# A framework to formalize multi-scales land-use generalization in the ScaleMaster 2.0

**Jean-François Girres[1], Guillaume Touya[2]**

[1] UMR GRED – Université Paul Valéry Montpellier, Route de Mende 34199 Montpellier Cedex 5, France
[2] COGIT - IGN France, 73 avenue de Paris 94165  Saint-Mandé, France

**Abstract.** This article presents a general framework to formalize the generalization of land-use coverages within the ScaleMaster2.0 model, a multi-scale automatic generalization system, developed to facilitate smooth transitions (i.e. cartographic continuum) between level of details in a multi-representation database (MRDB). The article describes the land-use simplification method developed, the strategy used to manage generalization on large areas, and the way to formalize multi-scales land-use generalization rules in the ScaleMatsre2.0. Finally, an experiment is provided on two datasets to illustrate the proposition: a land-use coverage in the area of Dakar (Senegal) and a map of sedimentary floor off the coast of Brittany (France).

**Keywords:** Generalization, MRDB, Land-use, ScaleMaster 2.0

## 1. Introduction

Multi-representation databases (MRDB) allow representing the same real world entities in different cartographic databases with their own level of detail. Thus, it enables navigating from a scale to another using multi-scales geoportals. To facilitate smooth transitions (i.e. cartographic continuum) between representations, cartographic generalization allows deriving intermediate representations, based on the most detailed database (Figure 1). In this context, the ScaleMaster2.0 (Touya & Girres, 2013) was developed as a multi-scales generalization system, allowing the derivation of intermediate levels of detail by formalizing generalization rules from a MRDB. The ScaleMaster2.0 is an automatic generalization system, developed as an extension of the ScaleMaster, initially proposed by Brewer & Buttenfield (2007). The ScaleMaster2.0 is implemented using the CartAGen library (Renard et al. 2010), a Java platform developed since 2009 by the generalization team of the COGIT lab. The parameterization of the ScaleMaster2.0

model is realized using three XML interfaces associated to the Java core: the ScaleMaster.xml file (to formalize generalization rules), the Parameters.xml (to specify the input databases used and the output results scale), and the Symbology.xml file.
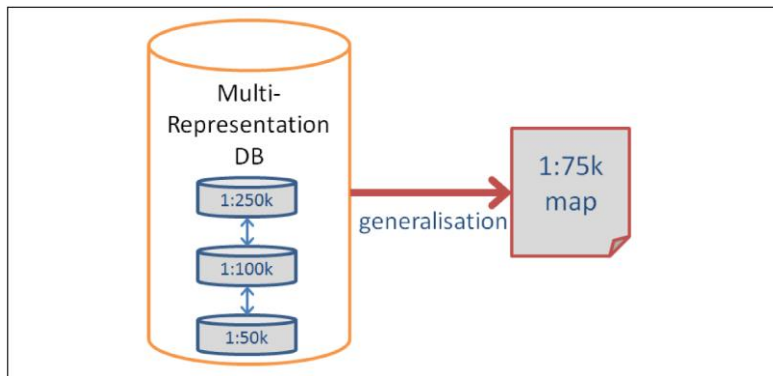


**Figure 1.** Intermediate scale (1:75k) derived by automatic generalization from a larger scale (1:50k) in a MRDB

Since the beginning of the development the ScaleMaster2.0, several advances in the field of cartographic generalization have been proposed and formalized in the framework, as for instance the generalization of unusual map themes from OpenStreetMap (Touya & Girres, 2014), or methods to derive intermediate scales by taking account of the consistency between level of details, called multi-representation aware (MR-aware) generalization (Girres & Touya, 2014). Nevertheless, the question of land-use (or more generally of categorical data) generalization in the ScaleMaster2.0 model has not been studied yet.
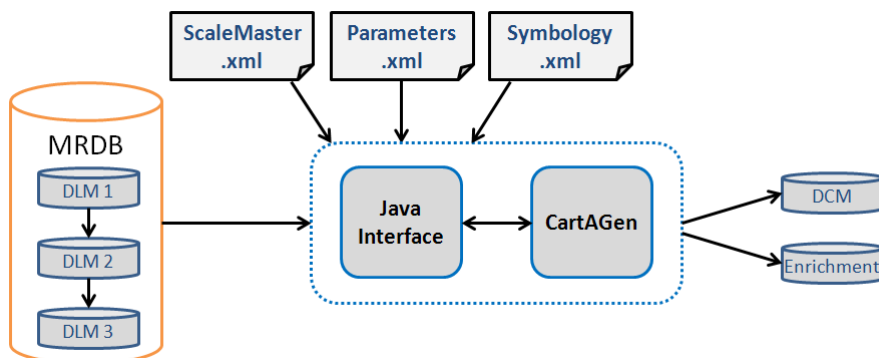


**Figure 2.** Implementation of the ScaleMaster2.0 model using the CartAGen library

Then, this article presents in section 2 the different strategies proposed to handle land-use generalization. Section 3 presents the framework proposed to manage land-use generalization in the ScaleMaster2.0. Finally, an experiment of the model will be proposed in section 4, before concluding and evocating further works.

## 2. Land-use generalization strategies

Land-use coverages, as far as all categorical data (e.g. geological or pedological maps) constitute a partition of the real world, where a land-use class is affected for each element of the partition. Land-use is generally represented in geographic databases as a single coverage composed of different land-use classes. But in some cases, geographical databases represent land-use as the joining of independent layers, each of them composing a land-use class.

In a land-use generalization process, each land-use class can be handled differently according to the desired output scale, the geographical context of the data, or the point of view of the map (i.e. road, touristic or military maps). For instance, in a given context, a small polygon can be enlarged if it represents an isolated urban area, but can be deleted if it represents a piece of land. Thus, land-use generalization processes require specific rules for each land-use class, according to the geographical context of the objects and the goal of the output map.

Several land-use generalization models have been proposed in the literature. We can specially differentiate:

- Models guided by the geometry, where for example small polygons are aggregated with bigger ones, or thin polygons are enlarged (van Oosterom, 1995)

- Models guided by the data schema, where two object classes are merged in the same class at a smaller scale, as proposed by Le Men (1996), Yaolin et al. (2003) or Revell (2007)

- Models guided by the functionnality, where new concepts are created by aggregation (Ruas & Lagrange, 1995). For instance, an airport is the aggregation of several buildings and other surfaces.

- Mixed approaches, where the model is guided by the geometry, the data schema and the functionnality (Galanda & Weibel, 2002; Haunert & Wolff, 2010).

To formalize land-use generalization in the ScaleMaster2.0, we follow a mixed approach. Indeed, in the context of MRDB, the data schema of a land-use coverage can be modified between two databases representing the

same real world entities with their own level of details. In the meantime, polygonal geometries are generally simplified (by aggregation and using simplification algorithms) or even collapsed to other geometries (for instance river areas become river lines under a given threshold).

The following section presents how land-use generalization is handled using this mixed approach in the ScaleMaster 2.0.

## 3. A framework to generalize land-use coverages in the ScaleMaster2.0

This section presents the general framework proposed to handle multi-scale land-use generalization processes in the ScaleMaster2.0 model. It describes the method developed to perform land-use coverages generalization, the strategy proposed to manage land-use generalization on large areas, and finally its formalization in the ScaleMaster 2.0.

### 3.1. Land-use coverages generalization method

This section presents the proposed method to simplify a land-use coverage at a given scale. As exposed before, we consider that each land-use class requires a minimum area to be conserved. Then, different methods are proposed to eliminate holes whether they are inside the polygon, at the border of the polygon, or at the border of the map area. Finally, a line simplification is performed using the Douglas & Peucker (1973) filtering algorithm.

The method starts by processing each land-use class separately, in order to remove polygons smaller than the given threshold. Thus, after an optional union of merged land-use classes (in the case of a process guided by the data schema) and a preliminary union of neighboring polygons, small polygons completely located inside the coverage are removed according to the parameterized threshold (Figure 3). Small polygons located at the border of the coverage are preserved as they are potentially part of bigger polygon in the neighboring coverage (see section 3.2).
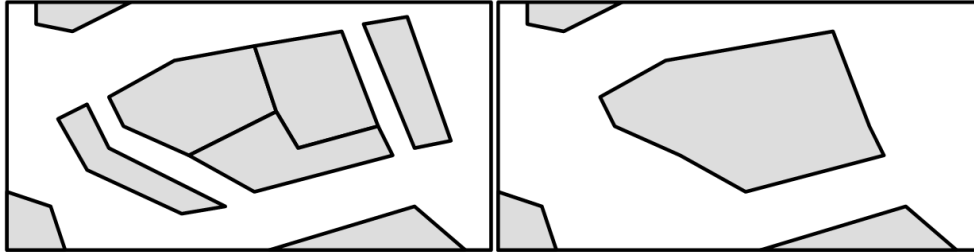
**Figure 3.** First-step of the method : Union of neighboring polygons in the same land-use class, and elimination of small polygons according to their size and location.

In the second step, the method merges all land-uses classes in a global coverage (converted in a topological map), and processes the distribution of the holes according to their characteristics (Figure 4). If the set of land-use classes doesn't provide a complete coverage of the area, an empty class is temporary created. Three categories of holes are differentiated and processed differently:

- Holes entirely located in a polygon are sealed and affected to the land-use class of the surrounding polygon (a)

- Holes located at the border between two polygons of different classes are divided in two parts (using the polygon's skeleton) and each part is distributed to the neighboring land-use classes (b)

- Holes located at the border between more than two polygons are divided in the corresponding number of neighboring polygons (using a partition computed with a triangulation around the polygon's centroid), and each part is distributed to the neighboring class sharing the longest border (c)
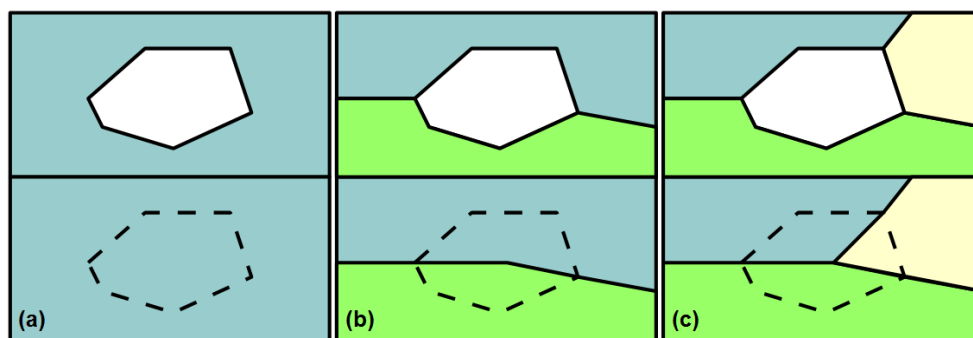


**Figure 4.** Second-step of the method: Three different strategies to eliminate holes in the coverage according to their neighboring land-use classes.

Once all holes are removed in the land-use coverage, the line simplification algorithm is applied to finalize generalization. At the moment, the Douglas & Peucker (1973) filtering algorithm is used in the model.

When land-use generalization is applied on large areas, the method presented in this section can be time-consuming, or can faces memory limitations for the processing. To avoid this problem, a partitioning method of the coverage is proposed in the following section.

## 3.2. Partitioning to handle generalization on large areas

In order to manage land-use generalization on large areas with the Scale-Master2.0, a partitioning of the coverage can be handled preliminarily. Different methods can be proposed to manage area partitioning, but they all face the problem of the consistency of the joining tiles.

Area partitioning can be achieved using different methods, as proposed in previous studies (Briat et al., 2011) and illustrated in the figure below: a delimitation layer (a), regular squares (b) or quad-trees (c).
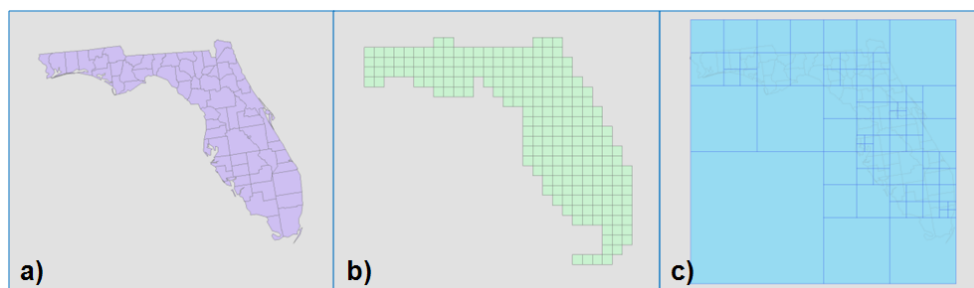


**Figure 5.** Three different methods for area partitioning, from Briat et al. (2011)

But other partitioning methods can be proposed using external data, or not. For instance, the Collagen model (Touya, 2010) proposes partitioning using delimitations which have a geographic signification (urban or suburban areas, mountainous, rural areas...). Other partitioning methods, based on the road network (Ruas and Plazanet, 1996) or on the road and river networks (Duchêne, 2012) have also been proposed. These examples show that no method can be considered as more efficient, as they all depend on the maximum number of objects tolerated in the memory, or computational time accepted to process an entire coverage.

The partitioning method using regular squares can be considered as the easiest to implement, but it supposes to define the appropriate size of the square in order to not extent the maximum number of tolerated object. The partitioning method using quad-trees can be relevant in this context, as it can guarantee a tolerated quantity of objects in each square, and a smaller number of them. But this method can be more difficult to handle, especially

in terms of consistency in the joining of land-use classes processed independently in each square.

Indeed, once the area partitioning and the processing of each tile is performed, the question of the consistency of the merge of land-use classes processed separately need to be resolved. To anticipate this problem, an extension of each square of the partitioning can be achieved, as proposed by Briat et al. (2011) and illustrated in figure 6, to generate a double covering of objects located at the border and facilitate their joining with neighboring objects. But the question of the parameterization of this extension needs to be resolved.
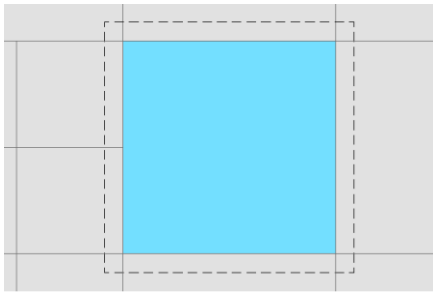


**Figure 6.** Extension of a partitioning area, from Briat et al. (2011)

Another method proposes to identify preliminarily the objects located at the border of the area, and performs specific methods a posteriori to manage the consistency between neighboring objects processed separately in the different tiles.

In the framework proposed to generalize land-use coverage in the Scale-Master2.0 model, the partitioning using regular squares was chosen, because of its facility to implement. The question of the consistency of the joining of neighboring land-use classes is still an ongoing problem. Then, to manage this problem, we decided, as exposed in the previous section, to preserve the small polygons (potentially eliminated) located at the border of the tile, as we consider they can be part of a bigger area, larger than the minimum area threshold. This solution supposes the development of a posteriori methods to manage these neighboring polygons, but these methods have not been implemented yet in the model.

### 3.3. Formalization in the ScaleMaster 2.0

The land-use generalization algorithm and the partitioning method have been integrated in the CartAGen library (Renard et al, 2010), in order to be used by the ScaleMaster2.0 model. To formalize their parametrization in the model, the ScaleMaster.xml and Parameters.xml file interfaces have

been adapted to manage the specific generalization rules of each land-use class and the partitioning of the coverage.

The developed method formalizes for each land-use class two parameters: a merging option to another class (to simplify the data schema) and a minimal area threshold (to eliminate small polygons). These parameters need to be defined for each scale range of the ScaleMaster.xml file if generalization has to be performed at different scales. An additional threshold for line simplification needs to be defined for the entire land-use coverage (all land-use classes are simplified the same way).

In Figure 7, two land-use classes are presented with their own land-use simplification parameters at the scale interval 1:50k-1:150k. In this example, lake areas smaller than 100.000 sq. meters and built-up areas smaller than 200.000 sq meters are eliminated. The line simplification parameter is specified for lake areas, but is also applied to all the land-use classes.
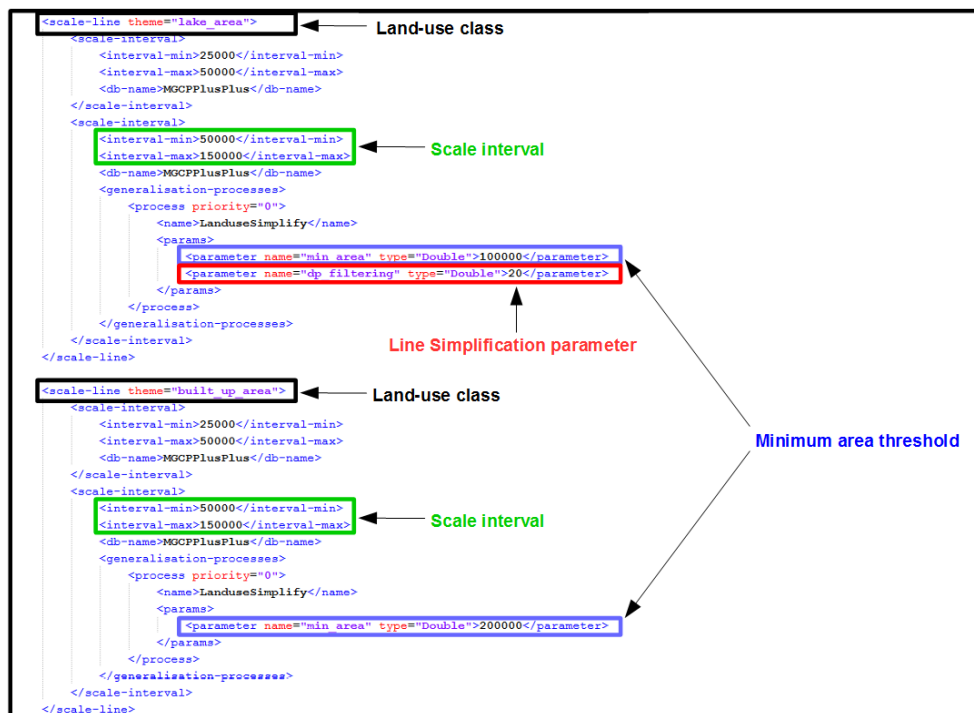


**Figure 7.** Formalization of generalization rules for each land-use class in the ScaleMaster.xml file.

To manage land-use generalization on large areas, the area partitioning option using regular squares is integrated in the Parameters.xml file. In Figure 8, the generalization is performed at the scale 1:100k, using a parti-

tioning of the area with regular squares of 0.2 decimal degrees (in the WGS84 coordinate system).



```
<?xml version="1.0" encoding="UTF-8"?>
<PEA-REP-generalisation-DV1-parametres>
    <echelles-sortie>
        <echelle>100000</echelle>                          ◄────────── Output scale
    </echelles-sortie>
    <BDs-entree>
        <BD>
            <nom>MGCPPlusPlus</nom>
            <chemin>C:\ScaleMaster\Data\Senegal\MGCP\</chemin>
            <partition-size>0.2</partition-size>           ◄────────── Square size
        </BD>
    </BDs-entree>
</PEA-REP-generalisation-DV1-parametres>
```

**Figure 8.** Parameterization of the area partitioning option, using regular squares, in the Parameters.xml file.

Finally, on a functional point of view, the developed land-use generalization process follows the steps: 1. Partition the land-use coverage in regular squares (if needed), 2. Merge the classes (if needed), 3. Convert the land-use coverage into a topological map, 4. Eliminate the small objects, 5. Apply line simplification algorithms. The next section presents an experiment of the model.

## 4. Experiments

The developed land-use generalization method was experimented on two different case studies: a land-use coverage in the area of Dakar (Senegal) and a map of sedimentary floor off the coast of Brittany (France).

The first experiment is performed on a land-use coverage composed of 12 classes in the area of Dakar, extracted from a spatial dataset with a reference scale of 1:50k. The output derived scales are the 1:100k and 1:200k with a line simplification parameterized respectively at 20 and 30 meters using the Douglas-Peucker algorithm. Minimal area thresholds used for each land-use class are defined in Table 1.

Results of land-use generalization in the area of Dakar at the scale 1:100k are shwon in Figure 9. These results show that the formalization of appropriate threshold for minimal area and line simplification can provide interesting results in the constitution of cartographic continuums, with acceptable time computation by the use of partitioning options. Nevertheless, the definition of the generalization rules parameters remains a difficult task, according to the dataset processed, its geographical context, and the point of view of the map.

| | Minimal area (in square meters) | |
|---|---|---|
| | 1:100k (dp=20m.) | 1:200k (dp=30m.) |
| River area | 100.000 | 200.000 |
| Lake area | 100.000 | 200.000 |
| Built up area | 200.000 | 300.000 |
| Quarry area | 300.000 | 400.000 |
| Wooded area | 300.000 | 400.000 |
| Grassland area | 300.000 | 400.000 |
| Crop area | 300.000 | 400.000 |
| Reservoir area | 100.000 | 200.000 |
| Thicket area | 300.000 | 500.000 |
| Marsh area | 300.000 | 500.000 |
| Soil surface area | 300.000 | 500.000 |
| Tidal water area | 300.000 | 500.000 |

**Table 1.** Minimum area for each land-use class generalized at different scales in the area of Dakar.
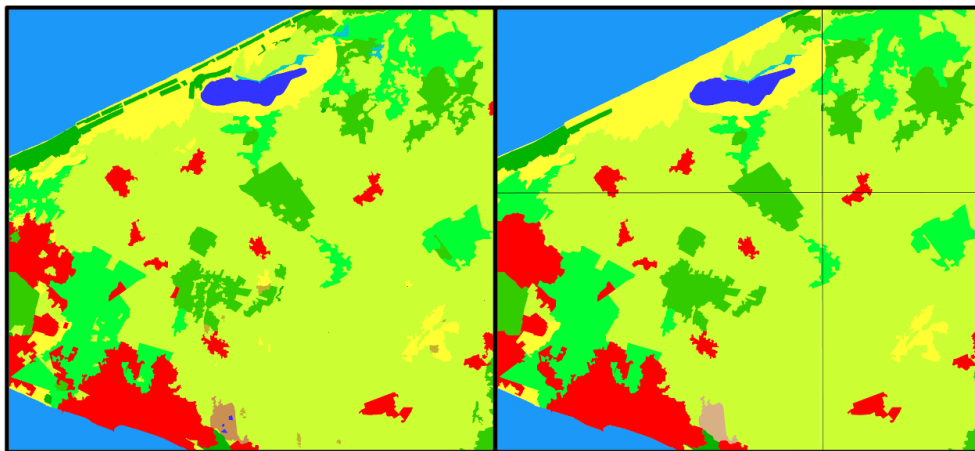


**Figure 9.** Original dataset (left) and derived land-use coverage at the scale 1:100k (right) in the area of Dakar (with the partitions limits displayed).

To illustrate a generalization process on another type of categorical data, the second experiment was performed using a sea floor dataset off the coast of Brittany (1:50k scale). The output derived scales are the 1:100k and 1:200k with a line simplification parameterized respectively at 20 and 30 meters using the Douglas-Peucker algorithm. Minimal area thresholds and merging options used for each sedimentology class are defined in Table 2.

| | Minimal area (in square meters) | |
|---|---|---|
| | 1:100k (dp=20m.) | 1:200k (dp=30m.) |
| Rock area | 100.000 | 200.000 |
| Pebble area | 200.000 | 400.000 |
| Pebble gravel area | 200.000 | Merged with Pebble area |
| Gravel area | 200.000 | 400.000 |
| Sand area | 200.000 | 400.000 |
| Fine Sand area | 200.000 | Merged with Sand area |
| Sand Gravel area | 200.000 | Merged with Sand area |
| Land area | 50.000 | 100.000 |

**Table 2.** Minimum area and filtering thresholds used for the generalization of sedimentology maps if the board of Brittany.

Results of the experiment on the dataset of sedimentology floor off the coast of Brittany are exposed in the figure below. We can see that using appropriates generalization rules on the geometries and in the data schema, combined with appropriate symbols, smooth changes can be created between intermediate representations in order to create a consistent cartographic continuum. It shows that a generalization system based on a ScaleMaster 2.0 is able to obtain results close to tGAP continuous representations of categorical data (van Oosterom et al, 2014).
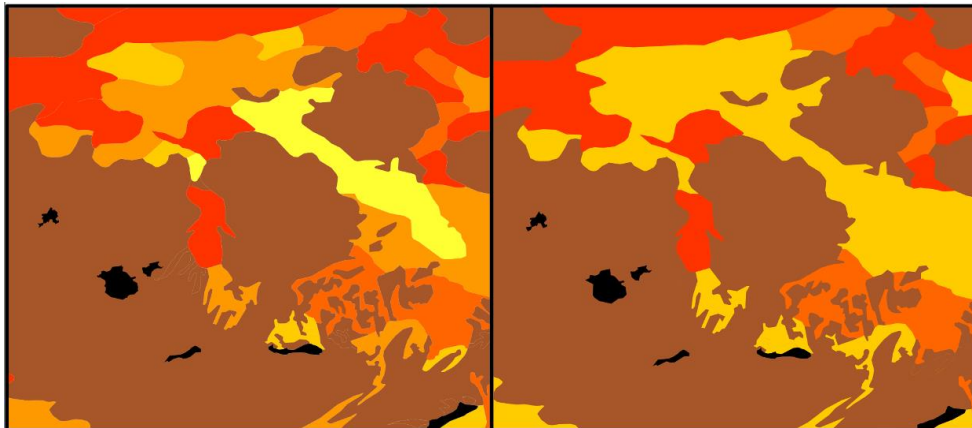


**Figure 10.** Results of the generalization of the sedimentology map of the coast of Brittany at the scales 1:100k (left) and 1:200 (right).

## 5. Conclusion and further work

This article presents a general framework to manage multi-scales land-use generalization in the ScaleMaster2.0 model. The proposed framework can be seen as a preliminary contribution in order to formalize derivation rules of categorical data at intermediate scales, in order to create cartographic continuums. The two experiments performed show interesting results to derive intermediate land-use coverages from a MRDB, with acceptable time computation by the use of partitioning options. Nevertheless, several improvements should be provided concerning the consistency between neighboring partitions of the generalized land-use coverages, using post-processing operations already proposed in this paper, but which have not been implemented yet in the model. Moreover, the integration of other generalization operations (e.g. amplification of polygons, or skeletonization) already implemented in the ScaleMaster2.0 should be added in the general process of land-use generalization. Finally, the preservation of the consistency with lower level of details, as evoked by Girres & Touya (2014), or with other generalized themes (road or river networks for instance) needs to be extended to categorical data in order to provide consistent cartographic continuums of land-use coverages.

## References

Brewer CA, Buttenfield B (2007) Framing guidelines for Multi-Scale map Design using databases at multiple resolutions. Cartography and Geographic Information Science 34(1):3-15

Briat MO, Monnot JL, Punt E (2011) Scalability of contextual generalization processing using partitioning and parallelization. 14th Workshop of the ICA commission on generalization and Multiple Representation, 2011 Paris

Douglas D, Peucker T (1973) Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. The Canadian Cartographer, 10 (2), 112–122.

Duchêne C, Ruas A, Cambier C (2012) The CartACom model: transforming cartographic features into communicating agents for cartographic generalization. International Journal of Geographical Information Science vol. 26 (9), pp. 1533-1562.

Galanda M, Weibel R (2002) An agent-based framework for polygonal subdivision generalization. In D. Richardson and P. van Oosterom (Eds.), Advances in spatial data handling, 10th international symposium on spatial data handling, pp. 121-136. Springer.

Girres JF, Touya G (2014) Cartographic generalization Aware of Multiple Representations, Proceedings of the 8th International Conference on Geographic Information Science , Vienna, Austria

Haunert JH, Wolff A (2010) Area aggregation in map generalization by mixed-integer programming. International Journal of Geographical Information Science 24(12):1871–1897

Le Men, H (1996) Généralisation cartographique pour l'occupation du sol : application au passage v1 à v2 pour la BD carto. Revue Internationale de Géomatique 6 (2-3), 227-248.

Renard J, Gaffuri J, Duchêne C (2010) Capitalisation problem in research - example of a new platform for generalization: CartAGen. Proceedings of 12th ICA Workshop on generalization and Multiple Representation, Zürich, Switzerland

Revell P (2007) Automated generalization and representation of ordnance survey polygonal landcover data at 1:10 000 scale. In Proceedings of ICC2007, the International Cartography Association Conference.

Ruas A, Lagrange JP (1995) Data and knowledge modelling for generalization. In J. C. Müller, J.-P. Lagrange, and R. Weibel (Eds.), GIS and generalization: methodology and practice, pp. 73-90. London: Taylor & Francis.

Ruas A, Plazanet C (1996) Strategies for automated map generalization, 7th International Symposium on Spatial Data Handling (SDH'96), Delft (Netherlands), 12-16 August, pp 319-336

Touya G, Girres JF (2013) ScaleMaster 2.0: a ScaleMaster Extension to Monitor Automatic Multi-Scales Generalizations. Cartography and Geographic Information Science 40(3):192–200

Touya G, Girres JF (2014) ScaleMaster 2.0: Generalising Unusual Map Themes from OpenStreetMap, Proceedings of 17th ICA Workshop on generalization and Multiple Representation.

Touya G (2011) Le modèle Collagen. Collaboration de processus automatiques pour la généralisation cartographique des paysages hétérogènes. Thèse de doctorat, Université Paris-Est.

van Oosterom P (1995) The GAP-tree, an approach to 'on the fly' map generalization of an area partitioning. In: Müller JC, Lagrange JP, Weibel R(Eds.), GIS and Generalization: Methodology and Practise, pp. 120–132. London: Taylor & Francis.

van Oosterom P, Meijers M, Stoter J, Šuba R (2014) Data structures for continuous generalisation: tGAP and SSC. In D. Burghardt, C. Duchêne, and W. Mackaness (Eds.), Abstracting Geographic Information in a Data Rich World. Springer International Publishing, pp. 83-117.

Yaolin L, Molenaar M, Ai T (2003) Categorical database generalization aided by data model. In Proceedings of the 21st International Cartographic Conference : Cartographic Renaissance, ICC 2003, Durban, South Africa, pp. 2308-2318. ICA.