5	0.2	0.3	0.5
congestion direction	1	0.2	0.3	0.5
building size	1.5	0.1	0.2	0.7
corner building	0.8	0.2	0.4	0.5
neighbours	1.0	0.1	0.2	0.7

Table 3: This table summarizes the parameters and weights used in our experiments on the elimination of buildings in a block.

As building elimination also applies to the generalisation of urban areas, we used the same dataset to experiment both ELECTRE and PROMETHEE based methods. Table 3 summarizes the parameters we used in this experiment for ELECTRE, and Table 4 summarizes the parameters used for PROMETHEE, the most weighted criteria being congestion and size. Figure 13 shows some visual results with ELECTRE for eleven adjacent blocks with various characteristics. The results are globally very good, as the method makes good decisions on the top of the list, as well as at the bottom of the list. We highlighted two questionable decisions in Figure 13: one in a block with only three buildings (the building listed third should be the first to be eliminated), but it is not so important as the free room in

Criterion	Weight	type of preference function	q	p
congestion	0.2	stair-shaped (type 4)	0.2	0.4
congestion direction	0.2	stair-shaped (type 4)	0.2	0.4
building size	0.2	piecewise-linear (type 5)	0.1	0.7
corner building	0.2	binary (type 2)	0.5	-
neighbours	0.2	piecewise-linear (type 5)	0.15	0.6

Table 4: Types of preference functions from [20] and associated parameters used for our five criteria.

the block enables a generalisation without elimination; one in a block with many small buildings (probably residential housing), where one building in the middle could have been chosen as the first to be eliminated by a human cartographer, but it is not a crystal-clear decision.

Figure 13: Multiple blocks with the elimination ordering computed with ELECTRE (1 means the first to be eliminated). Block delineations are displayed in blue, green highlights some good decisions, orange highlights questionable decisions.

4.4 Evaluation

The first evaluation carried out was to compare the methods to a manual selection of the buildings to eliminate first and last. We selected 50 blocks from the Lourdes dataset, using the output of the block classification method presented in this paper to make sure different types of blocks were selected. In each of the selected block, an expert cartographer chose the first building to eliminate, and the last. As only the first to be eliminated was selected, we used blocks were the number of buildings that should be eliminated is small: if more than ten buildings have to be eliminated, which happens frequently in very large suburban blocks, it is not so important if the first one is different from an expert choice; if only two should be eliminated, we have to make sure that the proposed algorithm selects the good ones. The expert choices were compared to the ELECTRE and PROMETHEE choices, and the Table 5 summarizes this comparison. To measure the distance between the methods output and the choices of the expert, we searched the position of the expert first (or last) choice in the lists and weighted the position by the number of buildings in the block. For instance, of the first choice of the expert in third a the list of 6 buildings, the distance is $2/6 = 0.33$.

The results of the evaluation are clearly in favour of the PROMETHEE method, with a particular difference on the last building to be eliminated where the PROMETHEE method often makes the same choice as the expert. Regarding

elimination method	mean distance first	mean distance last	median distance first	median distance last
ELECTRE	0.46	0.51	0.45	0.55
PROMETHEE	0.35	0.19	0.35	0.07

Table 5: Mean and median distances between the expert choices and the ouput lists of ELECTRE and PROMETHEE for 50 blocks in the Lourdes dataset.

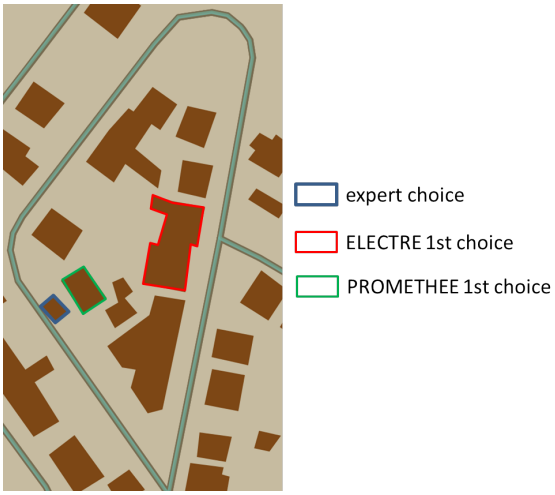

Figure 14: Example of disagreement on a ground truth block. The expert choice is more guided by building size while PROMETHEE is more a compromise between size and congestion. ELECTRE seems more guided by congestion.

the first choice to be eliminated, the agreement between both choices is heterogeneous (the standard deviation is 0.3 for PROMETHEE). There are many cases where the first choice of the expert is the same as the first choice from PROMETHEE (15 out of the 50 ground truth blocks), but there are also many cases where the first choice of the expert is too far in the PROMETHEE list to be eliminated (distance bigger than 6 in 14 out of the 50 ground truth blocks). An analysis of the cases where there is a disagreement between PROMETHEE and the expert shows that in most cases, the expert weighted much more the size criterion than PROMETHEE (Figure 14).

In this building elimination use case, we did not carry out any comparison with a machine learning technique as we had no mean to speed an annotation process up. But we compared with the existing elimination algorithm [37] (Figure 15). In order to compare the lists obtained with the three methods, three distances designed for string comparisons were used:

- the Hamming distance [44, 8], $hd \in [0, nb_{buildings}]$;
- the edit-distance with swaps, $ed \in [0, nb_{buildings}]$;
- the Jaro-Winkler distance [45], $jwd \in [0, 1]$.

Compared to the first two distances, the Jaro-Winkler distance weights more errors in the beginning of the list than errors at the end of the list. We carried out some comparisons on lists containing all the buildings in a block, and on lists containing only the first five to be eliminated (it is rare to have more than five buildings eliminated). When the full lists are compared, the distances were normalized between 0 and 1, by dividing the distance by the number of buildings in the block. The comparisons are summarized in Table 6. These quantitative results confirm the assessment with the experts, as ELECTRE seems very different than both other methods. However, the difference is reduced when we only consider the first 5 buildings of the blocks. And even if they are more similar, PROMETHEE and AGENT methods are quite different, confirming the variability in building elimination results.

5 Conclusion and Further Research

To conclude, this paper showed that multi-criteria decision techniques are useful for automated cartography, surpassing machine learning techniques for problems lacking massive training datasets. They can combine criteria based on

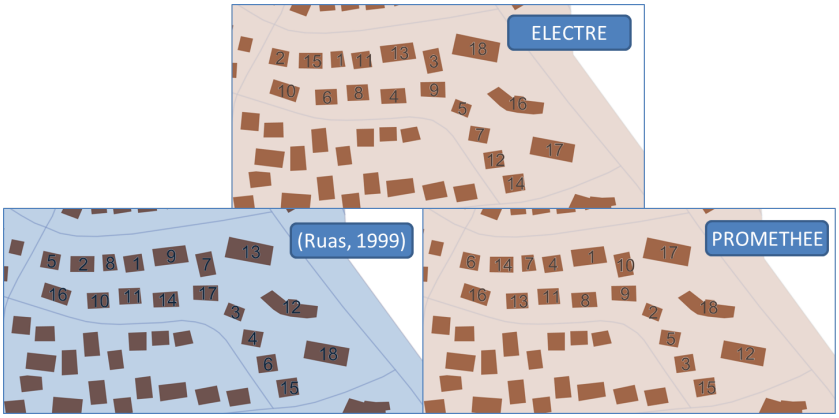

Figure 15: Different priority lists for elimination obtained on the same block with the three methods. 1 means that it is the first building to be eliminated.

compared lists	Hamming distance	edit-distance	Jaro-Winkler distance
ELECTRE vs PROMETHEE	0.95	0.92	0.36
ELECTRE vs AGENT	0.94	0.91	0.388
PROMETHEE vs AGENT	0.39	0.82	0.201
ELECTRE vs PROMETHEE (first 5)	3.25	3.14	0.569
PROMETHEE vs AGENT (first 5)	2.18	2.25	0.305
ELECTRE vs AGENT (first 5)	3.27	3.14	0.594

Table 6: Means of different distances computed on the priority lists obtained from the different elimination algorithms.

geometry, topology and semantics, which would be hard to compare or combine with other methods. As each block or each building is almost a unique situation given its own characteristics and its neighbourhood, machine learning model are more complex to train. The paper also showed that ELECTRE TRI and PROMETHEE perform well for our two specific use cases on urban block classification, and building elimination in a block. In the second use case, PROMETHEE outperformed ELECTRE with similar criteria. Consequently, the paper shows that there is not one method that is better for all map generalisation problems, and the different existing multi-criteria decision techniques, not only PROMETHEE and ELECTRE, should be investigated when tackling a new sub-problem of map generalisation. But in general, such techniques can be applied to improve the actual automated or semi-automated generalisation processes that are used in national mapping agencies to create large series of topographic maps at different scales (printed on paper or online maps).

To go further, the paper showed that PROMETHEE and ELECTRE are complex to parameterize when applied to map generalisation, and we believe that systematic parameter robustness analyses should be carried out to identify optimal parameters for a given sub-problem of map generalisation. This robustness analysis could rely on techniques for the optimal exploration of parameter spaces, adapted from research on spatial simulation [34, 33].

We also plan to use multi-criteria decision techniques for other map generalisation applications. The more important one would road network selection, i.e. choosing the important roads to keep in the map [7], where the existing techniques to combine multiple criteria based on rules, cost functions, or machine learning, are not completely satisfying. Furthermore, as ELECTRE was already successfully used to infer the level of detail of OpenStreetMap features [6, 10], it would be necessary to reuse the aim is not just to infer the level of detail but to generalise such data to a given level of detail.

References

[1] Dirk Burghardt, Cécile Duchêne, and William Mackaness, editors. *Abstracting Geographic Information in a Data Rich World*. LNG&C. Springer, Berlin Heidelberg, 2014.

[2] Cécile Duchêne, Blanca Baella, Cynthia A. Brewer, Dirk Burghardt, Barbara P. Buttenfield, Julien Gaffuri, Dominik Käuferle, François Lecordix, Emmanuel Maugeais, Ron Nijhuis, Maria Pla, Marc Post, Nicolas Reg-

nauld, Lawrence V. Stanislawski, Jantien Stoter, Katalin Tóth, Sabine Urbanke, Vincent van Altena, and Antje Wiedemann. Generalisation in Practice Within National Mapping Agencies. In Dirk Burghardt, Cécile Duchêne, and William Mackaness, editors, *Abstracting Geographic Information in a Data Rich World*, Lecture Notes in Geoinformation and Cartography, pages 329–391. Springer International Publishing, 2014.

[3] Anne Ruas and William A. Mackaness. Strategies for urban map generalization. In *Proceedings of the 18th ICA/ACI International Cartographic Conference*, volume 3, pages 1387–1394. ICA, 1997.

[4] William A. Mackaness and Geoffrey Edwards. The importance of modelling pattern and structure in automated map generalisation. In *Proceedings of the Joint ISPRS/ICA Workshop on Multi-Scale Representations of Spatial Data*, pages 7–8, 2002.

[5] Kusay Jaara, Cécile Duchêne, and Anne Ruas. Preservation and modification of relations between thematic and topographic data throughout thematic data migration process. In Buchroithner, editor, *Cartography From Pole To Pole, Selected Contributions to the 26th International Conference of the ICA*, Lecture Notes in Geoinformation and Cartography, pages 103–118. International Cartographic Association, Springer, 2013.

[6] Guillaume Touya and Carmen Brando-Escobar. Detecting level-of-detail inconsistencies in volunteered geographic information data sets. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 48(2):134–143, 2013.

[7] Guillaume Touya. A road network selection process based on data enrichment and structure detection. *Transactions in GIS*, 14(5):595–614, 2010.

[8] Maxim A. Rylov and Andreas W. Reimer. A comprehensive multi-criteria model for high cartographic quality point-feature label placement. *Cartographica*, 49(1):52–68, 2014.

[9] Ana-Maria Olteanu-Raimond, Sébastien Mustière, and Anne Ruas. Knowledge formalization for vector data matching using belief theory. *Journal of Spatial Information Science*, 10:21–46, 2015.

[10] Guillaume Touya and Andreas Reimer. Inferring the scale of OpenStreetMap features. In Jamal Jokar Arsanjani, Alexander Zipf, Peter Mooney, and Marco Helbich, editors, *OpenStreetMap in GIScience*, Lecture Notes in Geoinformation and Cartography, pages 81–99. Springer International Publishing, 2015.

[11] Patrick Taillandier and Franck Taillandier. Multi-criteria diagnosis of control knowledge for cartographic generalisation. *European Journal of Operational Research*, 217(3):633–642, 2012.

[12] M. Sester, Y. Feng, and F. Thiemann. Building generalization using deep learning. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4:565–572, 2018.

[13] José Figueira, V. Mousseau, and B. Roy. Electre methods. In José Figueira, S. Greco, and M. Ehrogott, editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, pages 133–162. Springer-Verlag, 2005.

[14] K. Srinivasa Raju, Lucien Duckstein, and Cecile Arondel. Multicriterion analysis for sustainable water resources planning: A case study in spain. *Water Resources Management*, 14(6):435–456, 2000.

[15] Rui P. Lourenço and João P. Costa. Using electre tri outranking method to sort momilp nondominated solutions. *European Journal of Operational Research*, 153(2):271–289, 2004.

[16] E. Georgopoulou, Y. Sarafidis, S. Mirasgedis, S. Zaimi, and D. P. Lalas. A multiple criteria decision-aid approach in defining national priorities for greenhouse gases emissions reduction in the energy sector. *European Journal of Operational Research*, 146(1):199–215, 2003.

[17] B. Roy, M. Présent, and D. Silhol. A programming method for determining which paris metro stations should be renovated. *European Journal of Operational Research*, 24(2):318–334, 1986.

[18] Wei Yu. *Aide multicritère à la décision dans le cadre de la problématique du tri : concepts, méthodes et applications*. PhD thesis, Université Paris Dauphine, 1992.

[19] J. Almeida-Dias, J. R. Figueira, and B. Roy. Electre tri-c: A multiple criteria sorting method based on characteristic reference actions. *European Journal of Operational Research*, 204(3):565–580, 2010.

[20] Jean-Pierre Brans and Bertrand Mareschal. Promethee methods. In José Figueira, Salvatore Greco, and Matthias Ehrogott, editors, *Multiple Criteria Decision Analysis: State of the Art Surveys*, volume 78 of *International Series in Operations Research & Management Science*, chapter 5, pages 163–186. Springer-Verlag, New York, 2005.

[21] Min Deng, Jianbo Tang, Qiliang Liu, and Fang Wu. Recognizing building groups for generalization: a comparative study. *Cartography and Geographic Information Science*, 45(3):187–204, 2018.

[22] Guillaume Touya and Marion Dumont. Progressive block graying and landmarks enhancing as intermediate representations between buildings and urban areas. In *Proceedings of 20th ICA Workshop on Generalisation and Multiple Representation*, Washington DC, USA, 2017.

[23] Stefan Steiniger, Tilman Lange, Dirk Burghardt, and Robert Weibel. An approach for the classification of urban building structures based on discriminant analysis techniques. *Transactions in GIS*, 12(1):31–59, 2008.

[24] Julien Gaffuri and Jenny Trévisan. Role of urban patterns for building generalisation: An application of AGENT. In *ICA Workshop on Generalisation and Multiple representation, commission on map generalisation*. ICA, 2004.

[25] William A. Mackaness and Omair Z. Chaudhry. Automatic classification of retail spaces from a large scale topographic database. *Transactions in GIS*, 15(3):291–307, 2011.

[26] Mark Thurstain-Goodwin and David Unwin. Defining and delineating the central areas of towns for statistical monitoring using continuous surface representations. *Transactions in GIS*, 4(4):305–317, 2000.

[27] Patrick Lüscher and Robert Weibel. Semantics matters: Cognitively plausible delineation of city centres from point of interest data. In *13th Workshop of the ICA commission on Generalisation and Multiple Representation*. ICA, 2010.

[28] Annabelle Boffet. Creating urban information for cartographic generalisation. In *International Symposium on Spatial Data Handling (SDH)*, 2000. event-place: Beijing, China.

[29] Zhiwei Wei, Qingsheng Guo, Lin Wang, and Fen Yan. On the spatial distribution of buildings for map generalization. *Cartography and Geographic Information Science*, 45(6):539–555, 2018.

[30] Omair Z. Chaudhry and William A. Mackaness. Automatic identification of urban settlement boundaries for multiple representation databases. *Computers, Environment and Urban Systems*, 32(2):95–109, 2008.

[31] Michael Batty and Paul Longley. *Fractal Cities: A Geometry of Form and Function*. Academic Press, San Diego, CA and London, 1994.

[32] Juste Raimbault. *An Urban Morphogenesis Model Capturing Interactions Between Networks and Territories*, pages 383–409. Springer International Publishing, Cham, 2019.

[33] Benoit Costes and Julien Perret. A hidden Markov model for matching spatial networks. *Journal of Spatial Information Science*, 18:57–89, 2019.

[34] Romain Reuillon, Mathieu Leclaire, Juste Raimbault, Hélène Arduin, Paul Chapron, Guillaume Chérel, Etienne Delay, Pierre-François Lavallée, Jonathan Passerat-Palmbach, Pierre Peigne, Julien Perret, and Sébastien Rey-Coyrehourcq. Fostering the use of methods for geosimulation models sensitivity analysis and validation. In *ECTQG 2019*, Mondorf, Luxembourg, 2019.

[35] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[36] Anne Ruas. A method for building displacement in automated map generalisation. *International Journal of Geographical Information Science*, 12(8):789–803, 1998.

[37] Anne Ruas. *Modèle de généralisation de données géographiques à base de contraintes et d'autonomie*. PhD thesis, Université Marne La Vallée, 1999.

[38] Monika Sester. Kohonen feature nets for typification. In *6th ICA Workshop on progress in automated map generalisation*. ICA, 2001.

[39] N. Regnauld. Contextual building typification in automated map generalization. *Algorithmica*, 30(2):312–333, 2001.

[40] D. Burghardt and A. Cecconi. Mesh simplification for building typification. *International Journal of Geographic Information Science*, 21(3):283–298, 2007.

[41] Ibrahim Bildirici, Serdar Aslan, Özlem Simav, and Osman N. Çobankaya. A generic application for building typification. In *Proceedings of 14th ICA Workshop on Generalisation and Multiple Representation*, Paris, France, 2011.

[42] Patrick Taillandier. *Modèle de révision de connaissances dans un processus de généralisation de données géographiques à base d'agents*. PhD thesis, Université Paris-Est, 2008.

[43] Jantien Stoter, Dirk Burghardt, Cécile Duchêne, Blanca Baella, Nico Bakker, Connie Blok, Maria Pla, Nicolas Regnauld, Guillaume Touya, and Stefan Schmid. Methodology for evaluating automated map generalization in commercial software. *Computers, Environment and Urban Systems*, 33(5):311–324, 2009.

[44] R. W. Hamming. Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2):147–160, 1950.

[45] W. E. Winkler. The state of record linkage and current research problems. *Statistics of Income Division, Internal Revenue Service Publication*, R99, 1999.