Conception of a GIS platform to simulate urban densification based on the analysis of topographic data

Anne Ruas, Julien Perret, Florence Curie, Anabelle Boffet-Mas, Anne Puissant, Grzegorz Skupinski, Dominique Badariotti, Christiane Weber, Pierre Gancarski, N. Lachiche, et al.

GeOpenSim :

Conception of a GIS-Platform to simulate urban densification based on the analysis of topographic data.

Authors : Anne Ruas; Julien Perret; Florence Curie; Annabelle Mas; Anne Puissant; Gregorz Skupinski, Dominique Badariotti; Christiane Weber ; Pierre Gancarski; Nicolas Lachiche ; Julien Lesbegueries ; Agnès Braud

1- Introduction

In this paper we present the current state of a research project named GeOpenSim funded by the French research funding agency (ANR) from December 2007 to March 2011. The aim of GeOpenSim is to build an OpenSource GIS platform to analyze the evolution of urbanization and to simulate it on specific areas. Compared to other researches on urban growth and simulation (Couclelis, 1997; Batty, 2007), we do not use cellular automata but vector topographic agents (such as buildings, streets and urban blocks) and we use historical data to build evolution rules. In our research, simulation should be considered as a powerful tool to build and to consolidate our own understanding on urban morphological evolution. It is not a predictive tool. At final state the GeOpenSim platform will be under an Open Source license to allow geographers to set and test their own evolution rules on real topographic data.

Practically we wish to see whether and how we can develop models and methods to analyze the landscape at different dates and to derive rules of evolution from these analyses. The platform is meant to be open to different criteria according to the richness of our data and knowledge. For example it should be able to simulate building densification with or without planning rules, such as the fact that during a certain time period this urban block can only be devoted to housing. Moreover, each simulation should generate a new state described by the parameters of this simulation. By the way, these states can be compared to facilitate again our visual analysis and understanding. During the project and for practical reasons, our experiments were limited to two different parts of France (around Orleans and around Strasbourg), from 1955 to 2008. Of course experiments can be performed on different regions with different dates provided ~~that~~ where vector topographic data are available. We have also limited our experiments on building and street densification and reconstruction because of time constraints (see below). In the paper we only present the building densification process.

The research team is composed of researchers from COGIT laboratory and university of Orleans, specialized on the analysis of vector topographic data and agent modelling, researchers from the geographical department of the University of Strasbourg specialized in urbanization process and researchers from the computer science department of the University of Strasbourg, specialized in machine learning techniques.

The present paper is structured in the following way:

- In section 2 we present how a simulation works : which input data are used, which functions are used to densify the space and how the simulation works, is tuned and run,
- In section 3 we present the densification method for each urban block illustrated with results,
- In section 4 we present the method used during the project to build the required knowledge for simulation,

- In section 5 we conclude and present the main research perspectives.

2- How does simulation work?

2.1 The global process

The main process is what we call 'densification or reconstruction'. We name densification the action of adding buildings (or streets) on an area that may or may not already contain buildings (or streets). We name reconstruction, the action of removing the existing buildings, redefining urban blocks boundaries and then proceeding to a densification on the new free space.

The simulation in itself follows a predefined sequence that consists in choosing the area where the simulation will take place, the initial and final dates of the simulation period, the rule bases used to perform the simulation, the kind of building and street patterns that will be used during the simulation. Then, the system uses its rule bases and some spatial analysis to define which areas will change, with what kind of patterns and up to which level of densification. Last, the simulation runs and the result is stored in the system with the parameters of the simulation (see figure 1). Of course the process requires that information already exist in the platform: the initial vector database at a specific date, some rules of evolution and patterns to densify the space.

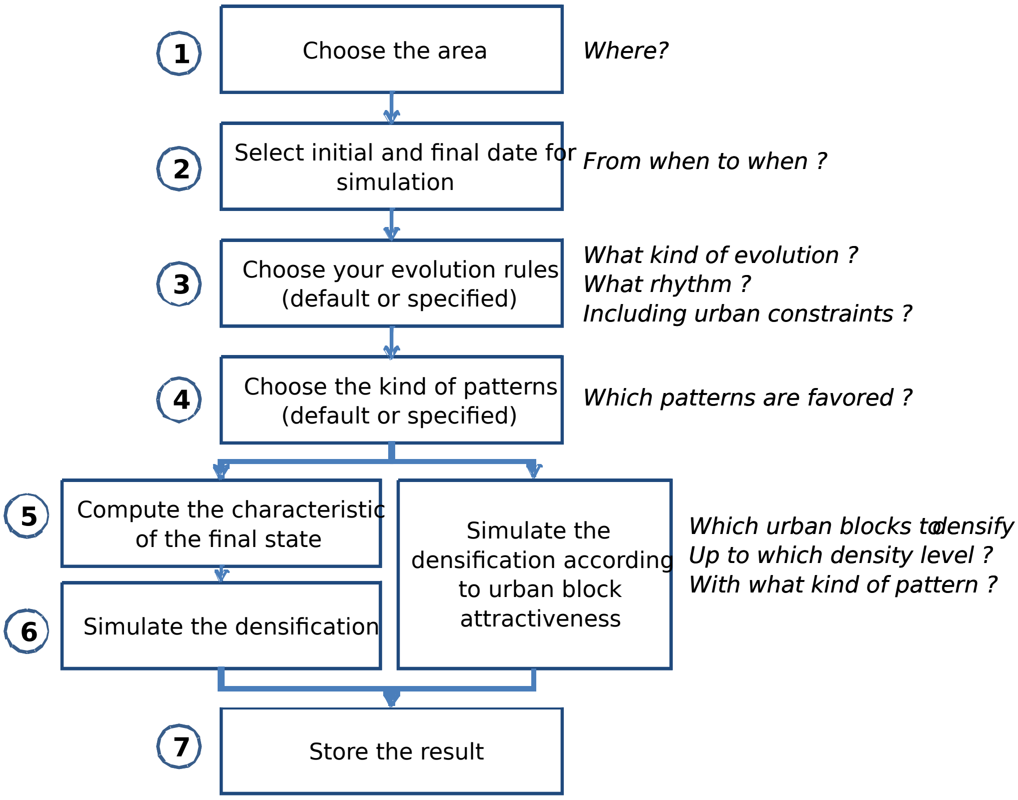

Figure 1: Sequences to run a simulation

Thus a simulation follows 4 macro-steps: the spatio-temporal specifications (steps 1 and 2 in figure 1); the selection of the parameters of the simulation (steps 3 and 4 in figure 1); the simulation itself (steps 5 and 6 in figure 1) and the storage of the results for comparison purposes (step 7 in figure 1). Two strategies of simulation are presented in the following subsection 2.2.

The quality of the simulation depends on three factors:

- our capacity to analyze the entire urban area on which we make simulation in order to detect the best candidates (i.e. the best urban blocks) to densify or to reconstruct (figure 2),
- the quality of the methods of densification that add streets and buildings not only onto free spaces but also on urban blocks that already have urban properties (see section 3),
- our capacity to build relevant rules of evolution (see section 4).

In a defined urban area composed of a set of urban blocks from this date to this date which urban blocks are good candidate to change?

In case of densification (add buildings to existing buildings)
- with same kind of buildings ?
- with another kind of buildings ? Which one ?
- up to which density ?

In case of reconstruction
- where might be the new urban block limit ?
- what kind of buildings ? Mixed or uniform types ?
- up to which density ?

Figure 2: Where and how to densify an urban area?

2.2 The simulation based on agent modelling

We have chosen to use the agent paradigm to control the simulation process. As for generalisation purposes, an agent is a geographical object that acts to reach its goals. It can be either meso - urban blocks – or micro – buildings (Ruas, 1999). We test two different strategies to densify an urban area:

- The first strategy is composed of two steps: We use a rule base to identify which urban blocks are good candidates to change and up to which density (step 5 in figure 1), then we use an agent based paradigm in order to let each urban block densify itself according to its characteristics and to its goals (step 6 in figure 1)
- The second strategy (steps 5 and 6, right side in figure 1) is under conception. It tries to follow an entire agent paradigm: each urban block has an *attractiveness index* to define itself as a good or bad candidate for densification. If it is a good candidate, it then follows the same process than the other strategy.

For both strategies an urban block agent uses its own method 1/ to characterize itself, 2/ to compare its properties to its goals properties 3/ to select an appropriate method to densify or to reconstruct itself according to its properties and its goals 4/ to densify or reconstruct itself and 5/ to trigger building agents for spatial readjustment purposes. A building agent, when it is activated, modifies its location (distance and orientation) towards its neighbouring buildings and street. The result is different for each block for three reasons: 1/ each urban block has different goals (box 5 figure 1) 2/ the densification method integrates the properties of each urban block 3/ the densification method integrates some random decisions to ensure a certain diversity.

It appears that the difficulty is neither to build an agent based approach nor to build a method of densification (see section 3) but to identify relevant knowledge to distribute evolution goals over a large set of urban blocks. This is due to the fact that urban growth depends on many factors including some for which we do not have any information. We thus have to propose methods that mix urban properties and unknown factors transformed into probability. This aspect is developed in section 4.

3- The method to densify urban blocks

As previously stated, the densification methods are essential. The principle is 1/to choose a kind of building and 2/ to locate a set of them in a more or less regular way (building distribution). Of course as town planner, the location of the new buildings integrates the shape of the urban block and the location of the already existing buildings. Basically two strategies exist and are proposed. The first one consists in extending the existing building pattern up to the density goal. To do so, the system analyses the properties of the existing buildings and reproduces them. The second strategy consists in adding a specified new pattern in the urban block. To do so we developed a library of patterns. Each pattern describes the properties of the buildings and their spatial distribution.

3.1 Building a relevant pattern library and method for building location

Our goal at this step is to build a generic pattern schema with relevant attributes that can describe a large diversity of real patterns, a schema that is used to automatically add buildings on an existing space. The method developed to locate the new buildings into an existing urban block is described in (Curie et al, 2010). We first identified a draft data schema that contains attributes that describe building shape, size and distribution and we tried to describe a set of well known building patterns with this schema. We made simulations of different patterns, we showed the results to geographers specialized in town planning and we improved gradually our patterns according to their comments. The data schema today used to describe patterns is presented in figure 3. We defined a classical library of building shapes (square, rectangle, U, L, T, stairs, etc.) following different research such as Mackaness and Rainsford (2002), and in order to obtain better visual results we set that a building pattern may be composed of two different kinds of shapes with different frequencies (figure 3). For the spatial distribution described by the *distance to the road* and *distance between buildings* the values can be fixed, or can follow a Gaussian law or can be randomly chosen between the *minimum* and *maximum* value.

Name of the pattern
Classical Time frame
Distance to the road : minimum value; maximum value; kind of distribution
Distance between buildings : min; max; kind of distribution
Orientation of the building : towards the closest street; towards close buildings
Type of building 1
  Shape (rectangle; U; L; T; square; stairs; ..)
  area : min; max; kind of distribution
  elongation [0,1]
  width
  frequency 100% or X% if type of building 2 exists
Type of building 2
  Shape (rectangle; U; L; T; square; stairs; ..)
  area : min; max; kind of distribution
  elongation [0,1]
  width
  frequency (100-X)%

Figure 3: the data schema of building pattern

During the project we proposed three types of patterns: individual houses, collective buildings and industrial buildings. More precisely we set 6 kinds of patterns because of their frequencies in the French landscape between 1950 to today. These patterns have been defined by means of a visual analysis of existing maps and databases and improved by means of geographers' expert knowledge and knowledge from literature (Lacoste 1963; Panerai et al, 2001; Gauthiez, 2003; Allain, 2004; Arnaud, 2008). The process of conception is iterative: patterns were proposed and used to simulate urban densification or reconstruction; results were shown to our experts and comments were used to improve the patterns. The 6 patterns conceived are the following:

- two patterns for individual houses:
  - **Individual and not planned houses** (called spontaneous) often complete an existing house pattern. Houses are mostly built along streets or road network
  - **Individual planed houses** are built during a limited time frame according to a house planning program which may include new access streets. Houses look the same, their locations are organised in a regular way along the streets. The point is to maximise the number of houses.
- three patterns for **blocks of flats** (collective buildings) :
  - the **Large-BF** were very frequently built in France during the 50^{ies} , these buildings are rectangles, from 300 to 900m², with high elongation value (between 0.1 and 03).
  - The **VeryLarge-BF** corresponds to important housing programs in the 60ies in France during an important housing crisis. These buildings are composed of an important quantity of lodgements (> 1000). A set of buildings are built together, clustered in the centre of an urban block. Each building is either a rectangle with important elongation and with an area up to 800 or even 1000m², or square, smaller (between 300 to 400m²) and very tall.
  - The **small-BF** appears in France in the 80ies and is still used today in housing programs. Buildings are smaller (between 300 to 1000m²) not so high (often between 3 or 4 floors) and shapes are more original, including stairs.

- One **industrial pattern** composed of large buildings, with compact shapes. The diversity of shapes is very important and as well as the rhythm of densification: some densifications are slow and regular while others are very quick.

The method to add buildings in an area is described in (Curie et al, 2010). To sum up for individual houses, houses are added one by one along the street network. For Large or Very large blocks of flats, the process seeks free space according to building size and shape. The shape, size and position of each house or building is computed from the parameters of each pattern (figure 3) including a range of values and a method (Gaussian law or Random) to compute the values (for example the size) inside this range.

3.2 Experimental results and comments

The densification has been tested on real datasets to evaluate and improve the quality of the pattern and of the object location. During the project we created historical databases at different dates on different areas. The method to create these historical databases has been presented in (Perret et al, 2009). To sum up, we used recent topographic databases, we copy cut them in the past, and we used scanned maps or orthophotographies to 'down-date' the data. As the main process is urban growth, we mainly removed houses, buildings and streets and created few of them. This historical dataset is essential to build evolution rules and to test our simulation. In figure 4 we simulate a house densification within an urban block from 1976 to 1989 and we compare it to the reality. This example illustrates the fact that we still have to be able to add dead-ends within urban blocks.

Figure 4: Experimental results for the densification with planned individual houses

Figure 5 illustrates a densification of a heterogeneous urban block with non planned individual houses. In reality the houses built in 1989 close to the isolated house in 1956 took the same orientation than this house whereas in our simulation, the new houses follow the orientation of the street.

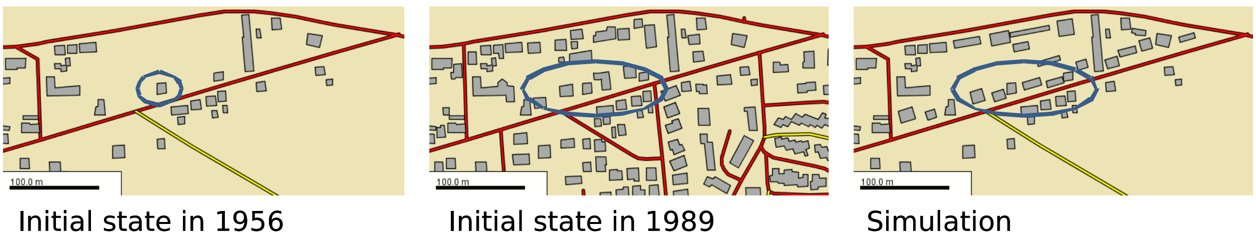

Figure 5: Experimental results for the densification of an urban block with spontaneous housing

Reality in 1989 (titre à changer dans l'image du milieu)

Of course the simulation is never meant to look like the reality but each difference is full of meaning either on the reality or on the difficulty of producing relevant results.

4- Building relevant knowledge from historical data

From the beginning of the project the aim was to build relevant knowledge related to urban growth from existing data but also to propose relevant methods for building knowledge on other areas, on other time frames.

We know that urban evolution depends on a large set of factors including time period, duration, existing patterns, urban plans, land readjustments and many other elements such as population growth. The first assumption of our study is that if there is a process of reconstruction and densification within an area, even if we do not know all informations, the densification is never random and that some areas are more likely to change than others. For example if the density of an urban block is very high, changes are rather unlikely, but some buildings might be reconstructed. On the other hand, the best candidates for densification are empty or nearly empty areas, close to dense or semi-dense areas. Discussions with geographers and urban planners enlighten the fact that the densification or reconstruction depends on the type of urban blocks and on the location of the blocks with respect to the city center and the rural area. In order to build rules such as "If an urban block that has *such and such properties* might change in *this way* during *this* time period"; we decided first to learn how to classify automatically urban blocks according to a predefined classification (section 3.1). The second step is then to learn evolution rules such as 1-"*this type* of urban blocks, with *such and such properties*, are good candidates for densification" which answers to the question *which urban blocks are good candidates for densification*? and 2-"this type of urban blocks tends to change in such a way" which answers to the question *How to densify an area ? With what kind of pattern?*

4.1 Automated urban block classification based on supervised learning techniques

Urban classification is often produced by image analysis with visual interpretation and field checking. It is of course a very good but time consuming process. Alternatively image processing can be used to reduce the production cost (Donnay et al, 2001; Blaschke et al., 2004). Supervised learning techniques can be used for that purposes (Bauer et al., 2001). In our case, we have to classify urban blocks at different dates, for which we do not have necessary images and obviously we cannot perform field completion. We have thus chosen to classify urban blocks from morphological properties computed from vector representation only. This method has already been tested for generalisation purposes in (Boffet, 2001).

The aim of our approach is to develop a classification method useful to study morphological evolution process, which can be easily performed on different areas, for different time frames. Thus we decided to use supervised learning techniques, to generate decision rules for classification and beforehand to provide a module of data labelling necessary to generate learning examples and that will be used to provide new classification rules on demand (Lesbegueries et al, 2009).

The first step of our research has been to visually analyse urban blocks and to define a classification that would be relevant for our research subject. This classification has to distinguish individual and collective buildings as well as the density since we are looking for the good candidates for reconstruction or densification.

Several meetings and many discussions between geographers were necessary to reach a complex compromise as the only real validation could only occur after the evaluation of simulations on large datasets. Our current classification is our ‘working classification’. It will have to be validated in the future. Seven classes have been defined:

Class 1-Continuous urban fabric (city center),

Class 2-Discontinuous urban fabric with individual houses,

Class 3-Discontinuous urban fabric with housing blocks (blocks of flats),

Class 4-Mixed urban fabric (including individual housing (type 2) and housing blocks (type 3)),

Class 5-Mixed areas (including types 1, 2, 3 and 6),

Class 6-High density of specialised areas (including industrial, commercial, hospital or scholar buildings),

Class 7-Low density of specialised areas (containing few or no buildings).

For learning purposes, datasets have been manually classified by geographers with the data labelling module developed during the project. The labelling process was very interesting to define each class and its boundary. Of course some urban blocks are always in between two classes. Because the project focused on suburban areas and the learning process requires a number of samples large enough for each class, the class 1 was not studied.

In a parallel way, we proposed a large set of measures on buildings and urban blocks. Classification tests showed that the most relevant measures were the following:

- Area, elongation and convexity of the buildings;
- Building orientation and relative orientation of a building towards its closest street;
- Number of buildings, total area of buildings, density of buildings inside an urban block.

The classification algorithm used is TILDE (Blockeel and De Raedt, 1998). It is a supervised and symbolic algorithm that produces decision trees. It has a better expressivity than C4.5 (Quinlan 1995). Several tests have been applied on Strasbourg area on historical data over the fifty last years. To assess the classification we performed cross validation and we also classify new and unused samples. The quantitative evaluation is between 75 and 80 %. The results depend on the quantity of samples used per class and the type of class, knowing that some classes such as individual housing have much more samples than very mixed areas (class 5).

The decision tree allows detecting rules such as:

- **If** the block density < 0.05 **Then** the urban block is in class 7
- **If** density > 0.05 and number of buildings (whose area < 105m²) < 3 and number of buildings (whose area < 185m²) < 1 and average-building area < 1205m² **Then** the urban block is in class 3.

This analysis allows finding relevant measures at the same time as threshold values. The results were also visually analysed to provide a qualitative evaluation. What we considered as the best rules according to quantitative and qualitative evaluation were applied on different areas and again visually evaluated (figure 6).

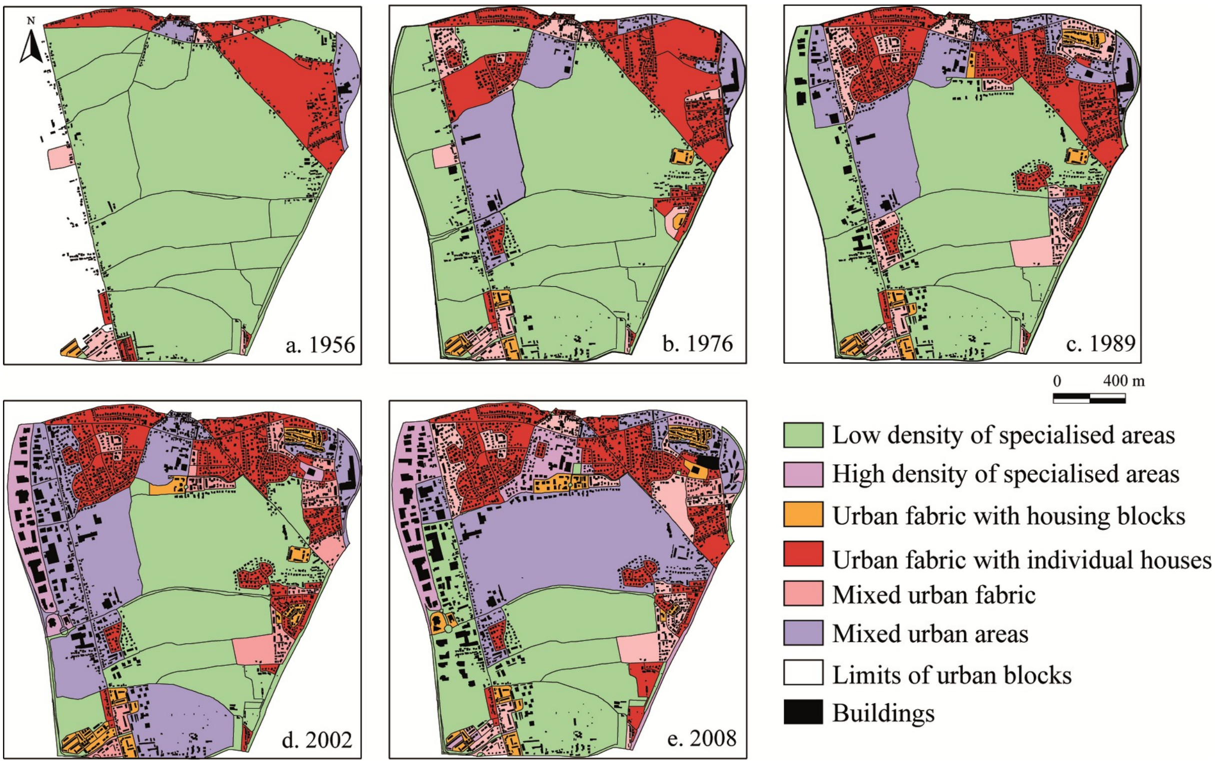

Figure 6: Urban blocks classification applied to 5 dates

Even imperfect, this classification is very interesting as it allows us to see urban evolution through a new perspective. It also allows studying the transition and stability of urban blocks along time period (section 4.2).

Other criteria such as the distance to the city centre or the fact that an urban block looks like its neighbouring blocks are under evaluation to improve the rules.

4.2 Learning transitions

In order to make the simulations as realistic as possible, we need statistics about, on the one hand, the transitions themselves and on the other hand, about the most frequent kinds of urban block evolutions. With this second information, we can study each kind of evolution in order to define it according to characteristics of urban blocks involved in the evolution. These analyses will help identify the best candidates to change depending on the nature of urban blocks and their previous states.

To do both analyses we have developed a new tool, named iVisualize (see Figure 7), that allows:

- to calculate the number of occurrences of each type of transitions (e.g., “Low density of specialised areas” to “High density of specialised areas”) or no transition (e.g., “Road” remains “Road”) according to a given length (i.e., number of dates taken into account) and a minimum threshold,
- to highlight classes of evolutions using learning algorithms.

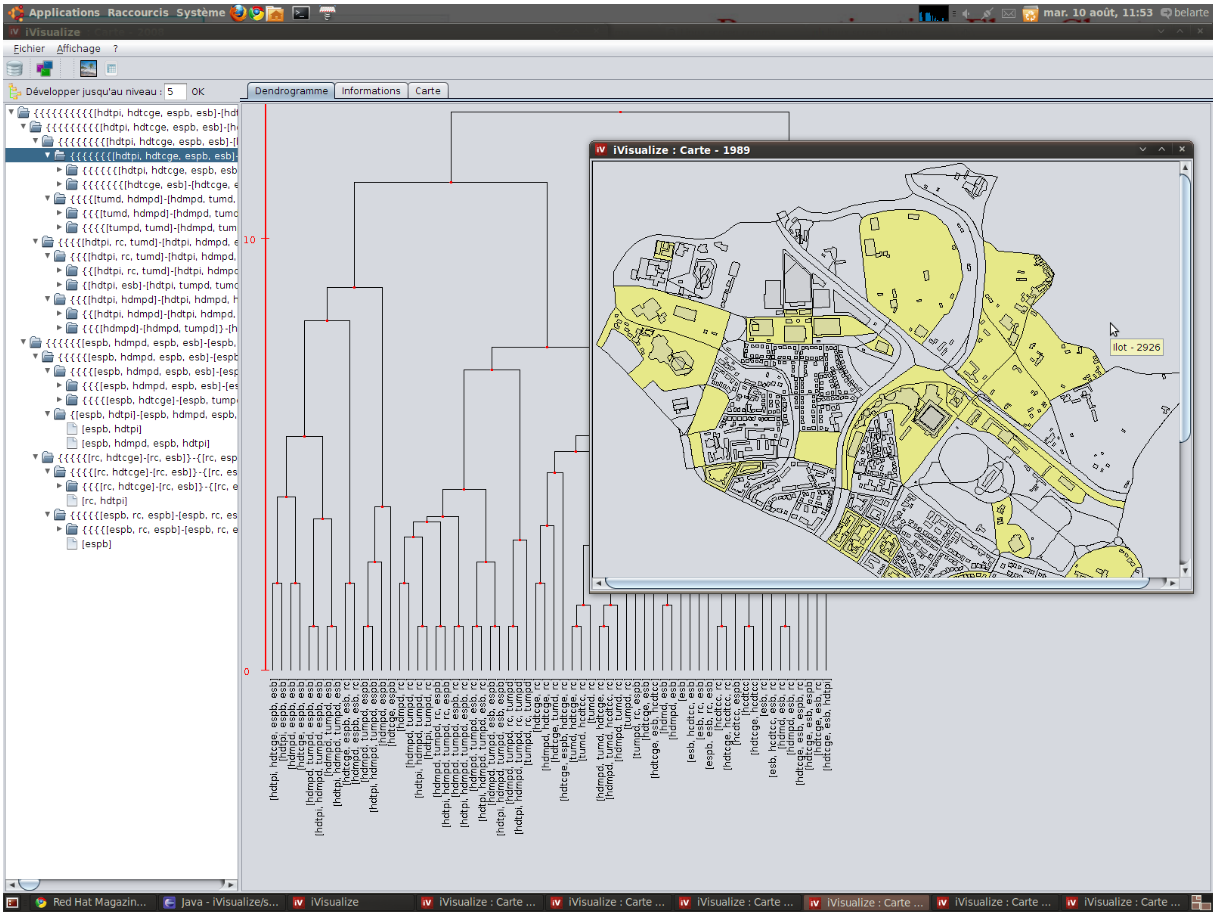

Figure 7: Urban blocks transitions and evolutions analysis tool

While the first aspect has not been a particular problem, the second one requires choosing the type of learning algorithms to be used.

It is very difficult and time consuming to define all examples needed for supervised algorithms. In fact, we choose to use unsupervised approaches and in particular distance-based algorithms (i.e., the Kmeans algorithm and the hierarchical agglomerative one) to highlight classes of evolutions. To apply these algorithms, in agreement with the geographers, we defined a *new semantic distance* between classes of urban blocks (from **0** for two objects from a same type **to 4** for a "Continuous urban fabric" object and a "Low density of specialized areas" object, for example). Finally, to apply the Kmeans algorithm, we use the similarity measure called dynamic time warping (DTW) proposed by (Sakoe et al., 1971) and (Sakoe et al., 1978) and the global averaging method for dynamic time warping (DBA) introduced in (Petitjean et al., 2010).

The first experiments carried out have produced promising results. The clusters of evolutions extracted by the algorithms have been studied by the experts and seem to correspond to thematic evolutions. These experiments have also enabled the experts to validate the semantic distance.

5- Conclusion and future work

Developing a GIS platform to study and to simulate urban evolution is very challenging. During this three years long research project, many results have been obtained. We developed:

- a data model to represent historical topographic data with explicit link between objects at different times (Perret et al, 2009),
- a module to create historical data from today topographic data base and old maps or orthophotographies (Perret et al, 2009),
- an agent engine that uses rules and densification methods to simulate urban growth, with the registration of each simulation for comparison purposes (section 2),
- several methods of densification that add buildings on free or semi-free areas with different building patterns (section 3),
- a building pattern library that can easily be enriched (section 3),
- a module to classify urban blocks based on supervised learning techniques associated with a labelling module to classify new blocks with new measures (section 4.1),
- a first set of evolution rules that focuses on the transitions of urban classification through time (section 4.2),
- a first evaluation module that compares states.

If the construction of relevant knowledge is a real difficulty, the results are improving and the association of different techniques for knowledge acquisition is promising. Current research aims at completing and improving the evolution rules by means of expert knowledge and statistics. We also intend to integrate urban planned constraints to study different kinds of urban evolution. A first module will be available in Mach 2011 on the GeOxygène opensource plateform. It should allow to enlarge the community of researchers testing their own urban evolution rules.

References

Allain R., *Morphologie urbaine - Géographie, aménagement et architecture de la ville*, Armand Colin, 2004.

Arnaud J.-L., *Analyse spatiale, cartographie et histoire urbaine*, Parenthèses, 2008 ;

Batty, M.: *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-Based Models, and Fractals.* MIT Press, 584 pages (2007)

Bauer T., Steinnocher K. (2001): *Per parcel land use classification in urban areas applying a rule-based technique*. In: GeoBIT/GIS 6(2001), pp. 24-27.

Blaschke, T., Burnett, C., Pekkarinen, A., 2004. *New contextual approaches using image segmentation for object-based classification*. In: De Meer, F., de Jong, S. (Eds.), Remote Sensing Image Analysis: Including the spatial domain. Kluver Academic Publishers, Dordrecht, pp. 211-236.

Blockeel H., De Raedt L., 1998, *Top-down induction of first-order logical decision trees*, Artificial Intelligence, vol. 101, n°1-2, 12p.

Boffet A., 2001, *Méthode de création d'informations multi-niveaux pour la généralisation cartographique de l'urbain*, http://recherche.ign.fr/labos/cogit/telechargementCOGIT.php - PhD Thesis, University of Marne la Vallée, France

Couclelis, H.: *From cellular automata to urban models: new principles for model development and implementation.* Environment and Planning B: Planning and Design 24 (2), 165-174 (1997)

Curie F., Perret J;, Ruas A., 2010, *Simulation of urban blocks densification* , proceedings of the 13th AGILE International Conference on Geographic Information Science, Guimarães, Portugal.

Donnay, J.P., Barnsley, M.J., Longley, P.A., 2001. *Remote Sensing and Urban Analysis*. Taylor and Francis, London and New York., 315 p.

Gauthiez B., *Espace urbain : vocabulaire et morphologie*, Editions du patrimoine, 2003.

Lacoste Y., *Un problème complexe et débattu : les grands ensembles* , Bulletin de l'association des géographes français, , no 318-319, 1963.

Lesbuegueries J., Lachiche N., Braud A., Puissant A., Skupinski G., Perret J., 2009, *A platform for spatial data labeling in an urban context*, International Opensource Geospatial Research Symposium, 11 p.

Mackaness, W. A., and Rainsford, D., (2002) *Template matching in support of generalisation of rural buildings* in: Richardson, D., and van Oosterom, P., (eds) : Advances in Spatial Data Handling. Berlin: Springer, pp. 137-152

Panerai P., Castex J., Depaule J., *Formes urbaines : de l'îlot à la barre*, Parenthèses, 2001.

Perret J., Ruas A., Boffet-Mas A., 2009, *Understanding urban dynamics: the use of vector topographic database and the creation of spatiotemporal databases*, International Cartographic Conference - ICC 2009, Santiago, Chile, November 2009, 11 p.

Petitjean F. , Ketterlin A. and P. Gançarski, 2010, *A global averaging method for dynamic time warping, with applications to clustering,* Pattern Recognition, doi:10.1016/j.patcog.2010.09.013 (2010) to appear.

Quinlan J. R., 1995, C4.5 : *Programs for machine Learning*, Morgan Kauffmann.

Ruas A., 1999 *Modèle de généralisation de données géographiques à base de contraintes et d'autonomie*. http://recherche.ign.fr/labos/cogit/telechargementCOGIT.php - PhD Thesis, University of Marne la Vallée, France

Sakoe H. and S. Chiba, 1971, *A dynamic programming approach to continuous speech recognition*, Proceedings of the Seventh International Congress on Acoustics vol. 3 (1971), pp. 65–69.

Sakoe H. and S. Chiba, 1978, *Dynamic programming algorithm optimization for spoken word recognition*, IEEE Transactions on Acoustics, Speech and Signal Processing 26 (1) (1978), pp. 43–49.