Simulation of Urban Density Scenario according to the Cadastral Map using K-Means unsupervised classification

Mojtaba Eslahi, Mateo Baira, Rani El Meouche

Research Institute, ESTP, 28 Avenue du Président Wilson, F-94230, Cachan, France
{meslahi, mateo.baira, relmeouche}@estp.fr

Keywords: Urban densification, Urban modelling and simulation, K-means classification, Cadastral Map, SLEUTH model.

Abstract

Due to the increasing trend of urbanization over the last few decades, city development and planning have become a major concern for the authorities. Many efforts have been made in the field of urban growth modelling by researchers and urban planners to investigate the factors and effects of urbanization and urban growth. The work presented here is part of a project that simulates urban densification scenarios. This work describes the automatic detection and classification of neighbourhoods based on vector data of buildings and parcels. During the urbanization process, an empty parcel is divided into sub-parcels to accommodate new buildings. In general, the sub-parcels created have the same characteristics (surface area, building height, etc.) as the already urbanized parcels in their vicinity. Here, the aim is to detect and identify the different neighbourhoods in the study area to determine the characteristics of the new parcels. This is done in three steps: first, classifying the urbanized parcels, then identifying the different neighbourhoods based on this classification, and finally dividing the parcels into sub-parcels with features depending on the surrounding neighbourhoods. The model developed here is designed to automatically identify different urban zones and simulate the division of parcels over time according to a variety of urban density scenarios. This model is also compatible with the creation of new buildings based on these levels of density.

1. Introduction

Urbanization and especially urban growth are the major concerns of urban planners and city authorities. A great deal of research has been carried out in the field of urban growth simulation, and various urban models have been developed. This research aims to develop a method that allows the automatic simulation of urban density in compliance with the cadastral plan. The method we created makes it possible to automatically determine different neighbourhoods of a city, then simulate the division of parcels over time based on different urban density scenarios, and finally create new buildings resulting from these scenarios. This work can be used to study the impact of different urbanization policies on urban expansion for each study area. The project is an extension of our previous efforts on urban growth simulation methods.

In urban growth simulations, each simulation can be configured using evolution techniques from automatic analysis of data. These techniques try to complete the transition rules of cellular automata (Curie et al. 2011). Here, we can consider the neighbourhood as a factor to differentiate and classify the cadastre. Some research has been carried out in this field, focusing on the definition of different urban sprawl scenarios and the creation of an urban growth simulation tool which respects the constraints of various scenarios. These scenarios correspond to restrictive urban expansion policies that have been verified by applying them to different study areas at several scales (Eslahi, 2019; Eslahi et al. 2019; El Meouche et al. 2021).

To carry out these simulations, we used SLEUTH which is a cellular automata simulation model (Clarke, 2008; Jantz et al. 2010; Project Gigalopolis, 2018). The SLEUTH model produces prospective results of urban growth from historical data. The term SLEUTH comes from its input maps including the Slope, Land use, Excluded area, Urban, Transport and Hillshade. The results of SLEUTH are binary raster maps representing areas where urbanization is expected to occur. These results do not directly represent the buildings that are going to be constructed, but the areas where they will appear.

In this research, we have proposed the methodology to modify the building creation model by integrating the cadastre. This work aims to consider the cadastre and to modify it to incorporate the evolution of the parcels over time when creating new buildings (Stoter and Salzmann, 2003; Oosterom et al. 2005).

To integrate the cadastre, we cannot simply cut out the urbanization pixels by the existing parcels because most of the time urbanization occurs on parcels with a very large area like agricultural or forest parcels which must be divided to be able to create buildings. The division of these parcels depends on the type of building we want to build. A single-dwelling parcel will be smaller than a collective building one. The shape and the surface of the parcels are homogeneous according to the neighbourhoods. Therefore, the first step of our work involves the identification of different neighbourhoods defined by the geometry of buildings and plots. We carry out an automatic classification process for the parcels by type of urbanization that can be applied to the different study areas.

Using a machine learning algorithm, here we only looking for unsupervised classification methodologies that make it possible to identify and classify neighbourhoods based on k-means classification. Unsupervised K-Means classification generally involves calculating initial class means that are distributed regularly across the data space, and iteratively clustering pixels into the nearest class by applying the minimum distance

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-4-W4-2024-57-2024

technique. Every iteration of the algorithm is performed to recalculate the class averages and reclassify the pixels according to the new averages (P. Sinaga and S. Yang, 2020).

After classifying urban parcels and identifying different neighbourhoods, a type of land use has been assigned to non-urban parcels. This is done by observing the neighbourhood of the parcel while respecting the overall probability of the appearance of a type of parcel considering our urban sprawl scenarios. Here, an agricultural parcel close to a residential area will have a higher chance of being subdivided into single dwellings.

Once the parcels to be developed have been assigned a classification, they need to be divided so that they comply with the characteristics of the type of neighbourhood assigned to them (surface area, density, etc.). Having determined that the parcels are appropriate, we can create the new buildings according to the neighbourhood characteristics and then perform 3D modelling of the results.

To visualize the results, we have created a 3D representation of the future buildings from the urbanization forecasting raster. Buildings are created by applying erosion to pixels, and then according to the type of neighbouring buildings of the pixel and the probability of occurrence of each building type in the studied area, a height is assigned to them.

The work carried out can be used as a basis for studying the impact of different urban development policies on city growth for each of the analysed zones.

In the next section, we present the study area and data used in our project. In section 3, the methodology of our developed model is defined. The results are described in section 4 and the model evaluation is done in section 5. Finally, the work is concluded in section 6.

2. Study areas and data used

We applied our developed model to different case studies with different scales to evaluate its efficiency. We have found this model is applicable to any study area. Here, we present the results of the model on the Saint-Sulpice-La-Pointe study area which is a small city located northeast of Toulouse in France (see Figure 1). Over the past 20 years, this small town has recorded significant population growth and, as a result, considerable growth in terms of urban area.

As discussed earlier the SLEUTH urban growth model is used to create the forecasting urban growth maps. A geospatial database and GIS are applied to create the input maps for the simulations. For SLEUTH modelling, slope and hillshade maps are generated from the RGE ALTI digital elevation model (DEM) provided by IGN (French National Institute for Geographic and Forestry Information). The urban areas, the excluded zones, land use and transport network maps are produced using BD TOPO and BD ORTHO (GeoServices IGN, 2020). And finally, the population and census data for the district area are extracted from the INSEE database (INSEE, 2016). The vector data used here are as follows:

- BD TOPO: Industrial building, undifferentiated building, cemetery, sports ground, vegetation zone, hydrographic surface, aerodrome, aerodrome runway, primary road, secondary road, station, railway, activity area.
- BD Parcellaire: Parcels

The SLEUTH model uses historical maps of urban planning, transport, and land use to calibrate the model. Here, historical maps from 2000 to 2017 have been used for modelling the prospective urban maps of 2050 via different scenarios.

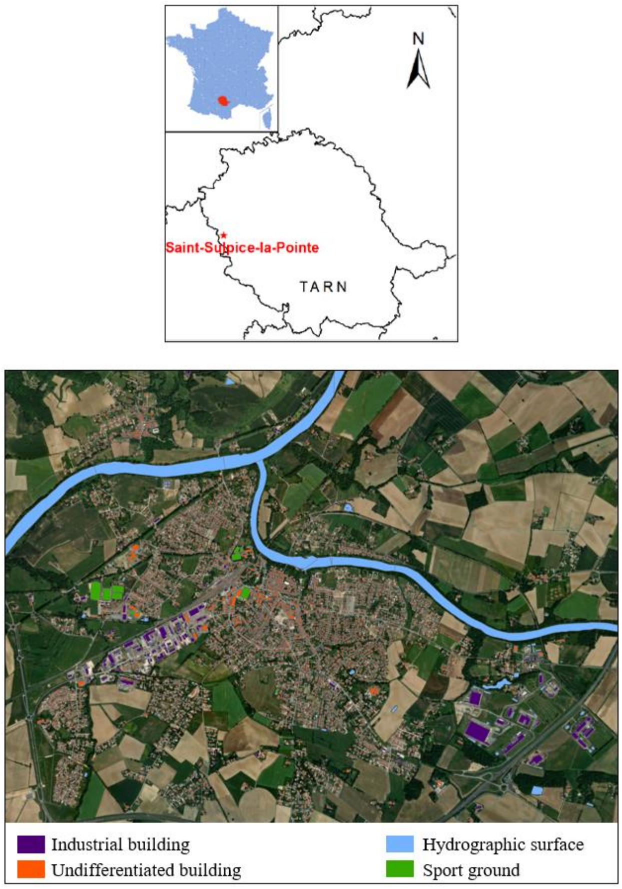

Figure 1. Location and extent of Saint-Sulpice-La-Pointe study area, source: World Imagery Esri

3. Methodology

The SLEUTH model used here is the version developed by us, which has more parameters than the basic SLEUTH model, including population growth and urban fabric scenarios. These scenarios provide a better overview of how the simulated urban area could be used, and how residents could be accommodated in the new simulated zones. The scenarios are defined based on building classification and on-site observation of current buildings. The scenarios are based on the combination of different building types, such as single dwellings, low-rise buildings, and high-rise buildings as follows:

- The low-density urban growth scenario represents 50% single-dwelling units and 50% high-rise buildings.
- The moderate-densification urban growth scenario consists of 30% single-dwelling units and 70% high-rise buildings.
- The high-density urban growth scenario involving three building types represents 45% single-dwelling units, 45% low-rise buildings and 10% high-rise buildings.

For this project, we used SLEUTH results comprising 2D raster data. The methodology carried out in this research consists of three main steps including the detection and classification of neighbourhoods, determining zones for urban development, and parcel division and building creation.

Figure 2 illustrates the process of our model. As mentioned, our 2D urban growth simulation leads to the acquisition of new urban plots. Finally, the three steps described above allow us to position the building and create a 3D representation of the prospective city.

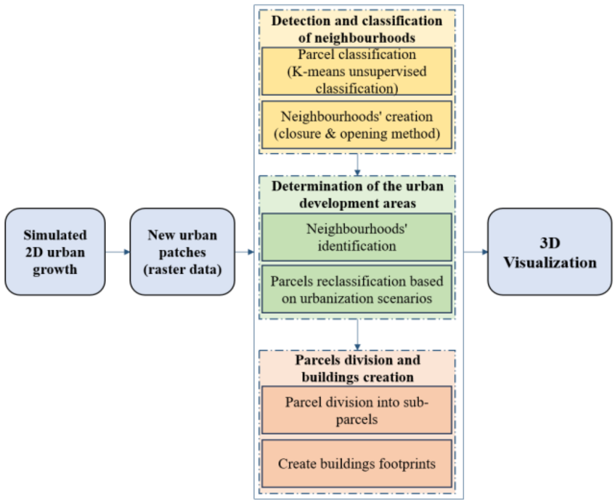

Figure 2. Model process

3.1 Detection and classification of neighbourhoods

To better understand neighbourhood detection and classification, we first need to define what a neighbourhood is in the context of our study. In the literature, the neighbourhood is defined as "a part of the city having certain characteristics or a certain unit" (Larousse Dictionary, 2009). Others define the neighbourhood as an entity made up of connected or complementary islets, an urban entity with very strong cognitive significance, but often blurred physical boundaries (Boffet, 2003). Here, we define the neighbourhood as a homogeneous spatial unit having a length and a width greater than a hundred meters.

To simulate urban expansion by creating new buildings, we first need to know the distribution of different types of urbanization over the study area. The first step in this work is to identify different types of neighbourhoods from vector data. These neighbourhoods will then allow us to determine the characteristics of the new buildings resulting from the urbanization process (Long and Kergomard, 2005; Puissant et al. 2011; Bernabé et al. 2013).

One of the main constraints of this project is developing a model in which all processing can be carried out automatically in any study area. We are therefore only looking for unsupervised classification methodologies that make it possible to identify and classify neighbourhoods without ground truth. That is why we propose a clustering method (k-means).

The k-means algorithm cannot determine the ideal number of classes for a study area, so we must calculate it upstream. To determine this, we perform a series of partitioning for an increasing number of partitions, and each time we calculate the partitioning quality index. This index increases as the number of partitions increases. However, after a certain threshold, it increases less significantly, which means that the new classes do not allow better separation of the dataset (see Figure 3). We will then use this threshold as the number of classes in our study area.

Figure 3. Quality of the score as a function of the number of classes for Saint-Sulpice-La-Pointe

Once the parcels are classified, we must group the parcels of the same class to form neighbourhoods. The method of creating neighbourhoods developed here is based on the use of buffers. To do this, we use the dilation-erosion method by applying two buffers with different signs for the distance values. The goal is to train buffers from parcels of the same class and then remove all buffers formed too small to represent a neighbourhood. The operations are broken down into two steps:

- Closure (a buffer zone with a positive distance value followed by a buffer zone with a negative distance value): A closure on parcels of the same class makes it possible to create zones representing a continuous presence of a certain type of parcel. For each class, the mean of the distance of the nearest neighbour to the centroids of the parcels of the same class is calculated. This distance is then used as the closing distance for each class of parcel.
- Opening (a buffer zone with a negative distance value followed by a buffer zone with a positive distance value): An opening eliminates areas resulting from the closure that do not meet the dimensions of a quarter (minimum 100 meters in length and width). So here a 50-meter opening is used.

3.2 Determination of the Urban Development Areas

Having identified the various neighbourhoods that will determine the characteristics of future buildings in the simulation, the next step is to identify the parcels in which urbanization will take place. To do this, we use the results of SLEUTH, which produces raster data that represent urbanized areas in pixel form. These pixels have different sizes depending on the study areas and sometimes they can intersect places where there is no possibility of urbanization such as roads or existing buildings. It is therefore necessary to process these pixels of urbanization to keep only those areas where urbanization is possible.

Various types of constraints, both natural and artificial, can represent an obstacle to the development of a building's construction. Here, we consider the following constraints:

- Existing buildings including industrial, undifferentiated, and remarkable buildings.

- Roads and rail infrastructure including highways, stations, railways, and marshalling areas.
- Areas of activity
- Cemeteries
- Sports grounds
- Natural constraints: rivers, lakes, ponds, vegetation

We should neither build on these limitations nor build too close to them. Therefore, for each type of constraint, we create a non-buildable buffer zone. The unbuildable distance is determined according to the proximity of the undifferentiated building to these constraints. For each constraint class, we calculate the median distance to the nearest building and this distance is then applied as the non-buildable buffer distance around all features of that constraint type. Water and vegetation surfaces of more than one hectare are processed differently. They are considered to have their distribution from the buildings around them. Then we calculate the average distance between them and their neighbouring buildings (buildings within a radius of 10 times the distance from the nearest building).

After creating different non-buildable zones, they are merged and all portions of urban pixels that intersect these non-buildable zones are removed. All parcels that intersect filtered urbanization pixels are intended to be urbanized, which means that the buildings must be built there. Some of these parcels have no classes because they do not contain any buildings. Therefore, we determine their class based on the surrounding neighbourhoods.

In addition, our methodology allows us to simulate the creation of buildings for different urbanization scenarios. Thus, we can define the types of buildings that we want to appear in certain proportions. Therefore, all the parcels intersecting urbanization pixels are reclassified so that the distribution of the parcels respects the desired urbanization scenario.

The parcel reclassification methodology is divided into three steps:

- Calculation of the number of parcels to be allocated to each type of neighbourhood, based on densification scenarios. Denote this number "N_i" with "i" the index of a type of neighbourhood.
- For each parcel, we calculate the probability of being reclassified according to each type of neighbourhood. This probability is calculated as a function of the distance between the parcel and the corresponding neighbourhood. The closer it is to a neighbourhood, the more likely it is to get its class.
- For each type of neighbourhood (from the smallest "N_i" to the largest), we select the "N_i" parcels with the highest probability of belonging to that type of neighbourhood, and then we assign them that class.

To give a new classification to the parcel types, we start by calculating their probability of belonging to the surrounding neighbourhoods. A probability of "P_i" for each neighbourhood is calculated on every parcel (where "i" is the index of a type of neighbourhood). To calculate this probability, we perform the following steps:

- For each parcel, we calculate the distance "D_i" that separates the parcel from the different neighbourhoods.
- Parcels that intersect a neighbourhood have a probability of one, and the rest have a probability of zero.
- If a parcel does not intersect any neighbourhood, we calculate the inverse of the distance separating it from each type of neighbourhood. The probability of a parcel belonging to a neighbourhood type is then the inverse of its distance to the neighbourhood divided by the sum of the inverse distances.

Therefore, the probability of belonging to a type of neighbourhood increases with proximity to that type of neighbourhood. The probability of parcels being classified as "industrial zones" for Saint-Sulpice-La-Pointe is shown in Figure 4.

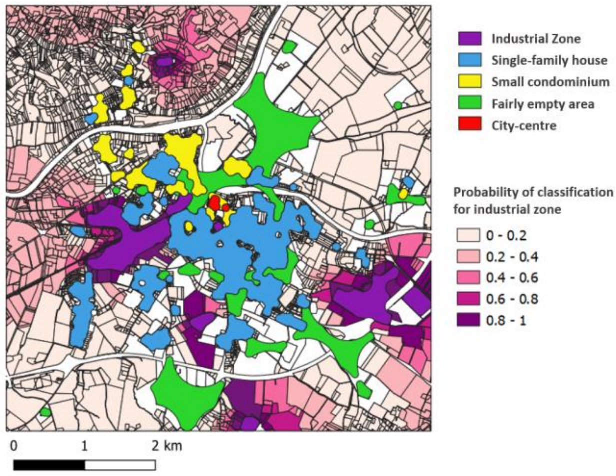

Figure 4. Probability of parcel reclassification as industrial zone

As illustrated in Figure 4, the parcels that are closer to industrial zones are more likely to become industrial than others. However, we can't simply assign each parcel to the class with the highest probability. Indeed, we must respect the overall probabilities of class appearance.

3.3 Parcels division and buildings creation

Once the class of parcels to be urbanized has been determined, one step remains to make the representation of buildings' footprints, i.e. parcel division. In this step, each parcel is divided into sub-parcels, and then building footprints are made in each sub-parcel that intersect the simulated pixels.

To divide the parcels, we use ArcGIS Pro's "Subdivide Polygon" data management tool. The subdivision method used to divide polygons is called equal area. In this method, polygons are divided into a specified number of parts of a certain area and a remaining part. The Subdivision type is based on stacked blocks where polygons are divided into stacked sets.

Indeed, when a new class is assigned to a parcel, it does not necessarily meet all the criteria of the class, particularly the area. For example, a large agricultural parcel that requires to be urbanized can be reclassified as a suburban parcel. The parcel can then be divided into sub-parcels of the area of a sub-parcel, and buildings can be placed in these sub-parcels. The division of the parcels may involve an overlap between two sub-parcels, with a building from the original parcel.

All the sub-parcels intersecting the same constructions are then merged and considered unbuildable. The divided parcels intersect urbanization pixels, but not all sub-parcels resulting from the division intersect urbanization pixels. We therefore only construct new buildings in sub-parcels whose area of intersection with an urbanization pixel is greater than or equal to the area of a building of the class of the parcel.

Once the constructible sub-parcels have been identified, we can proceed with the creation of the building's footprints. Here we use the geometry of the sub-parcels to create the new buildings. The first idea was to erode the sub-parcel to achieve the desired building surface geometry, but this method creates buildings that are too elongated. To correct this excessive stretch, we used the following methodology:

- Dividing the sub-parcel into "mini parcels" with areas equal to twice the area of the desired building
- Select the mini parcel with the lowest elongation (closest to a square).
- Erosion of the selected mini parcel to reach the desired building surface.

This method makes it possible to obtain buildings with an increase in the length of BDTOPO buildings. However, there may still be a few buildings with excessively large elongations, so all buildings with an elongation greater than four are removed. 99% of the buildings in our study area in the BDTOPO database have an elongation of less than four.

4. Results

We used ArcGIS Pro to create the input maps of SLEUTH as well as the simulation of urban densification scenarios. The objective here is to develop a model that can identify and classify neighbourhoods based on vector data of buildings and parcels.

Figures 5 and 6 illustrate the classification obtained by applying our classification methodology to the Saint-Sulpice-La-Pointe study area, as well as the neighbourhoods obtained after the morphological dilation-erosion process, respectively. The K-means calculated here are based on the height of the building, the density of the built-up area, the surface of the parcels, the built-up area ratio and building elongation.

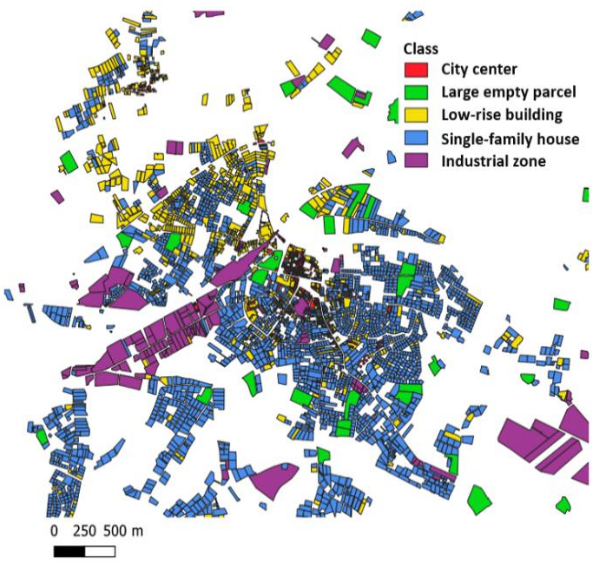

Figure 5. K-means classification of Saint-Sulpice-La-Pointe parcels for 5 classes.

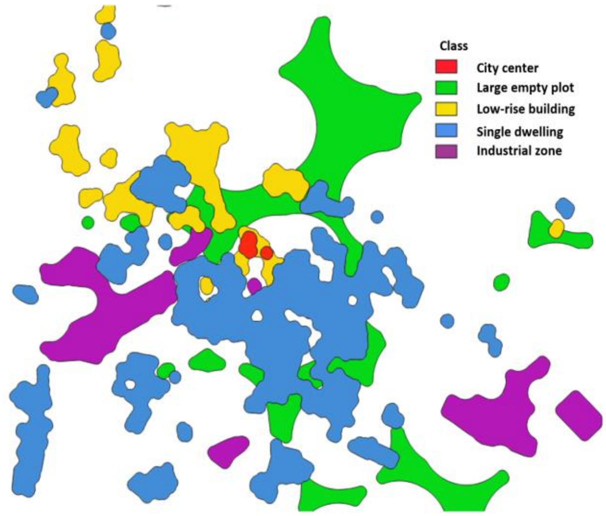

Figure 6. Neighbourhoods obtained for Saint-Sulpice-La-Pointe.

We observe that the different classes are spatially well separated. In addition, the obtained neighbourhoods correspond to them, which can be visually observed from the parcel classification.

Table 1 shows the characteristics of the five classes of parcels that we have identified.

Color	Qualification	Height (m)	Undifferentiated built area (%)	Industrial built area (%)	Surface (m²)
Yellow	Low-rise dwelling	7.2	17	0	1 159
Red	City center	7	75.8	0	185
Green	Large & quite empty parcel	5.6	5.7	0	11 702
Blue	Single dwelling	4.2	17.7	0	1 122
Purple	Industrial area	5.7	3.4	20.9	6 350

Table 1. Characteristics of the different classes of parcels identified for Saint-Sulpice-La-Pointe.

As can be seen, the parcels are well separated, and each class corresponds to a specific combination of built height and built area density that can easily be given meaning.

The buildings obtained from our model have adequate shapes. Their average elongation for Saint-Sulpice-La-Pointe is 1.60 versus 1.64 for the buildings of the BDTOPO provided by IGN. Figure 7 shows the footprints of buildings created for sub-parcels.

Figure 7. Buildings created from mini-parcels

Figures 8 to 10 illustrate the 3D representation of prospective buildings obtained for an area of Saint-Sulpice-La-Pointe according to different densification scenarios for the year 2050. These figures illustrate three types of buildings including the single dwelling in blue with an average height of 4.4 meters, low-rise houses in green with a height of 6 meters and high-rise buildings in red with 7.5 meters in height.

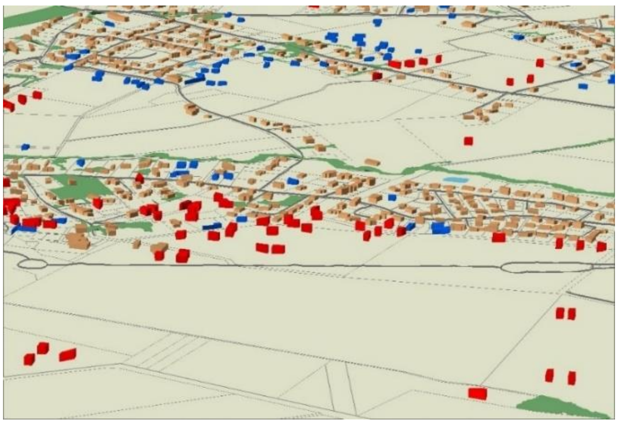

Figure 8. 3D representation of prospective low-density urban growth scenario (50% single dwelling, 50% high-rise buildings)

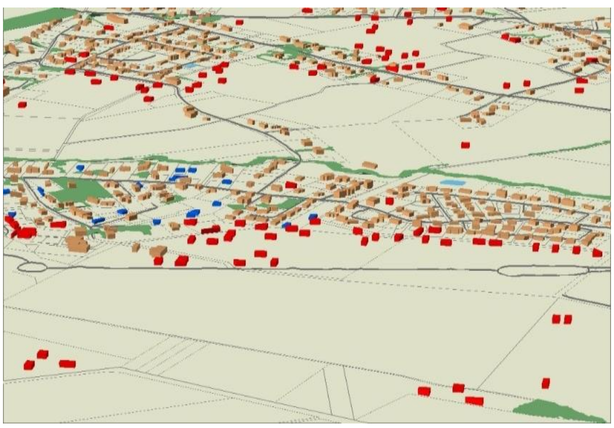

Figure 9. 3D representation of prospective moderate-density urban growth scenario (30% single dwelling, 70% high-rise buildings)

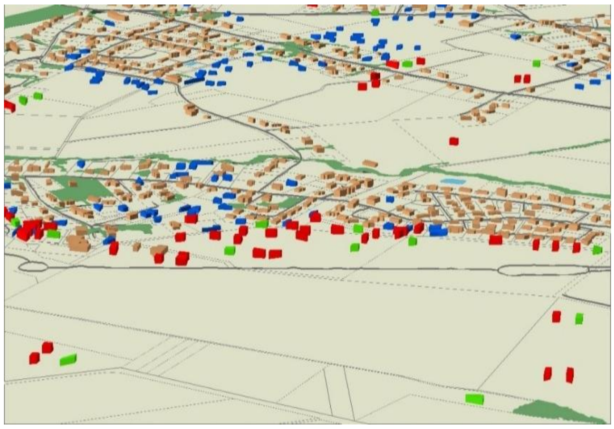

Figure 10. 3D representation of prospective high-density urban growth scenario with 3 types of buildings (45% single dwelling, 45% low-rise buildings, 10% high-rise buildings).

5. Evaluation

The results obtained are acceptable for the identification and classification of neighbourhoods as well as for the creation of buildings. However, there is a slight gap between the proportions of desired building classes in the scenarios and the number of buildings obtained. This is explained by the fact that we use the scenarios to classify the parcels that will accommodate the new buildings. These parcels perfectly respect the planned scenarios but then we divide them into sub-parcels and the number of sub-parcels to be urbanized varies according to the area of the urbanization pixels in the original parcel and their arrangement. As a result, all the divided urbanization parcels do not give the same number of buildings, and this can lead to a slight deviation from the proportions of new buildings constructed with the densification scenarios, but, this situation is attenuated in larger study areas with more parcels to be urbanized. Table 2 shows the proportions of each class of new buildings obtained for Saint-Sulpice-La-Pointe.

Scenario	Expected low-rise buildings (%)	Obtained low-rise buildings (%)	Expected high-rise buildings (%)	Obtained high-rise buildings (%)	Expected single dwelling (%)	Obtained single dwelling (%)
Mixed	45	46.7	10	9.1	45	44.2
Low densification	50	52.2	-	-	50	47.8
Moderate densification	30	25.6	-	-	70	74.4
High densification	-	-	-	-	100	100

Table 2. Distribution of building classes according to the scenarios and actual buildings.

6. Conclusion

In this research, the goal was to develop a method that would allow the automatic simulation of urban densification by observing the cadastral layout. The developed model makes it possible to automatically determine the different neighbourhoods of a city, then simulate the division of parcels over time according to different urban densification scenarios, and finally create new buildings resulting from this densification. This work can be used to study the impact of different urbanization policies on urban sprawl for any study area.

To go further, it would be interesting to work with building data at a higher level of detail to be able to forecast not only the height and the area of new buildings but also more geometric parameters like the shape of their roofs, etc. This would have enabled us to present a more realistic simulation.

It would also be desirable to modify the algorithm for reclassifying the parcels according to the scenario so that the proportions of the classes of the new buildings correspond exactly to those predicted by the scenarios.

References

Bernabé, A., Calmet, I., Musy, M., Bocher, E., Andrieu, H., 2013: Classification au-tomatique des tissus urbains par la méthode des nuées dynamiques, 31e Rencontre de i'AUGC. French.

Boffet, A., 2003: Généralisation cartographique et description des espaces urbains (Cartographic generalisation and description of urban areas). Bulletin de l'Association de géographes français. French. 80. 207-219. 10.3406/bagf.2003.2328.

Clarke, K. C., 2008: A Decade of Cellular Urban Modeling with SLEUTH: Unresolved Is-sues and Problems. In Brail R.-K. (eds.), Planning Support Systems for Cities and Region, Lincoln Institute of Land Policy, Cambridge, MA, 47–60.

Curie, F., Mas, A., Perret, J., Puissant, A. & Ruas, A., 2011: *Simulation d'un processus de densification du tissu urbain à base d'agents, Revue Internationale de Géomatique*, French, vol. 21, n. 4, pp. 489-511, doi:10.3166/rig.21.489-511.

Dictionnaire Larousse, 2009: https://www.larousse.fr/

El Meouche, R., Eslahi, M., Ruas, A., Sammuneh, M.A., 2021: 3D Urban Growth Simulation Using Human Settlement Capacity SLEUTH Model (HSCS). In: Grueau, C., Laurini, R., Ragia, L. (eds) Geographical Information Systems Theory, Applications and Management. GISTAM 2020. Communications in Computer and Information Sci-ence, vol 1411. Springer, Cham. https://doi.org/10.1007/978-3-030-76374-9_4.

Eslahi, M., 2019: Urban growth simulations in order to represent the impacts of constructions and environmental constraints on urban sprawl. Modeling and Simulation. Doctoral Thesis, Paris-Est university, 2019. English. ⟨NNT: 2019PESC2063⟩. ⟨tel-02493929⟩.

Eslahi, M., El Meouche, R., Ruas, A., 2019: Using building types and demographic data to improve our understanding and use of urban sprawl simulation, Proc. Int. Cartogr. As-soc., 2, 28, https://doi.org/10.5194/ica-proc-2-28-2019.

GeoServices IGN, 2022: https://geoservices.ign.fr/

INSEE, 2016: https://www.insee.fr/fr/statistiques/2520034#consulter

Jantz, C.A., Goetz, S. J., Donato, D. and Claggett P., 2010: Designing and Implementing a Regional Urban Modeling System Using the SLEUTH Cellular Urban Model, Computers, Environment and Urban Systems, Vol. 34, No. 1, 2010, pp. 1-16. doi:10.1016/j.compenvurbsys.2009.08.003

Long, N., Kergomard, C., 2005: Classification morphologique du tissu urbain pour des applications climatologiques. Cas de Marseille. Revue internationale de géomatique. 15. 10.3166/rig.15.487-512. French.

Oosterom, P., Stoter, J., Lemmen, C., 2005: Modelling of 3D Cadastral Systems. 28th Cadastral seminar.

Project Gigalopolis, 2018: http://www.ncgia.ucsb.edu

Puissant, A., Lachiche, N., Skupinski, G., Braud, A., Perret, J., Mas, A., 2011 : Classification et évolution des tissus urbains à partir de données vectorielles. Revue internationale de géomatique. 21. 513-532. 10.3166/rig.21.513-532. French.

Sinaga, K. P., Yang, M. -S., 2020: Unsupervised K-Means Clustering Algorithm, in IEEE Access, vol. 8, pp. 80716-80727, doi: 10.1109/ACCESS.2020.2988796.

Stoter, J., Salzmann, M., 2003: Towards a 3D cadastre: where do cadastral needs and technical possibilities meet?, Computers, Environment and Urban Systems, Volume 27, Issue 4, 2003, Pages 395-410, ISSN 0198-9715,https://doi.org/10.1016/S0198-9715(02)00039-X.