ASTUS
Modelling urban sprawl with FORESIGHT and LUCSIM
Methodology report

Deliverable D.T2.2.4

November 2018

Contact of the Lead Partner
Auvergne-Rhône-Alps Région
Spatial planning and Montain Department
+33 4 26 73 46 67
astus@auvergnerhonealpes.fr

Table of contents

1 Introduction

The EU-funded ASTUS project for Alpine Smart Transport and Urbanism Strategies has been broken down into Work Packages (WPs) and divided up among various regional partners. WP2 consists of co-constructing low-carbon tools and developing decision aids for the pilot sites' projects and action, and making sure that the tools are transferable across Europe.

In the Auvergne-Rhône-Alpes region, ASTUS WP2 has set out to produce a CO2 minimiser tool as a transnational methodology for low-carbon scenarios, in cooperation with the TUM. There are also plans for Cerema Sud-Ouest to test urban sprawl modelling tools on the Thonon Conurbation and the Bauges Regional Nature Park areas.

For this more targeted initiative, the objectives are to:

- Develop a life-cycle costing approach to urban development, based on the costs generated by urban sprawl;
- Inform the development and mobility choices made by regional decision-makers by modelling the more or less long-term impact of their decisions on urban sprawl and land use;
- Test the software for modelling changes in land cover and land use on concrete examples, namely the Bauges Regional Nature Park and Thonon Conurbation areas.

This report describes the methodological aspect of the tests conducted and the feedback collected.

A report was also produced for each of the test areas, setting out the results in greater detail.

2 Presentation of the software used

2.1 Choice of the type of modelling software

A preliminary opportunity study conducted by Cerema Sud-Ouest in 2017 identified two main types of modelling tools suitable for use in the ASTUS project, based on the results they generate in terms of evaluating urban sprawl and changes in land use:

- Integrated land use and transport (LUTI) models: "An integrated land use and transport model is capable of simulating concurrent changes in land use and the transport system, since transport system efficiency is one of the variables on which households and businesses base their choice of location. An integrated model takes into consideration the fact that households choose their location on the basis of exogenous variables (an area's history) and endogenous variables (price, accessibility and amenities), and that both their choices and the variables are likely to change as a result of changes in the transport system and urban development policies"[1].
- Cellular automata (CA): In a cellular automaton model, the region is represented in its entirety by a network of cells, each of which is defined by a given state, such as a type of land use, which can then change over time in accordance with a series of rules known as transition rules.

The constraints laid down in the ASTUS project, notably in terms of implementation time frames and the goal of being able to transpose the methods and tools deployed in the test areas to European level, resulted in the fact that, while the LUTI models are perfectly suited to their modelling objectives, they are disproportionate and too unwieldy for use in this context.

Accordingly, we have chosen to use cellular automaton models instead: even if they cannot, in principle, answer all of the questions we might have, they are relatively easy to use and we know that we will be able to rapidly model different change scenarios for the area's land use. One of the main strengths of these models is that they can model complex behaviours on the basis of very simple operating rules.

[1] Mathieu Saujot (2013). Analyse économique et simulation prospective dans la planification de la ville sobre en carbone : Application à Grenoble du modèle TRANUS+. Économies et nuances. École Nationale Supérieure des Mines de Paris. 470 p.

2.2 Presentation of the two models chosen

2.2.1 FOREcasting Scenarios for citIes using GeograpHic daTa/FORESIGHT

Software name	FORESIGHT
Owner	Toulouse Tech Transfert (www.toulouse-tech-transfer.com)
References	Houet T., Aguejdad R., Doukari O., Battaia G., Clarke K., (2016) Description and validation of a 'non path-dependent' model for projecting contrasting urban growth futures, *Cybergeo*, 759 http://cybergeo.revues.org/27397
Licence	Paid (free when used for research or academic purposes)
Objective	To model various prospective urban-sprawl scenarios

Table 1: FORESIGHT's features and objectives

Taken from the FORESIGHT user guide:

- This model was developed to simulate a city's urban sprawl based on predefined, qualitative scenarios (narratives). It was developed by the UMR 5602 GEODE CNRS-UT2J laboratory as part of the ACCLIMAT project (adaptation of the Toulouse urban area to climate change). The general framework we refer to optimally uses this model (a scenario-based approach) and is presented by Houet et al. (2010, 2016).
- The model is based on the existing SLEUTH model (Clarke et al. 1997, Clarke and Gaydos 1998), which has been widely tested and recognised by the scientific community. The SLEUTH model is a program developed in C language, running under UNIX operating system and using a standard gnu C compiler (gcc). Model and source codes are freely available (cf. Gigalopolis Project).
- Unlike SLEUTH, FORESIGHT incorporates an additional spatial parameter, assimilated to an attractiveness factor in this example. It is not path-dependent and does not require pre-defined (e.g. co-constructed) scenarios. This version was compiled and encapsulated for non-expert end-users. It runs under a Windows operating system using JAVA®.
- FORESIGHT owes much to the SLEUTH (Clarke et al. 1997, Clarke and Gaydos 1998) and SLEUTH (Houet et al. 2016) models and is specifically dedicated to long-term scenario-based urban planning studies.

The FORESIGHT user guide is very thorough and goes into greater detail about both the theory and practice of the tool's operation, along with some of the finer points of the software's configuration. We drew heavily on the user guide to construct the methodology report that follows, while also delivering our feedback on the trial we conducted.

2.2.1.1 Operating principle

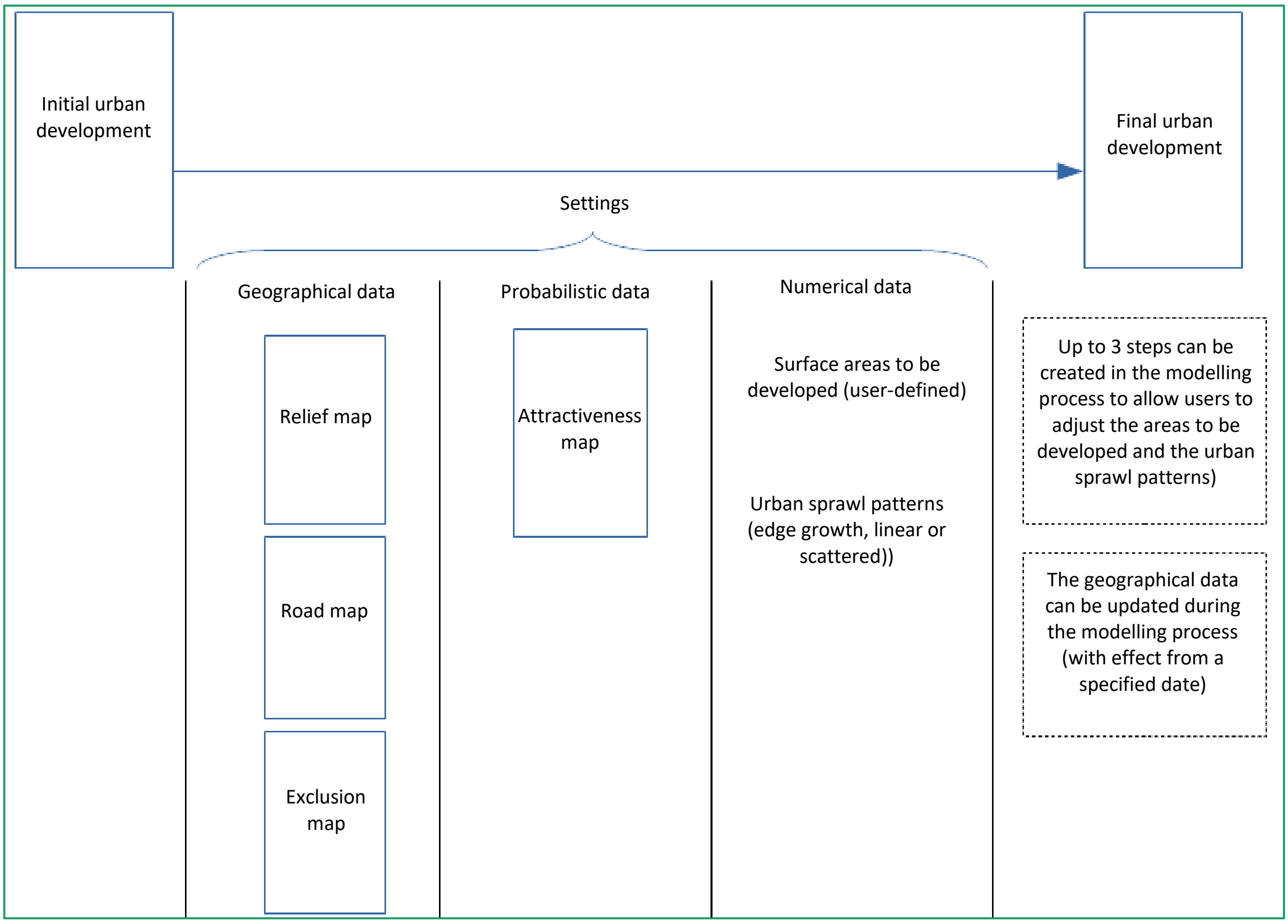


Figure 1: Block diagram of the FORESIGHT software

2.2.1.2 Interface

The FORESIGHT interface provides access to six tabs:

- *Pre-processing:* prepares the image data required for simulation;
- *Attractiveness generation:* a simple way of generating an attractiveness map;
- *Initial settings:* checks that the required image data is available in the appropriate folder;
- *Scenario parameters:* defines the numerical parameters for simulation, such as the modelling time horizon or the urban sprawl patterns;
- *Land management strategies:* updates certain map data during modelling;
- *Output:* launches the simulation.

Figure 2: FORESIGHT interface

2.2.1.3 Pre-processing tab

2.2.1.3.1 How the Pre-processing tab works

FORESIGHT has a pre-processing module for preparing the geographical input data. If the data generated is in the right format (*.GIF), there is theoretically no need to use this module.

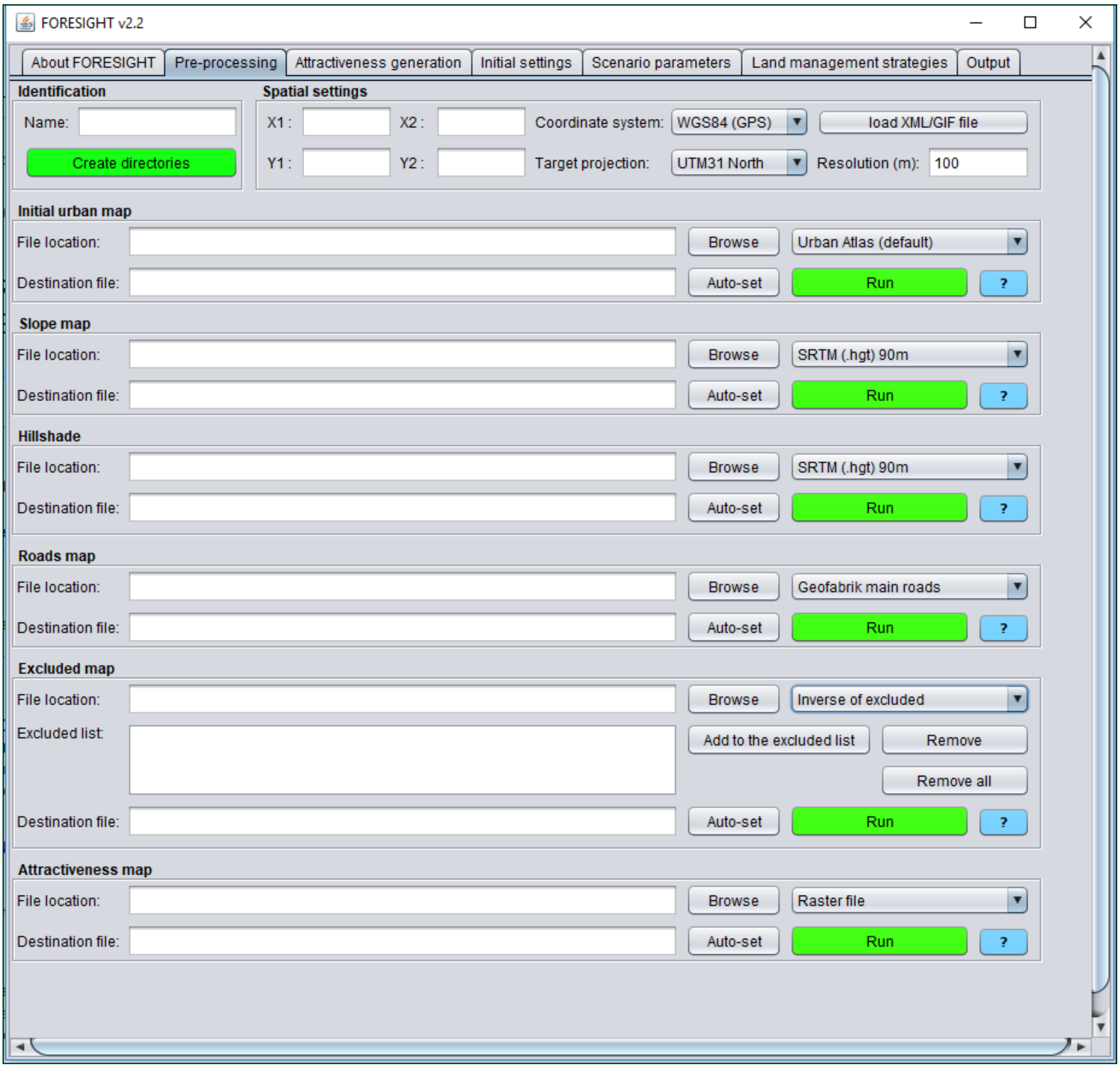


Figure 3: Foresight interface

Before using the FORESIGHT software, the first essential step is to create a working directory in which the software will implement the data required for modelling, along with the modelling results.

Once the directory has been created (at the root of the folder containing FORESIGHT), two other folders have to be created: *Input*, to contain all of the image data input required for modelling, and *Output*, to contain all of the images produced during a simulation. If another modelling operation is launched without changing the name of the directory, all of the earlier data will be overwritten by the new data. It is a good idea to store the data created in a different folder for subsequent access if necessary.

In the same tab, the geographical coordinates of the study area (*Xmin*, *Ymin, Xmax* and *Ymax*) have to be entered in the *Spatial settings* pane. It is important to specify the coordinate system in which the input data is provided and that of the "target projection", which means the projection that will be used to match the data and generate urban sprawl maps. The possible projections that are admissible as input are WGS84, Lambert 93 and UTM31 North. The latter two are also available for the "target projection" coordinate system.

The model's resolution is also required information. The resolution sets the pixel size for reading input data. It is theoretically independent of the modelling. If the resolution used is greater than the quality of the input data, the data will be simplified.

The software's default minimum mapping unit is 100 metres square (i.e. a 1 ha pixel).

The next pre-processing step is to generate each of the maps requested at the outset by the simulator.

A source file is requested for each map. Once the file is selected, the program defines a default output file name. Changing the output name can be useful for generating, for example, a new, updated version of an input file, which can then be used in the *Land management strategies* tab. The *Auto-set* function restores the default name.

Note that only a file with the default name can be recognised by the software for subsequent modelling.

The *Run* function generates the file (in *.gif format) in the *Input* folder.

2.2.1.3.2 Initial Urban Map

The software draws on the various available sources to generate an initial map of urban development, without specifying what urban development means: it might be sealed surfaces (as in CORINE Land Cover or the Urban Atlas, for example) or the urban footprint, whatever method is used to define that concept. This initial urban map is generated from a GIS vector GIS (*.shp format) obtained from one of the following three sources:

- CORINE Land Cover (CLC)
- Urban Atlas
- My Own File

FORESIGHT recognises raw (i.e. not pre-processed) data from CORINE Land Cover or the Urban Atlas and can extract from it the information required to construct the initial urban development map. To do so, it either selects, by default, all of the classes contained in the sealed areas (first level of nomenclature) or lets users select the land covers they want to include in the map.

To use a different source, users must first pre-process the data in a GIS. This entails using a source file to create a vector layer with one of two values, 0 or 1, in the numerical field, where the value 1 represents the urban fabric we want to include in the *Initial Urban Map*.

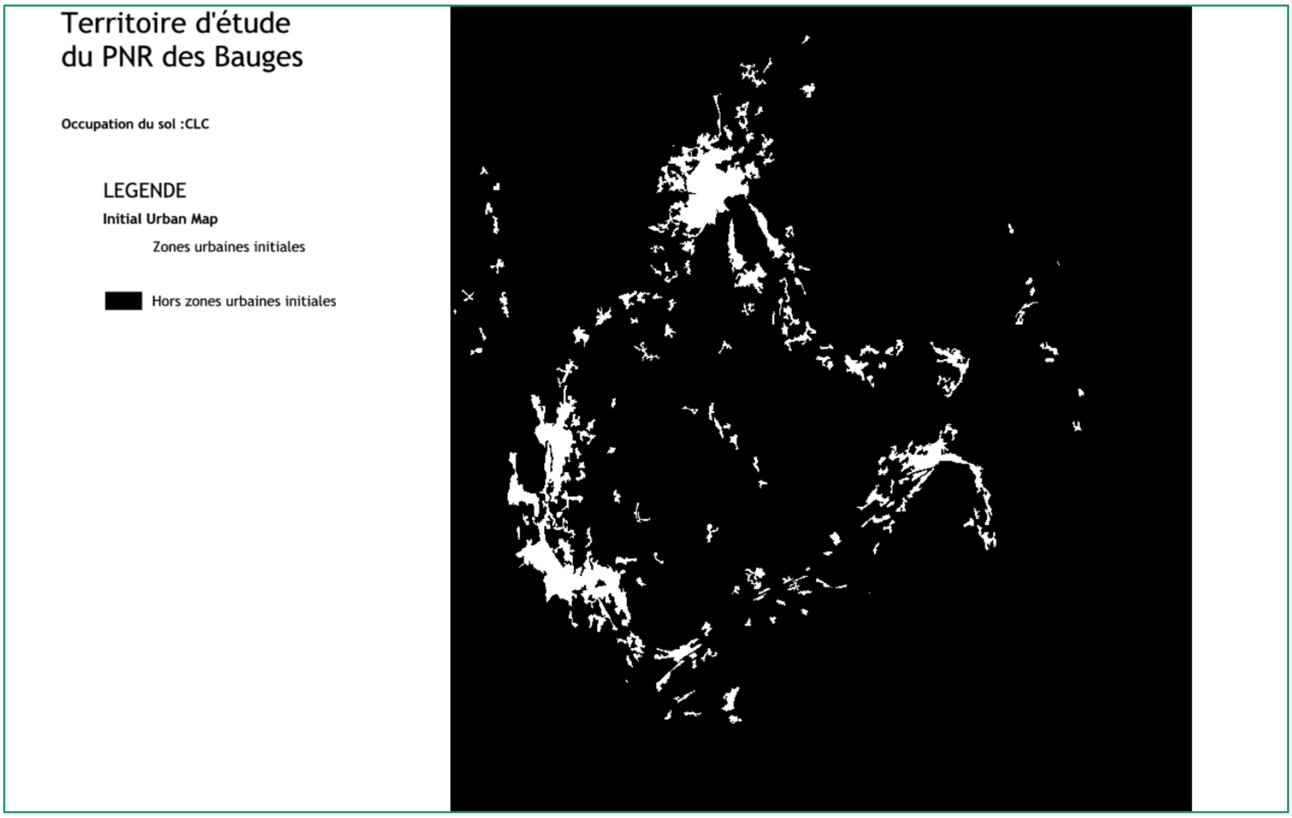


Figure 4: Initial Urban Map calculated with the CLC data for the Bauges Regional Nature Park

2.2.1.3.3 Slope Map and Hillshade

The software generates a relief map and a base map.

These maps can be generated from any digital terrain model (DTM). By default, the software uses the same source to generate both types of map.

The maps are generated from raster files obtained from a variety of sources:

- Shuttle Radar Topography Mission (SRTM) (*.hgt format)
- IGN data
- EPSG data

These sources must use a Lambert 93 projection to be assimilated by the software.

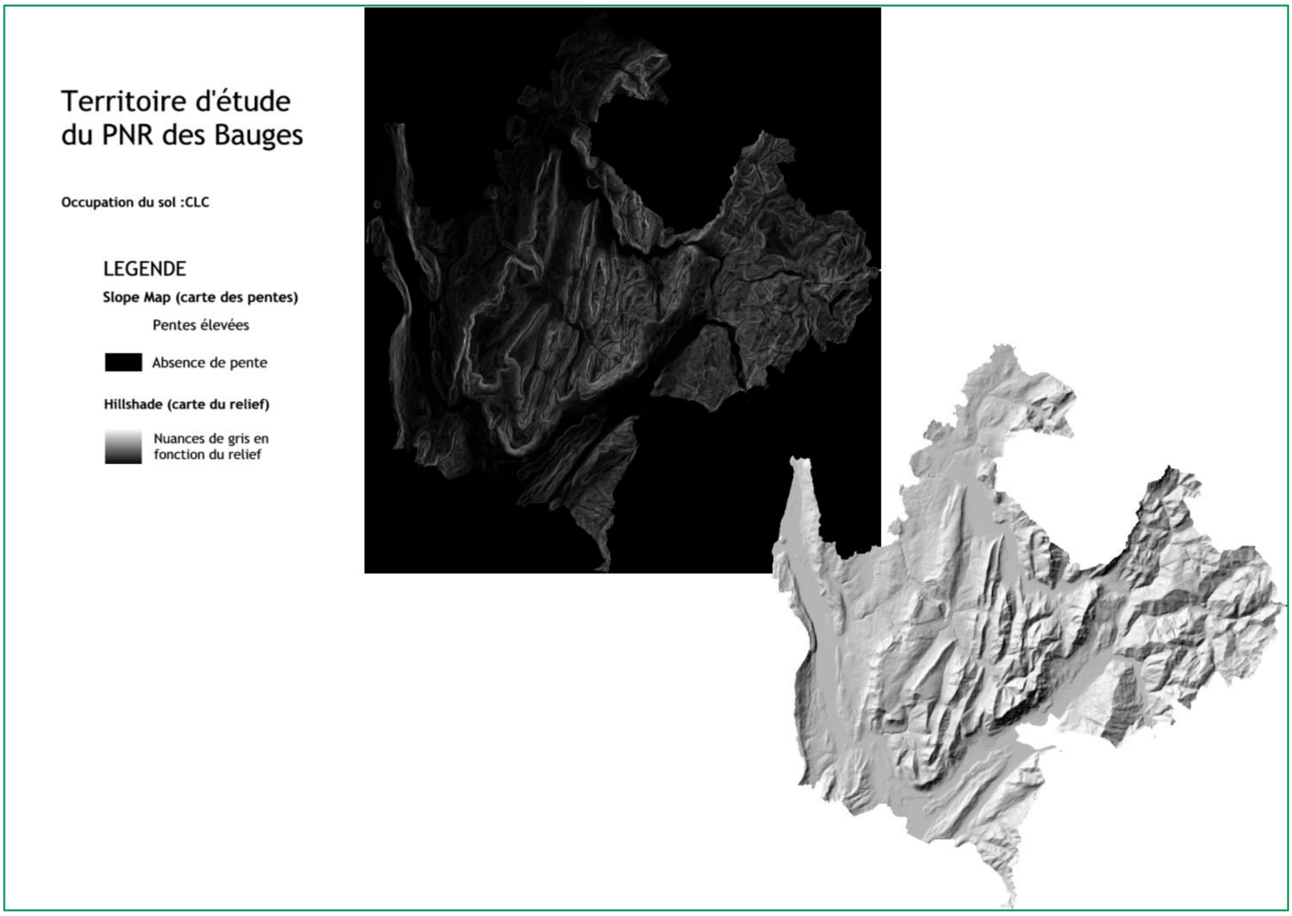


Figure 5: Slope map (top) and Hillshade map (bottom) for the Bauges Regional Nature Park area, calculated with EU-DEM data

The finer points:

- A *.hgt file, even if not derived from the STRM, can be recognised if the first option is selected.
- Any raster file and any DTM sourced from EPSG can be incorporated.
- The base map can be left empty (select the Blank option).

2.2.1.3.4 Roads Map

FORESIGHT uses this map to model linear patterns of land use. The choice of the type of network (principle, secondary, etc.) to be included in the map is left to users, who should consider to what extent the network will actually influence urban sprawl and in particular linear urban sprawl.

To construct the road map, the software's default option is to process OpenStreetMap (OSM) data. OSM data for each region can be downloaded free of charge from the Geofabrik[2].

[2] http://download.geofabrik.de/

Important

Apparently the OSM attribute table has changed since the FORESIGHT software was developed and the latest OSM data is not recognised by FORESIGHT. The road type (motorway, main network, secondary network, etc.) must now be displayed in a column entitled TYPE. To work around this problem, users can simply rename the column currently entitled FCLASSE.

To construct this map, FORESIGHT users can select *Main roads* such as *Motorway*, *Trunk* and *Primary*, along with *Secondary* roads and/or *Tertiary* roads. Users can also make their own selection from among the various road types.

To construct this map from any other source, users must first create a vector layer (lines only) with one of two values, 0 or 1, in the numerical field, where the value 1 represents the road network to be included in the map.

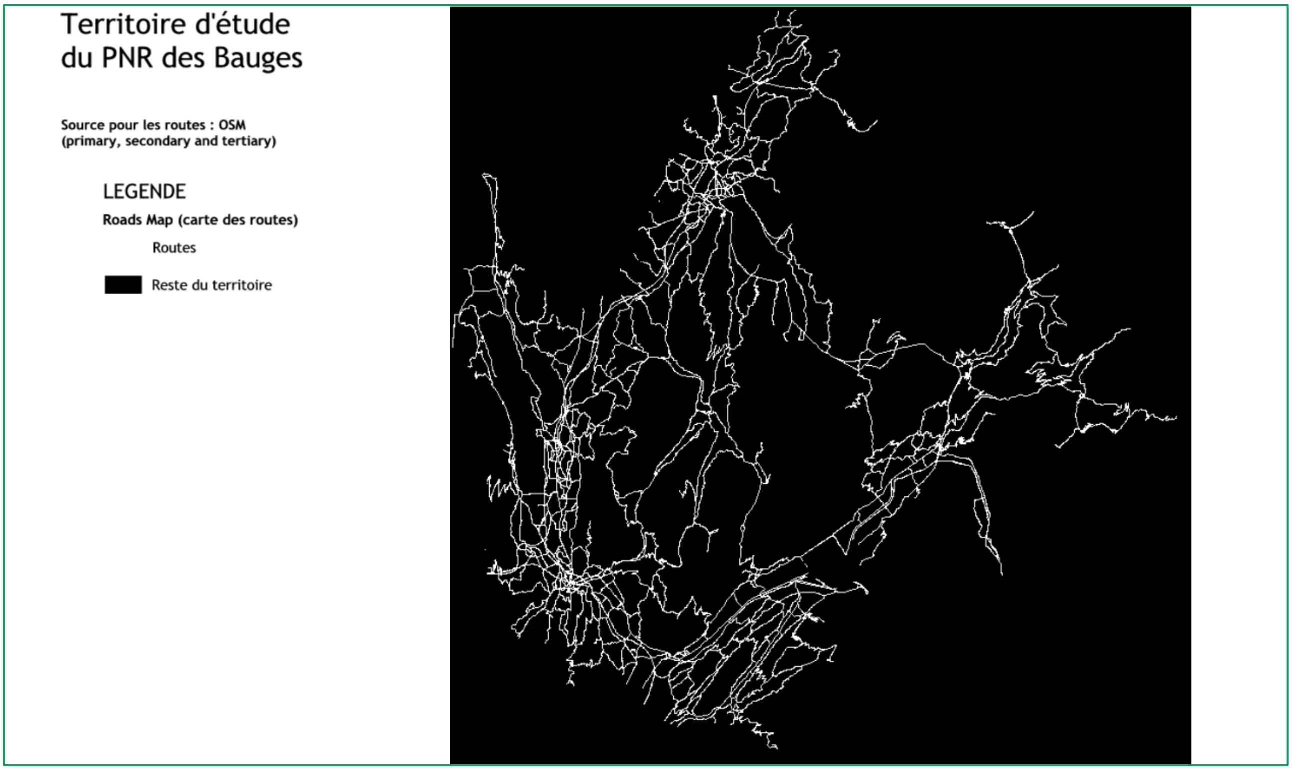


Figure 6: Road map calculated with OSM data (primary, secondary and tertiary road networks) for the Bauges Regional Nature Park study area

2.2.1.3.5 Excluded Map

The Excluded map shows the sectors for which FORESIGHT will not be able to create new urban development pixels. It can be used to show the boundaries of the study area, i.e. the area outside which no urban development is possible, or areas covered by water, or any other type of specific zoning, such as flood zones or protected areas, if the latter are considered sufficiently restrictive to exclude any urban development for the entire period being modelled.

This map is generated by superimposing as many layers as desired. A number of vector files (polygons only) can be placed on the *Excluded* list. The software also provides an *Inverse of excluded* mode, which can be used to restrict modelling to the study area only. Note, however, that these two modes cannot be combined in FORESIGHT. It is possible and advisable to prepare a single layer beforehand containing all of the areas to be excluded from modelling.

2.2.1.3.6 Attractiveness Map

The last map in the pre-processing tab - the attractiveness map -supplies the software with probabilistic information used to influence the generation of new urban development pixels. A specific FORESIGHT module uses a generic method to create this map (cf. below).

Users can produce an attractiveness map with their own resources (*.tif format), then incorporate it through this pre-processing module. To do so, the raster file produced must be encoded in shades of grey, using whole numbers between 0 (excluded) and 100 (maximum attractiveness).

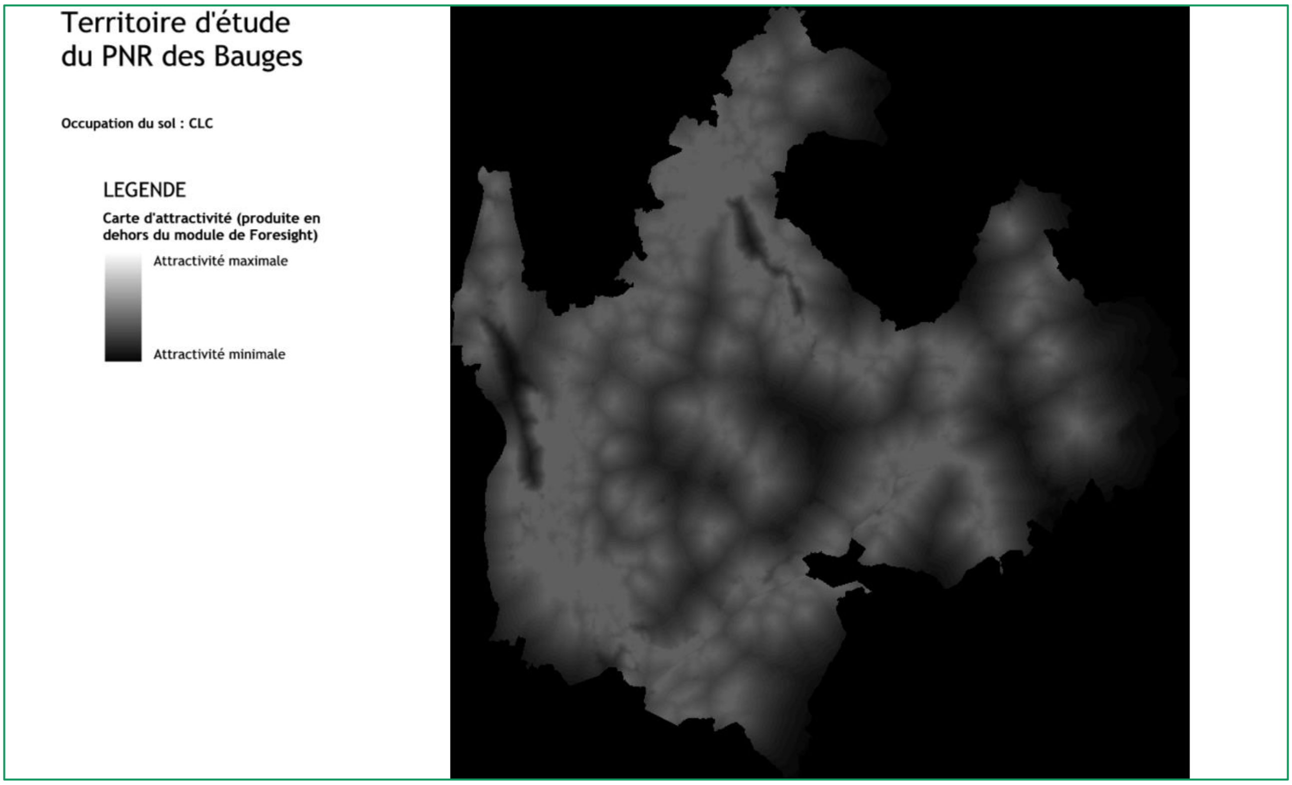


Figure 7: Attractiveness map produced with CLC data for the strategic scenario and for the Bauges Regional Nature Park study area

Important

We have tested this option, but so far have not been able to get the FORESIGHT[3] software to recognise our attractiveness map.

[3] After further testing, we were able to incorporate one of our attractiveness maps in FORESIGHT. There is no way to do this through the *Pre-processing* tab. However, it was possible to generate the *.gif file for one of our attractiveness maps directly through ArcGIS. Users will need to ensure that the resolution and extent are strictly identical to those in the files generated by FORESIGHT in the *Input* folder for the *Initial Urban Map* data, for example. The file was then copied into a model's *Input* folder, where it was recognised by the software *(Preview* in the *Initial settings* tab*)*, then used to generate a model based on our own attractiveness map.

2.2.1.4 Attractiveness generation

This map is drawn up on the same methodological principle as a cost-distance map:

- Bear in mind that the "cost" of developing a pixel is not the same throughout the area: it will be higher in agricultural, forest or protected areas than in areas that have already been developed or areas earmarked for development in an urban planning document. Starting from a given point (or set of points and/or polygons) in the area, the software then assesses the "cost" entailed in moving from that particular point to any other point in the area, thereby establishing the "cost" of urban development at any point in the area.

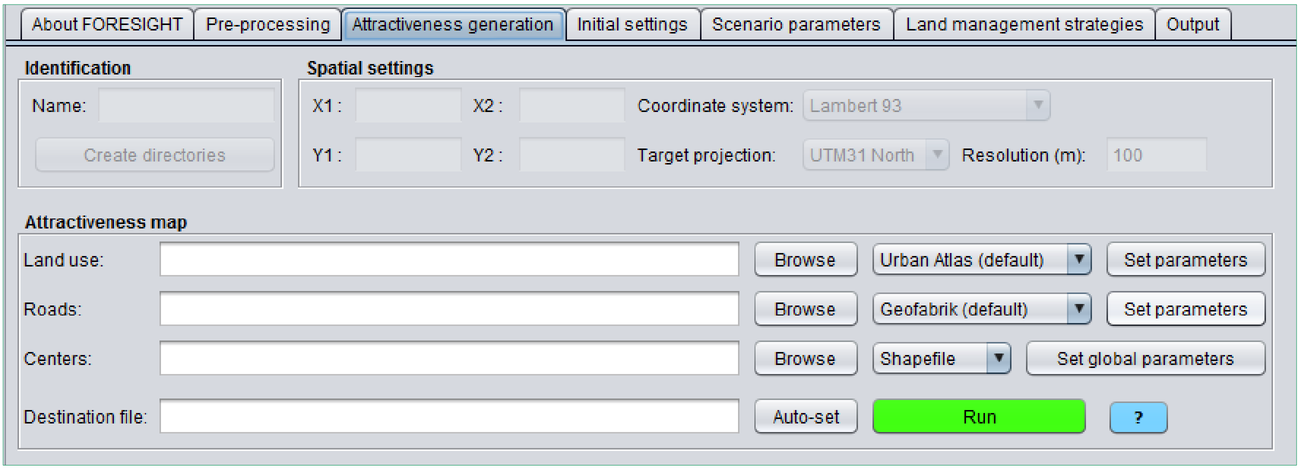


Figure 8: FORESIGHT interface

The FORESIGHT module used to establish the attractiveness map takes into account:

- Land use (sources: CORINE Land Cover or Urban Atlas): a friction factor can be defined for each land use class.
    - Friction is the opposite of attractiveness. The friction factor is an attempt to capture the notion of "cost" mentioned above. In other words, the "resistance" of a particular type of land use to urban development;
- The road network (source: OpenStreetMap): a friction factor can be defined for each type of road, to reflect the road network's influence as a vector of urban development.
- Centres: the module can use a network of points or polygons as baselines for evaluating the cost of developing all of the pixels in the area.

The cost of crossing can be defined for each type of land use and each type of road. Any value whatsoever can be entered, from 0 (no cost whatsoever to cross the pixel) to 255. Values equal to or above 255 and negative values are treated as an inaccessible area. Note that the specific procedure involved in generating attractiveness with this module is designed to detect excluded areas and treat them as natural boundaries.

To produce this map, the module uses the same sources as for the *Pre-processing* tab: the Urban Atlas and CORINE Land Cover data for the land use map, OpenStreetMap for the roads and a user-supplied shapefile file for the map of the centres. Each pixel corresponding to a centre will have a displacement value equal to 0 and therefore an attractiveness value equal to 100%. Default values are displayed (in default mode) to define the cost of a road pixel or a land use pixel. Depending on the chosen data source, an advanced mode lets users manually define the cost of crossing each pixel of road and the cost of using the available land (short list or full list).

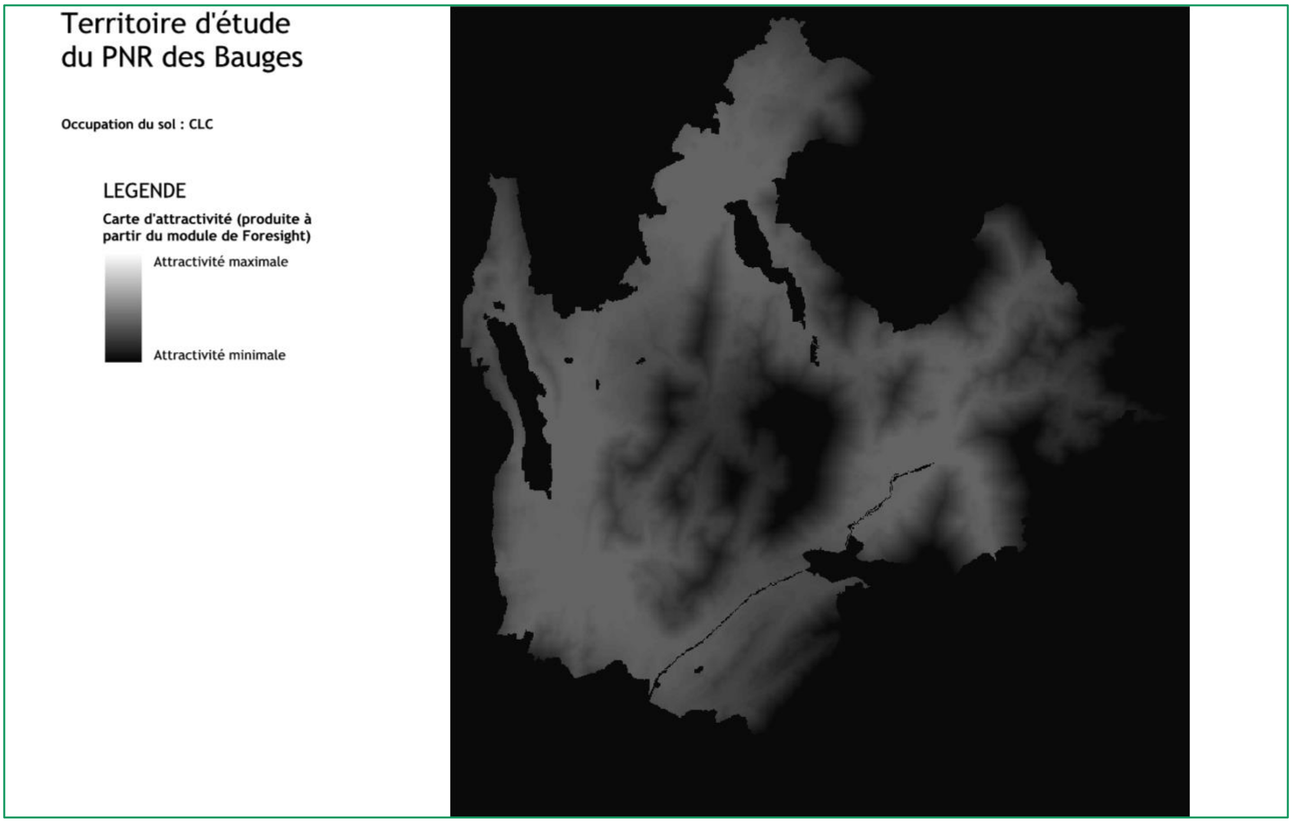


Figure 9: Attractiveness map produced with the dedicated FORESIGHT module (land use with CLC, roads with OSM) for the strategic scenario and for the Bauges Regional Nature Park study area

From a theoretical viewpoint, the map of simulated costs presents values ranging from 0 (centres) to potentially unlimited values, depending on the size of the study area. The attractiveness map, on the other hand, contains only values ranging from 0 (nil attractiveness) to 100 (maximum attractiveness). To generate the attractiveness map from the costs map, the software has three complementary parameters *(Set global parameters);*

- *Minimal attractiveness:* the attractiveness of the most distant pixel is set by the user to the minimum attractiveness value.
- *Quantile to min.:* sets a percentage of pixels whose values are the lowest for the entire costs map, for which the attractiveness value will be set at the minimum level of attractiveness *(minimal attractiveness)*
- *Quantile to max.:* sets a percentage of pixels whose values are the highest for the entire costs map, for which the attractiveness value will be set at the maximum level, i.e. 100% attractiveness.

All pixels whose value is not directly defined by the above parameters have an attractiveness value that lies between minimum attractiveness and 100, and which distributes the pixels according to a linear function (see explanatory diagram in the user guide).

2.2.1.5 Initial settings

Once the model's input data has been generated through either the *Pre-processing* tab or the *Attractiveness* tab, and is available in the required format in the *Input* folder in our working directory, we can confirm the data's availability in the *Initial settings* tab. The purpose of this tab is to define the input map data to be used to run a simulation. It also provides a preview of the various maps used *(Preview).*

It is quite possible to retrieve map data from other folders if, for example, users want to reuse information already produced elsewhere. It should be borne in mind that, during the simulation, all of the input data will be copied again to the *Input* folder in the study directory, without overwriting the data already present, produced through the previous tabs. All of the simulated data will then be found in the *Output* folder.

2.2.1.6 Scenario parameters and Land management strategies

FORESIGHT is designed to assist with the spatial representation of a long-term forecasting approach. Unlike the SLEUTH tool from which it was developed, the FORESIGHT software is not designed for baseline or automatic operation. In the *Scenario parameters* tab, it is the user who sets:

- The scenario duration;
- The land take (in hectares);
- The urban sprawl patterns, which are summed up in 4 patterns[4] (numerical data whose sum must be equal to 100%):
  - Spontaneous growth;
  - New urban centres, from which urban sprawl again begins to spread;
  - Linear growth;
  - Edge growth, continuing on from existing urbanisation.

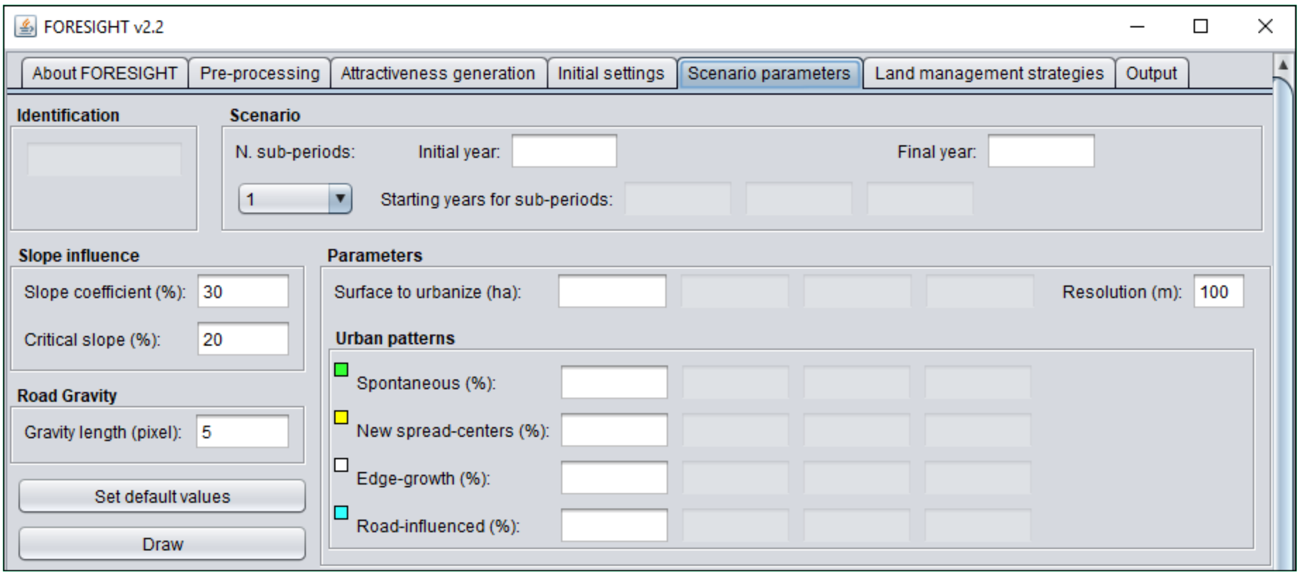


Figure 10: FORESIGHT interface

Users can also define up to three sub-periods in the model. These sub-periods can factor in changes over time in urban sprawl patterns and in land take.

Please note that the resolution must be defined here if the input data was prepared previously with a different resolution to the one users want to use for modelling. The software's default minimum mapping

[44] The FORESIGHT user guide provides a detailed definition of each of these parameters.

unit is 100 metres square (i.e. a 1 ha pixel). Any whole value is accepted by the software but, as stated in the user guide, only values between 50 m and 200 m are actually acceptable from a theoretical viewpoint: values outside this range are highly questionable. We have therefore set the default minimum mapping unit for modelling purposes at 100 metres square.

In addition to these parameters, another two criteria are used to specify the influence of the relief and the road network on the preferred location for urban development (for detailed information, please refer to the FORESIGHT user guide). FORESIGHT sets default values for these parameters, which we did not change for our tests.

Next steps: land management strategies

In the *Land management strategies* tab, FORESIGHT allows users to change some of the scenario's mapping data. When users want to draw up a change scenario for the area, they may need to update certain geographical information. This might mean adding a new infrastructure to the road map, for example, or taking the infrastructure changes into account in the area's attractiveness map. In this tab, users can update the road map, the exclusion map and the attractiveness map. The changes will be factored into the model at the date they desire and, in principle, as many times as necessary.

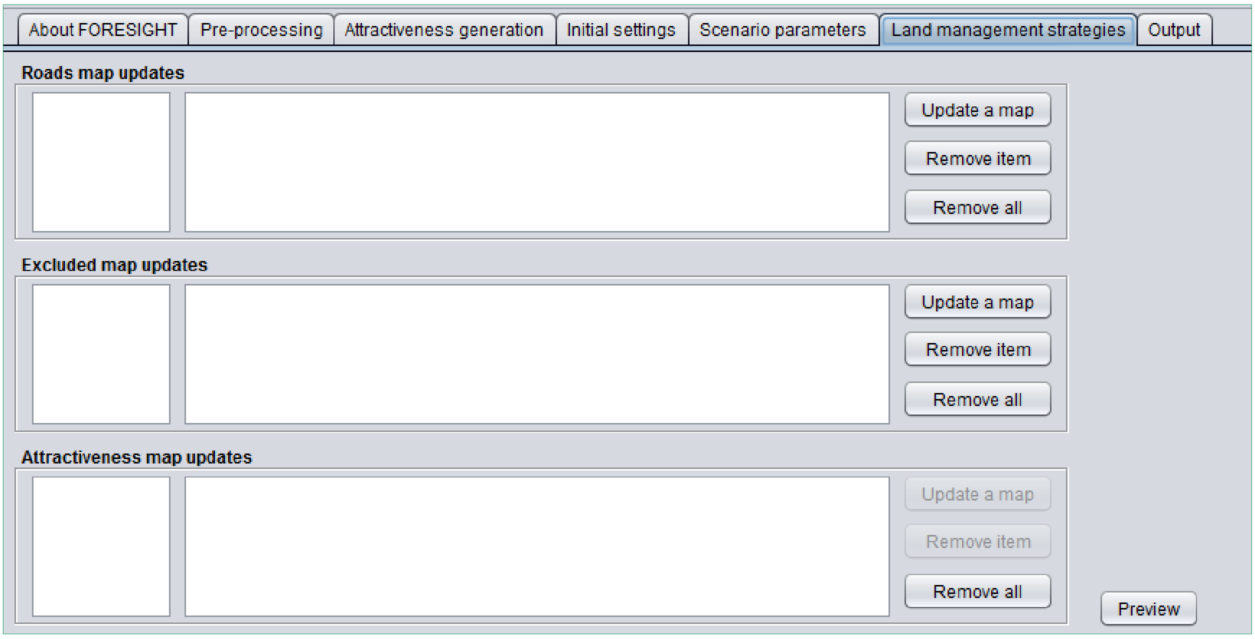


Figure 11: FORESIGHT interface

To be taken into account by FORESIGHT, these maps must be formatted as input data (*.GIF). If this data is updated in the GIS, the *Pre-processing* tab can be used to prepare the *.GIF images for inclusion here. In the *Input* folder, the software will then create a copy of the formatted information for inclusion by the modelling module. Note also that the updated information must match the full updated layer, i.e. to update the road map, a map of the entire road network to which the new infrastructure was added must be submitted to the software for inclusion.

During our tests of FORESIGHT, we did not use the possibilities offered in the Land management strategies tab.

2.2.1.7 Output data

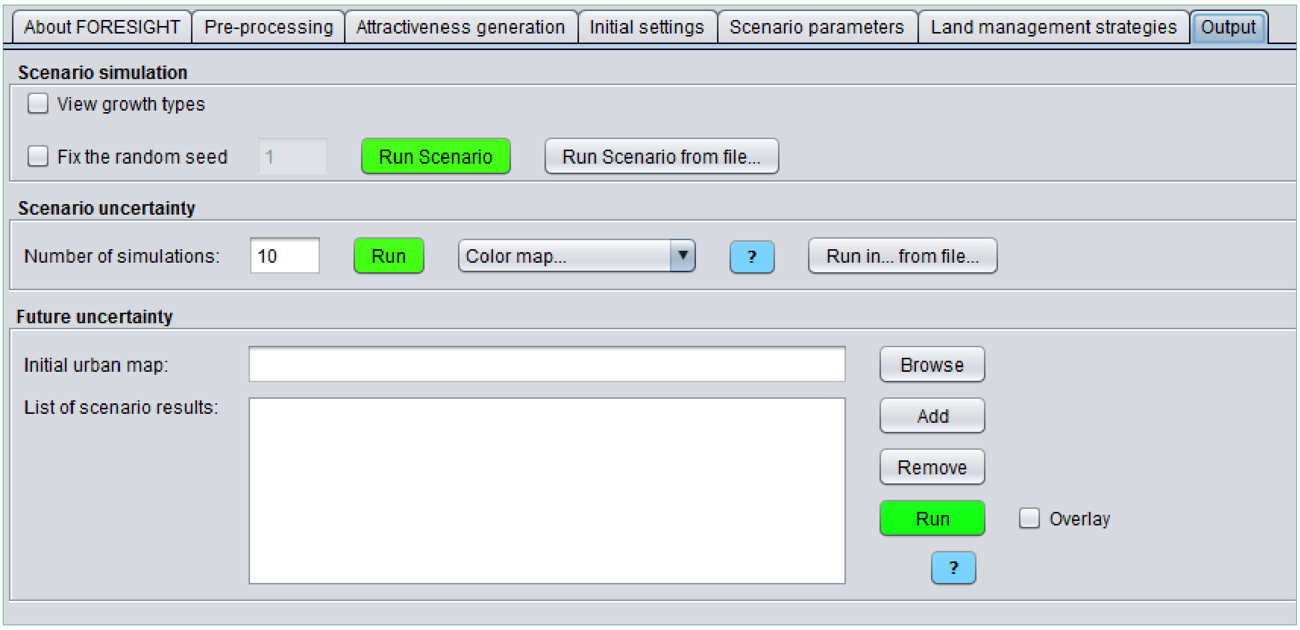


Figure 12: FORESIGHT interface

For a given scenario, FORESIGHT produces (in the *Run scenario*) a map for each year of the entire modelling period, showing the new urban development pixels each time. The software also provides a computer-generated image showing all of the pixels developed throughout the scenario, along with the map showing all of the urban development (initial + simulated).

If the user ticks the *View growth types* option, the software also generates maps that differentiate, for each year, the pixels created from the various *Patterns* set in the *Scenario parameters* tab.

By default, each model is different to the previous one (random simulation). The *Fix the random seed* option can be used to always run the same simulation[5].

In the *Scenario uncertainty* pane, users can run several successive modelling operations based on the same scenario. The output is available in a new folder in the working directory *(Future uncertainty).* It provides as many maps as there were models requested; each one represents the cumulative urban development (initial + simulated).

[5] Further information is provided in the FORESIGHT user guide.

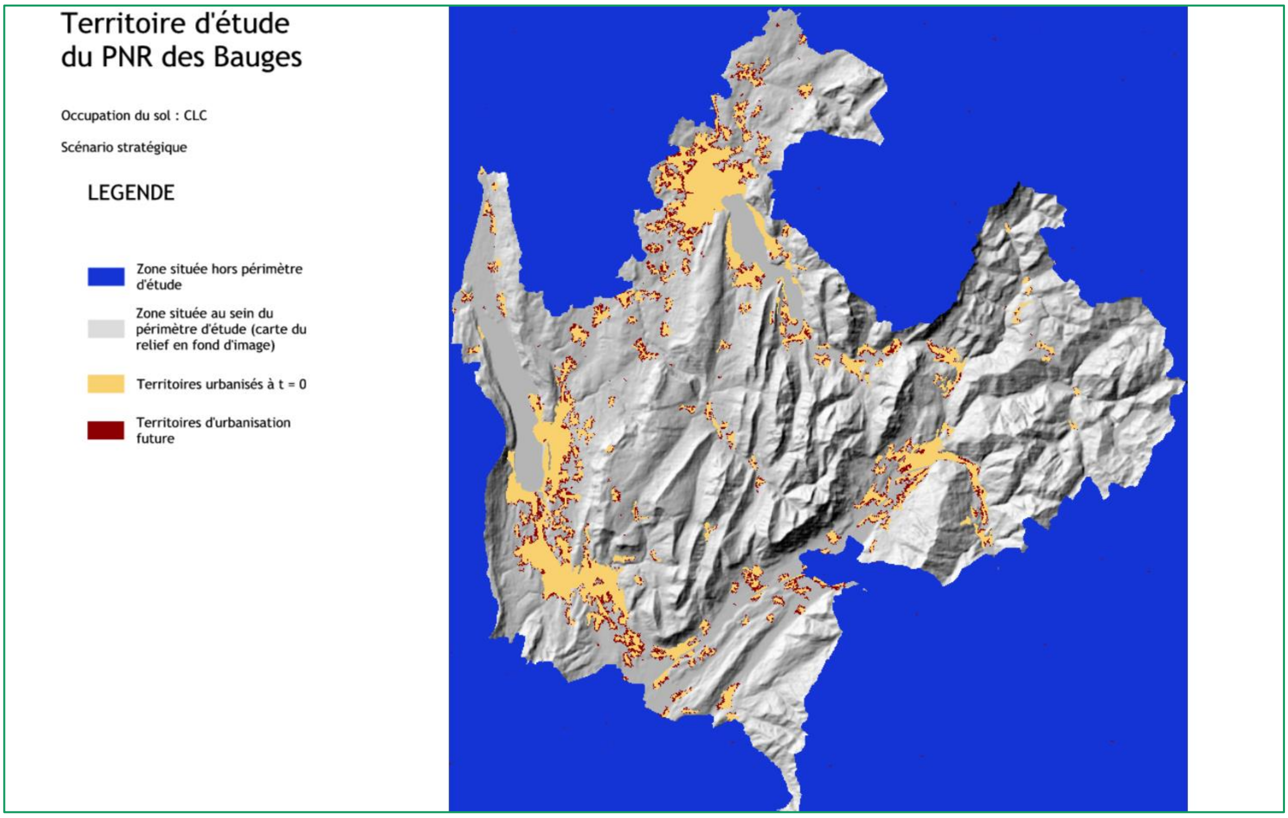


Figure 13: Model of urban sprawl through to 2050 in the Bauges Regional Nature Park study area (strategic scenario, CLC land use data)

Before discussing the features of the final pane, *Future uncertainty*, it is important to point out and remind users that the urban sprawl generated by the FORESIGHT software reflects the influence of all of the user-defined parameters, but is nevertheless randomly created. There is not necessarily an answer to the following questions: Why was this specific pixel developed in the area? Why this one rather than the neighbouring one?

To avoid this "pixel by pixel" analysis, the *Future uncertainty* pane provides a more probabilistic approach to urban development. To apply this option, users need to provide several input data:

- The initial urban map: this image in *.gif format can also be found in the *Input, Output* and *Future uncertainty* folders.
- Several cumulative urban development maps for one or more scenarios: these maps in *.gif format are taken from the *Output* (*cumulate_urban.gif*) or *Future uncertainty* (*cumulate_0.gif*, *cumulate_1.gif*, etc.) folders.

Based on these various models, the software will generate a map showing the probability, for each pixel in the area, of the pixel being developed. This corresponds to the relationship between the number of models in which the pixel was developed and the total number of models taken into account. As mentioned above, it should be pointed out this map can be generated, for example:

- From 20 models produced for a given scenario, using the *Scenario uncertainty* pane: we will then have information telling us, for a given scenario, for which sectors the probability of urban development is close to 100%, almost certain, and those for which the data is far more uncertain or random.
- From models generated by different scenarios: we will then have information telling us in which sectors urban development is highly likely to occur, whatever the area's change scenario.

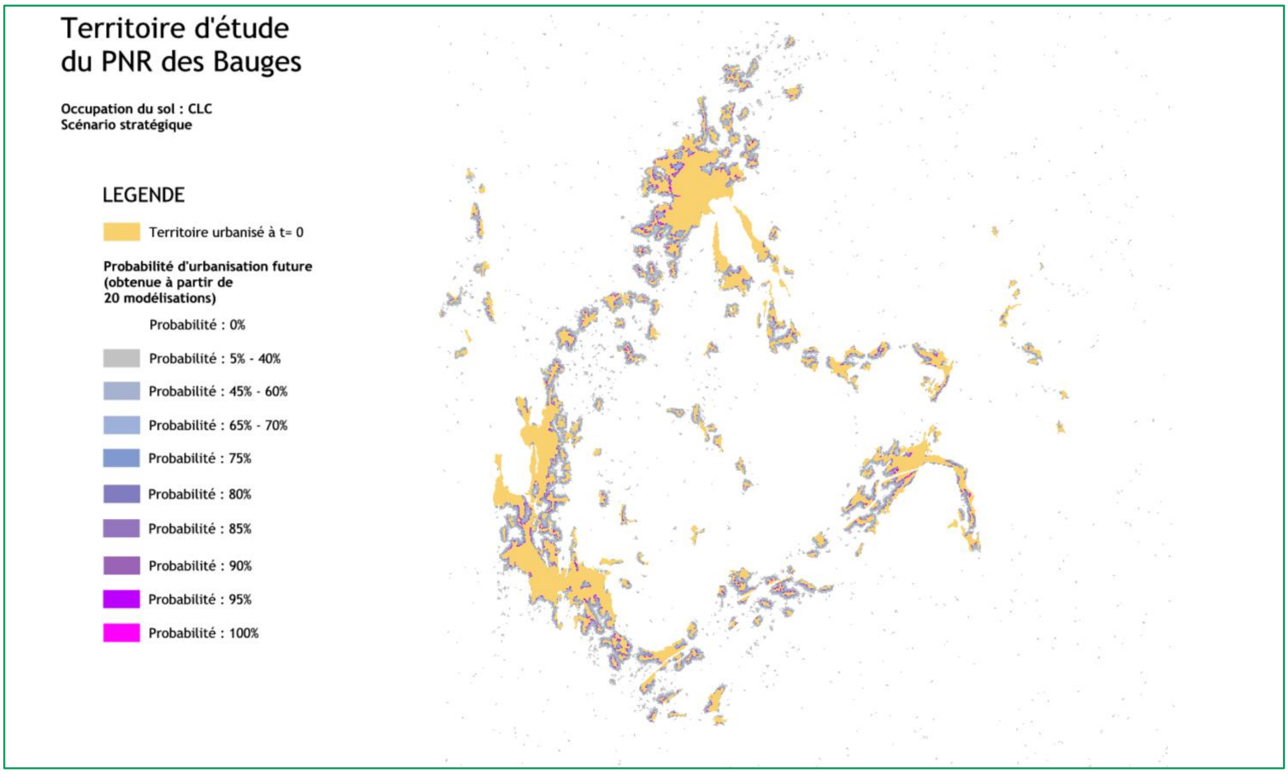


Figure 14: Probability of urban development, obtained from 20 models for the Bauges Regional Nature Park study area (strategic scenario, CLC land cover data).

2.2.2 Land Use Cellular Automata Simulation (LUCSIM)

Software name	LUCSIM
Owner	Université de Bourgogne-Franche-Comté / Laboratoire ThéMA
Website	https://sourcesup.renater.fr/LUCSIM/
Licence	GNU (free, open-source software)
Objective	Model changes in land use

Table 2: LUCSIM's features and objectives

Taken from the LUCSIM user guide:

- "The goal of the project is to give users a tool for simulating land use changes with cellular automata, to help them understand spatial dynamics. This software is a development environment for cellular automata modelling. More specifically, it is designed to simulate urban development from an image representing land use (building, forest, road, etc.) and transition rules for different land use states."»

2.2.2.1 Operating principle

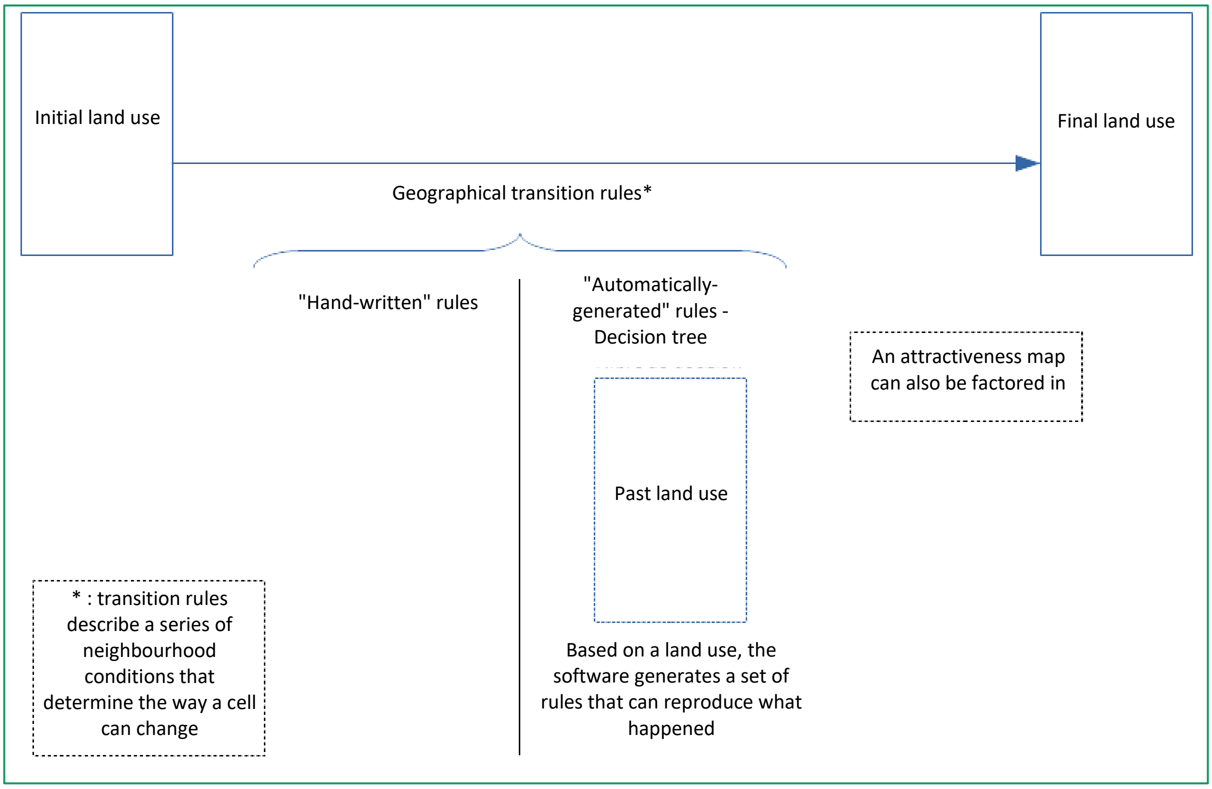


Figure 15: Block diagram of the LUCSIM software

2.2.2.2 Interface

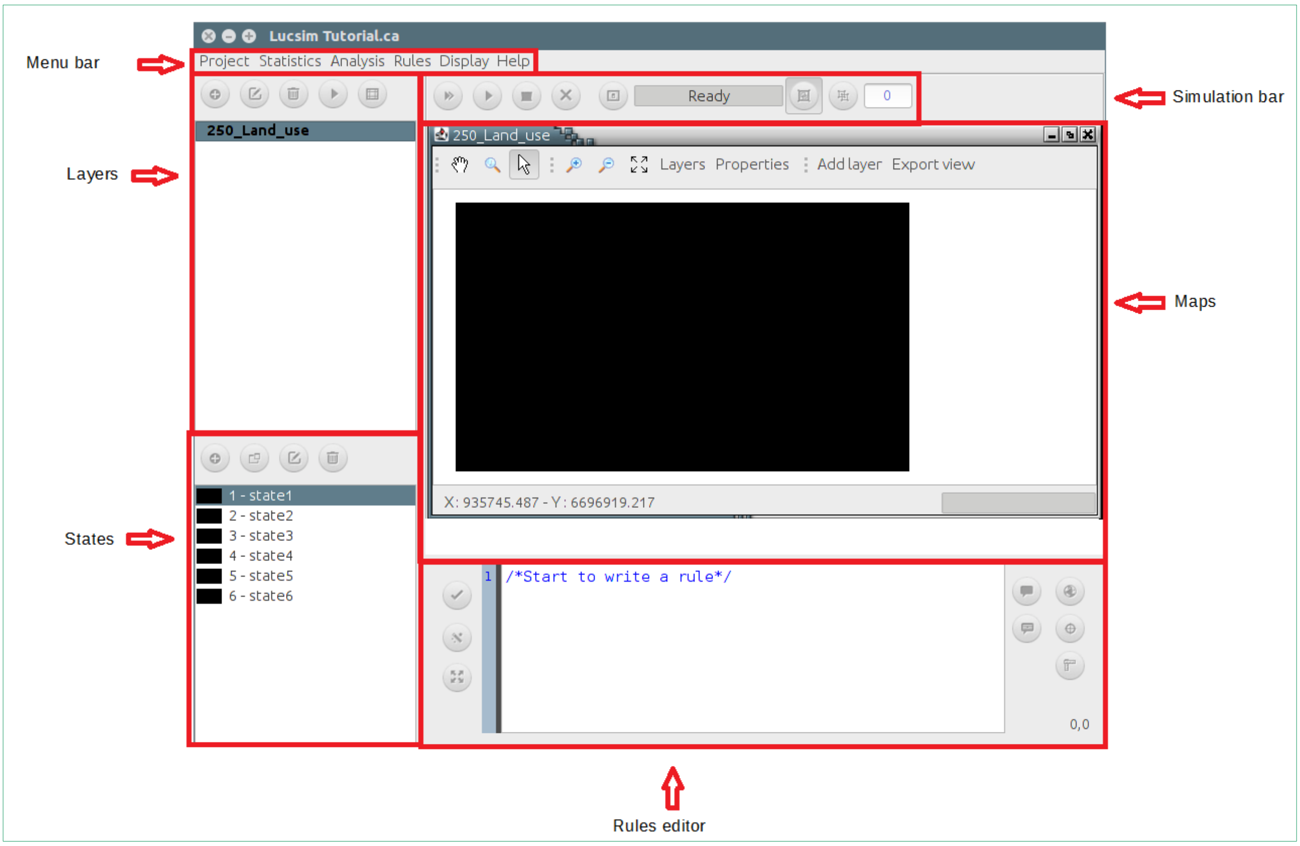


Figure 16: LUCSIM interface

LUCSIM is written in Java and, for now, operates only with version 8 of the *Java Runtime Environment* (JRE).

2.2.2.3 Input data

To model changes in land use for a given area, users must have at least a land use map for a given date. These maps are *.tif images, in which the pixel codes correspond to different categories of land use.

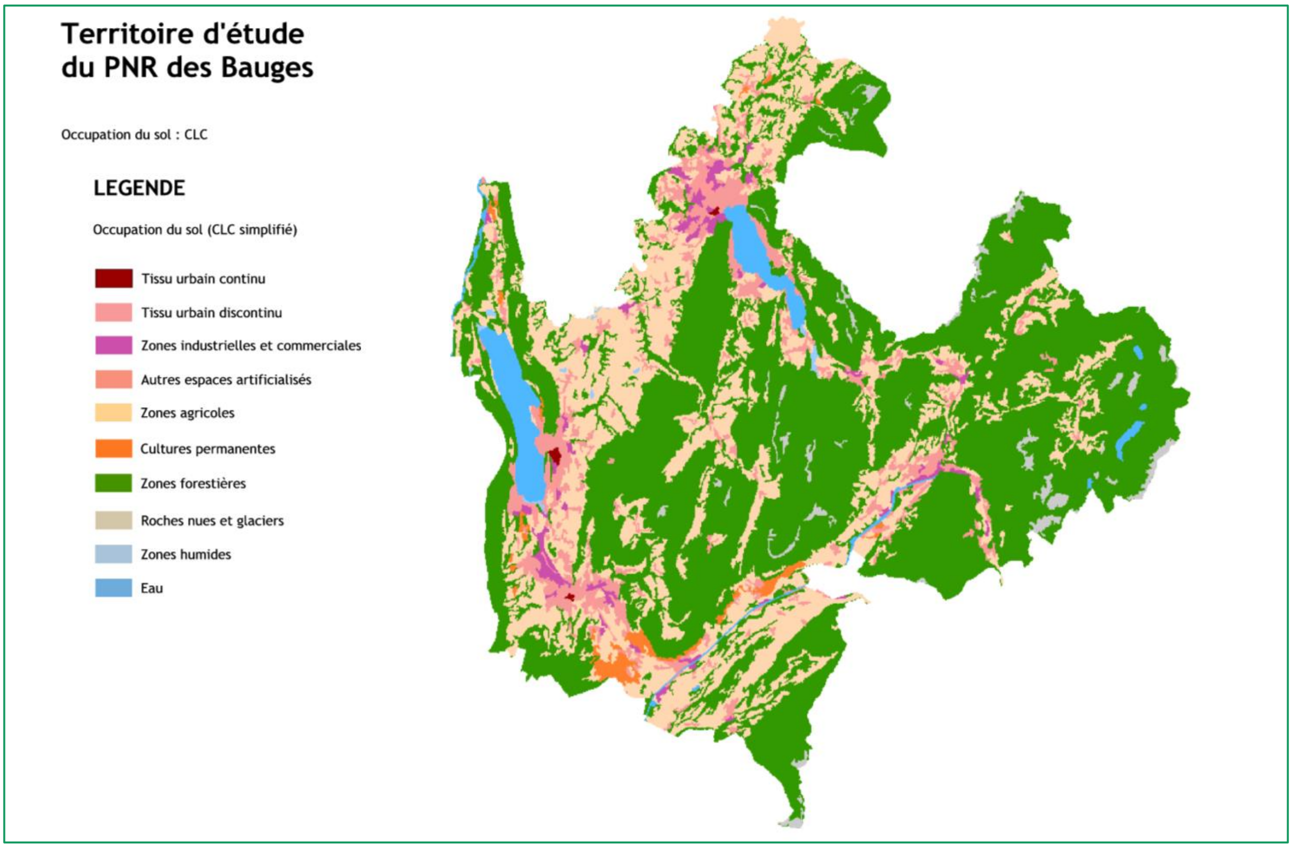


Figure 17: Land use (simplified CLC data) in the Bauges Regional Nature Park study area in 2012

Various types of input data can be used to create a project in LUCSIM. They do not all bring the same information to the software.

LUCSIM makes a distinction between the raster layers it will use for the cellular automaton, i.e. the "State layers" it uses for modelling, and other raster layers, which it might use, for example, to calibrate the transition rules. The first map incorporated is automatically added as a *State layer* and will serve as the baseline for modelling. When another raster layer is incorporated into the program, LUCSIM asks the user whether it is a *State layer.*

Images in *.tif format can be added to the project by clicking *New project* in the initial interface or *Add layer to the list* from the *Layers* pane. To be assimilated by the software, each pixel of the image must be characterised by a natural whole number between 0 and 255. In the raster layers, the value 255 is understood as NoData.

When users want to incorporate a second *State layer*, there are two points to watch:

- The value of the pixels in each layer: the pixels must be coded in exactly the same way as the initial map for the software to be able to identify the different types of land use;
- The alignment of the two layers: the layers must be aligned. This means that it must be possible to overlay each pixel of one layer on a pixel of the second layer.

Note that if a land use feature exists in only one of the two editions, the software will treat it as an appearance (or disappearance).

Once incorporated, the image appears in the *Maps* pane and all of the land use categories in the *States* pane. At this point, users can rename the different categories and change their colours (by double-clicking them or clicking *Modify state*).

It is also possible to add new images, but they must be in *.tif format. These layers do not contain the various categories of land use, but they do provide other types of information, such as an exclusion map, a *Suitability map* (attractiveness map) or other information such as the road network (raster version). These layers can be used to calibrate the transition rules but not to run simulations.

We tried to incorporate our own attractiveness maps *(Suitability Maps)* into the software. To be operable, the *Suitability map* must, in principle, be calibrated on the map produced by the *Potential model* (see below), i.e. a map for which the highest values (no upper limit, in principle) are the most attractive, and vice versa (zero corresponds to a pixel devoid of attractiveness.

This attractiveness map can be incorporated as an additional layer in the decision tree. This may have an impact on the calculation of the transition rules, or it may be applied as a modelling constraint through the *Rules/Set constraints* menu. This constraint adds an extra condition to all of the transition rules concerning the minimum value in the map, below which the software will not be able to change the land use. This minimum value is obtained by sampling.

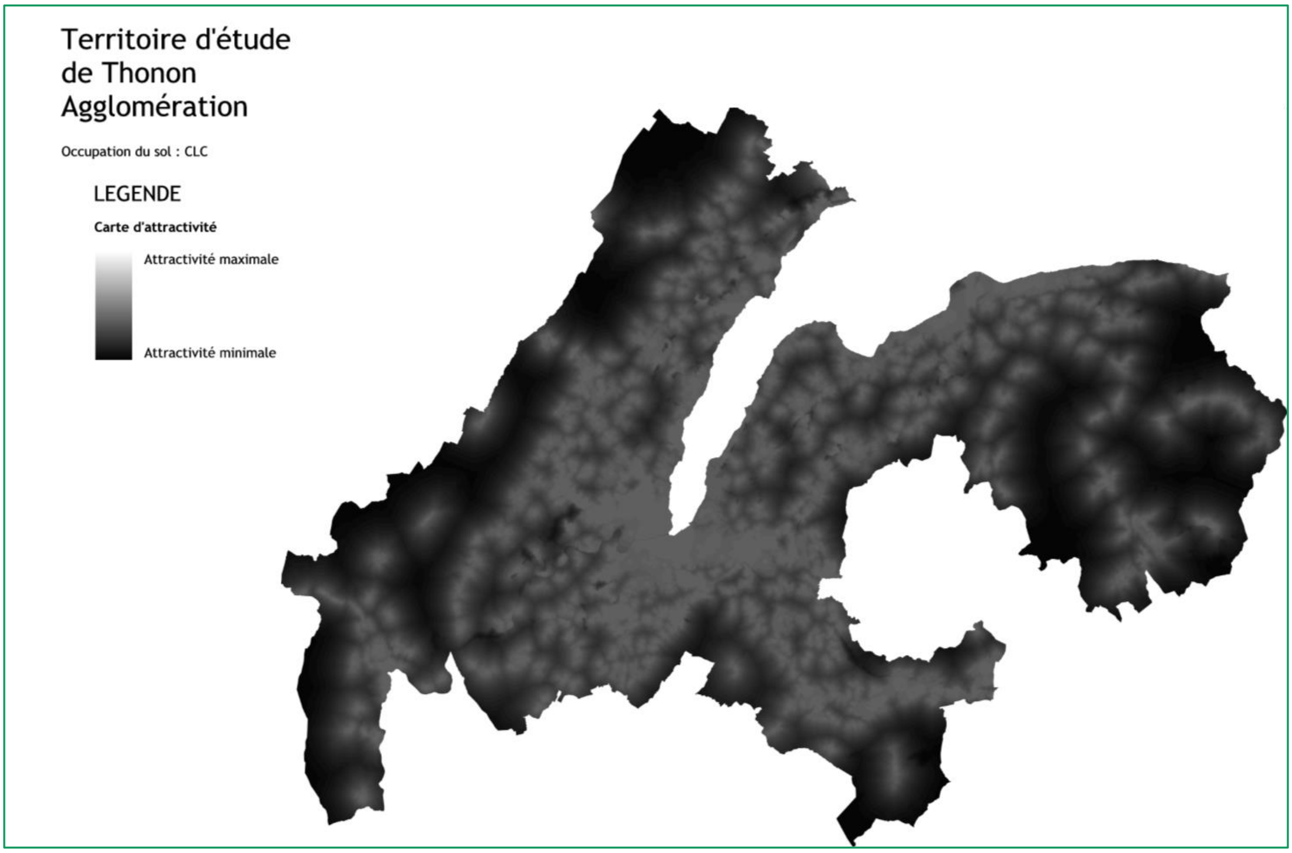


Figure 18: Attractiveness map produced with CLC for the strategic scenario in the Thonon Conurbation area

Next steps

Vector files (*.SHP format) can also be added to LUCSIM. They are used to display additional geographical information on the project layers. However, no processing can be run on these layers.

To add vector layers, users click *Add Layers* in the *Maps* window. Once the layer has been incorporated, users can change the way it is displayed (symbols, colours, etc.) by right-clicking the desired option in the top left-hand pane, named *Layers.*

This feature was not included in our tests of the LUCSIM software.

2.2.2.4 Analysing the information from different raster layers

Under LUCSIM, users can create statistical data from the incorporated raster images. There are several types of statistical tools for processing raster images in LUCSIM:

- Summary graph. From the *Display/Graph* menu, users can display a graph recapitulating the number of cells for each land use state (e.g. pie chart, histogram, etc.). The graph can be exported in *.png or *.csv format.
- Layer comparison (Matrix comparison). From the *Statistics/Layer comparison* menu, users can produce matrices comparing the various *State layers* included in LUCSIM, or even compare them with one of the images "simulated" by the software. A comparison matrix can be used to quantify the changes that occurred between two reference images: for each state in the first image, the matrix provides details of the values taken on by these pixels in a second image, which serves as a comparison. Users can export the image in CSV format or copy the chart's content directly to a spreadsheet.

Next steps

For more detailed statistical analyses, the *Statistics/Neighbourhood extraction* tool provides a view of the neighbourhood surrounding the cells that underwent a specific transition.

This feature was not included in our tests of the LUCSIM software.

2.2.2.5 Constraints

To calibrate the modelling of changes in land use, users can set constraints on the modelling process, using:

- A potential model, which is equivalent to factoring in raster data such as an attractiveness map *(Suitability map)*;
- A Markov chain, which limits changes in land use by setting the amounts of change that can occur, given the changes observed between two given dates.

These constraints can be applied to the model through the *Rules/Set constraints* menu. Note that, if this option is activated, users can no longer modify the rules entered in the *Rules editor* pane.

2.2.2.5.1 Potential model

The potential model is used to analyse the capacity for expansion of a particular type of land use, based on the concepts of spatial interaction. As such, it is equivalent to incorporating an attractiveness map. Potential models are calculated through the *Analysis* menu. The tool prompts the user to assign a weight to each type of land use. The higher the weight, the greater the attractiveness.

Once generated, the potential model can be incorporated as an additional layer in the decision tree. This may have an impact on the calculation of the transition rules, or it may be applied as a modelling constraint through the *Rules/Set constraints* menu. This constraint adds an extra condition to all of the transition rules concerning the minimum value in the map, below which the software will not be able to change the land use. This minimum value is obtained by sampling.

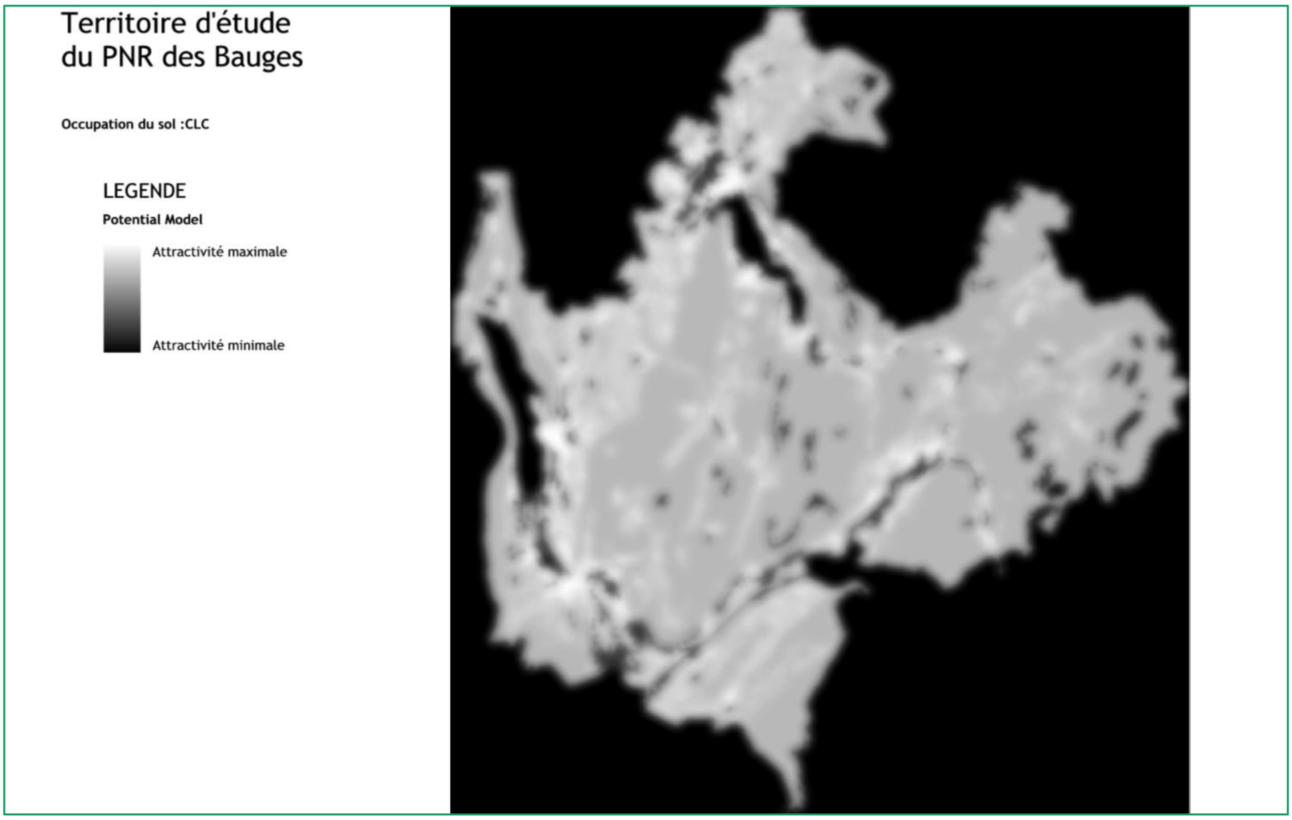


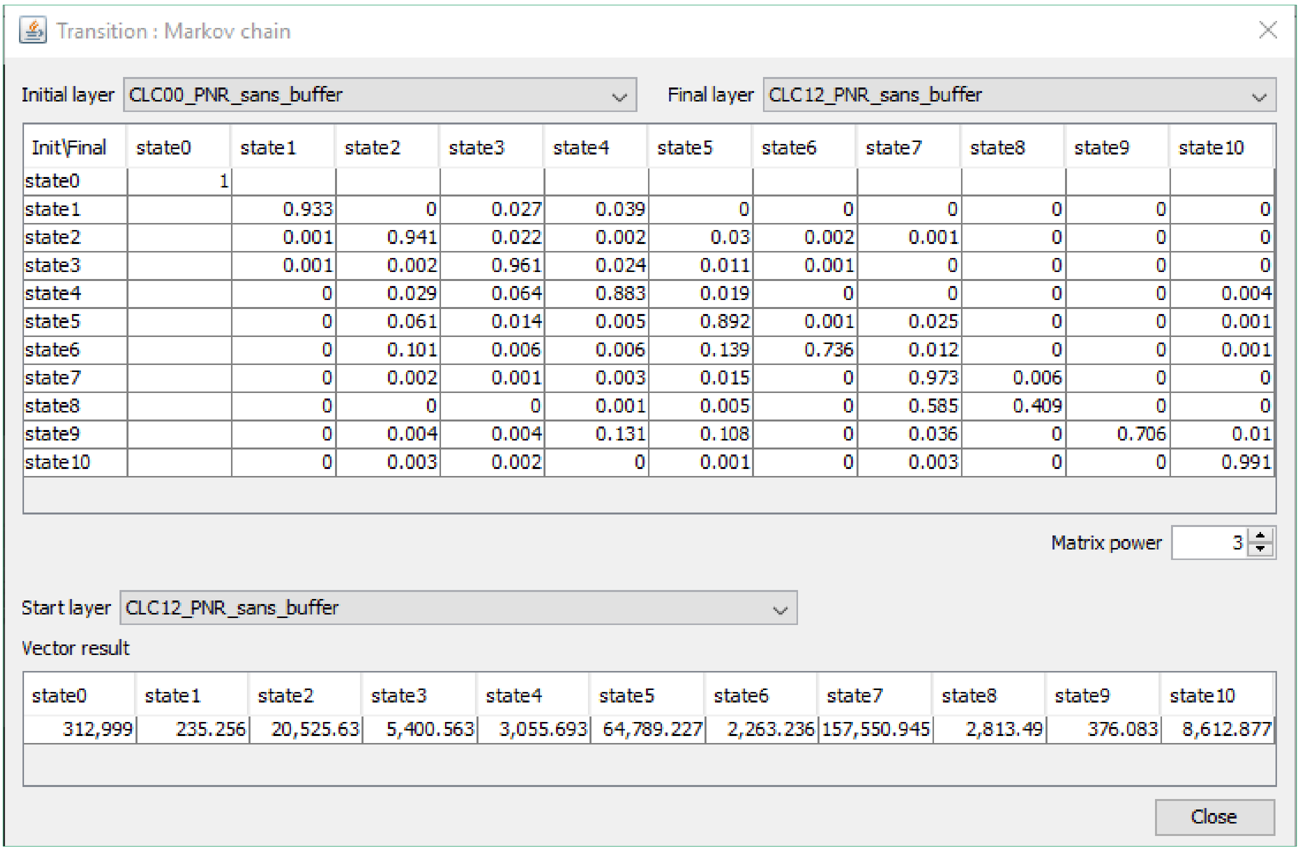
Figure 19: Potential model for the Bauges Regional Nature Park study area

2.2.2.5.2 Markov chains

Markov chains are a means of constraining the amount of change that can occur in a simulation over time. When applied to the transition rules, the cellular automaton's simulation stops if the number of cells stipulated by the Markov process is reached in one or more states.

To use this function, users must have two layers of land use from different points in time, then calculate the Markov chain through the *Analyse/Markov chains* menu. When the *Initial layer* and *Final layer* are selected in the dialogue box, a transition matrix is automatically calculated. This matrix contains the probability of a land use cell transitioning from one state to another in a simulation step, such as that observed between the initial layer and the final layer. A null probability indicates that the transition is impossible.

The *Matrix power* option lets users raise the matrix to the desired power, so that the changes observed between the two reference layers will be reiterated a specified number of times.



Init\Final	state0	state1	state2	state3	state4	state5	state6	state7	state8	state9	state10
state0	1										
state1		0.933	0	0.027	0.039	0	0	0	0	0	0
state2		0.001	0.941	0.022	0.002	0.03	0.002	0.001	0	0	0
state3		0.001	0.002	0.961	0.024	0.011	0.001	0	0	0	0
state4		0	0.029	0.064	0.883	0.019	0	0	0	0	0.004
state5		0	0.061	0.014	0.005	0.892	0.001	0.025	0	0	0.001
state6		0	0.101	0.006	0.006	0.139	0.736	0.012	0	0	0.001
state7		0	0.002	0.001	0.003	0.015	0	0.973	0.006	0	0
state8		0	0	0	0.001	0.005	0	0.585	0.409	0	0
state9		0	0.004	0.004	0.131	0.108	0	0.036	0	0.706	0.01
state10		0	0.003	0.002	0	0.001	0	0.003	0	0	0.991

state0	state1	state2	state3	state4	state5	state6	state7	state8	state9	state10
312,999	235.256	20,525.63	5,400.563	3,055.693	64,789.227	2,263.236	157,550.945	2,813.49	376.083	8,612.877

Figure 20 LUCSIM interface

To make these transition probabilities applicable, users must then incorporate them as a modelling constraint through the *Rules/Set constraints* menu. This adds a condition to the set of transition rules governing the maximum number of cells that can be attained for a given state (by using the *nbCell* function).

2.2.2.6 Writing transition rules

Transition rules are written with a special syntax that is specific to the software and uses 14 different functions. The software has a module to assist with creating these rules.

The introduction of rules is based on a complex syntax purpose-designed for LUCSIM. Basically, the syntax requires users to define an initial state and a final state, a transition function combined with two parameters (a land use and a distance in pixels) and a condition expressed as a percentage of land use or as a number of cells. For any one rule, several conditions can be combined with logical operators (and, or, etc.)[6].

The rules thus produced depend on geographical considerations and apply as follows:

- At any point in the area, the software will check whether a rule may or may not be applied, i.e. whether a pixel from the initial type of land use can develop into a second type of land use;
- For any give rule, the software will check each condition;

[6] LUCSIM software user guide

- Each condition is based on a count of the number (or proportion) of pixels of a specific type of land use, which is compared with a given value (lower, lower than or equal to, equal to, higher than or equal to, higher than);
- For any given rule, if all of the conditions are met, the pixel changes.

Next steps

Transition rules can be written by hand in the *Rules editor* pane, with or without the help of the *Rules generator* available in the same pane. They can also be generated "automatically" with the help of a decision tree through the *Rules/Creation/Decision tree* menu.

This feature was not included in our tests of the LUCSIM software.

2.2.2.6.1 Decision tree

Based on an analysis of the differences in land use between two different dates, the software calculates transition rules, then applies these rules to reproduce the changes that occurred over this period as accurately as possible. The rules thus created can be modified by hand.

To be able to use the decision tree, users need to have two maps in order to see how land use changed in the past and extrapolate its future developments. The *Rules/Creation/Decision tree* menu brings up a dialogue box in which users must configure the decision tree before launching the calculation. Selecting *Set parameters* brings up the dialogue box shown below.

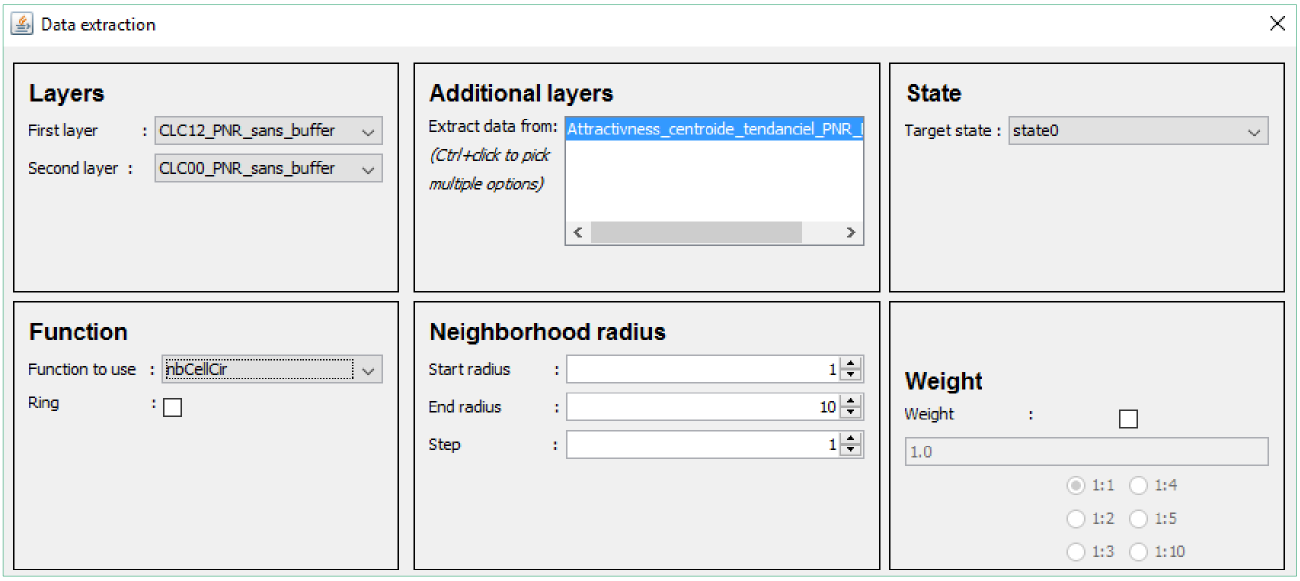


Figure 21 LUCSIM interface

Left to right and top to bottom:

- *Layers:* users must select the two reference layers from the drop-down list, *First layer* being the most recent edition.
- *Additional layers:* users can select different raster layers, which are not the reference land use layers and which, in principle, contain information that has an influence on the changes made in land use,

such as an exclusion map, a *Suitability map* (attractiveness map) or the potential model calculated by the software.

- *State:* from a drop-down menu, users must select the type of land use for which the decision tree will be calculated.
- *Function:* users must define the type of transition function that the decision tree will use to write the transition rules. The condition is expressed as a number of cells contained in a circle (*nbCellCir*) or in a square (*nbCellSq*).
- *Neighborhood radius:* users must define the neighbourhood in which the decision tree is to calculate the transition rules for the type of land use we are interested in, based on a list of the different types of land use present in that neighbourhood.
- *Weight:* when this option is enabled, this last pane allows users to increase the weight of the changes that occurred between the two reference layers. By enabling this function, uses instruct the decision tree to overvalue the number of changes that have occurred. In principle, though there is no guarantee, this helps make the rules generated more accurate. Experience gained at the Université de Bourgogne-Franche-Comté suggests that it is better to enable this function and that better simulations are obtained with a ratio lower than 1.

Once the various parameters have been entered, users can simply close the dialogue box and launch the calculation of the decision tree *(Start).*

The outcome is the following dialogue box:

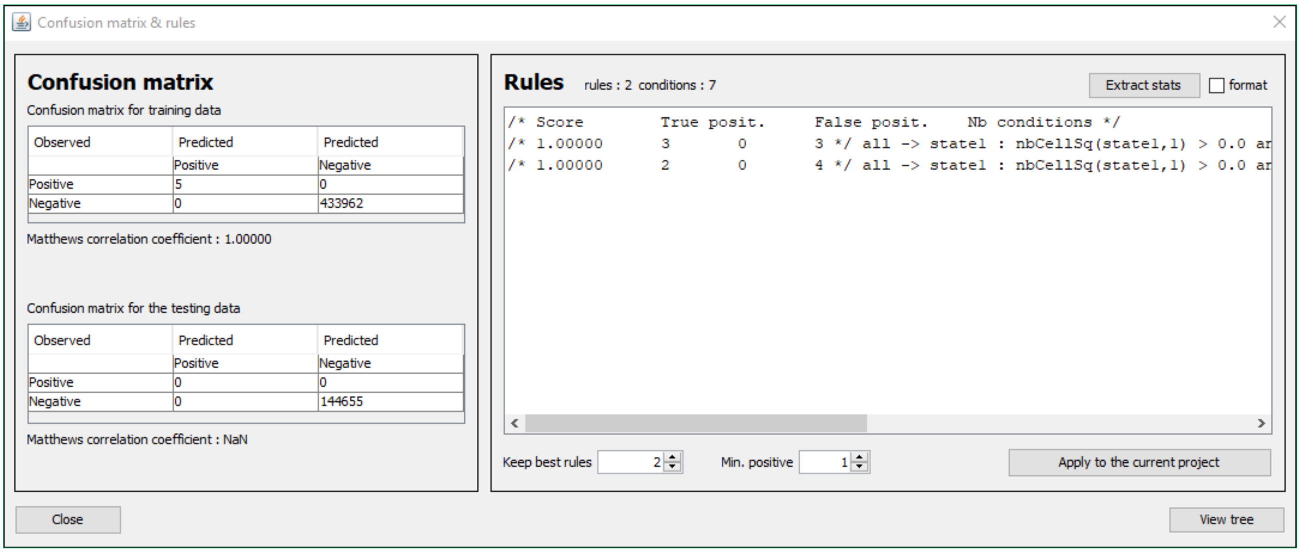


Figure 22 LUCSIM interface

This dialogue box contains all of the rules defined by the decision tree, along with the following information:

- *Confusion Matrix:* this matrix lets users compare the results obtained by modelling the reality, taking the oldest layer as the baseline. Based on this layer, the software runs a simulation of the changes in the type of land use being studied, based on the rules it has identified, and compares the results obtained with the most recent layer. The Matthews correlation coefficient, which is calculated by the software, is an indicator of the quality of the model produced: the higher the coefficient, the closer the simulation is to the reality.
- *Rules:* the dialogue box displays the rules calculated by the decision tree.
- The *Apply to the current project* button lets users enter the rules thus generated in the *Rules editor* pane. They can then copy and paste the rules from this pane to save them in a notepad application, for example. This is a useful procedure if users might subsequently want to produce a decision tree for a different category of land use, because the software does not retain any of the information: the *Apply to the current project* command automatically deletes all of the rules written in the *Rules editor* pane.

Next steps

The number of rules retained *(Keep best rules)* can be adjusted downwards, then returned to the original number if desired. This action has a direct impact on the Matthews correlation coefficient, which is automatically recalculated. Sometimes the coefficient is higher if fewer rules are kept. It is advisable, therefore, to retain the number of rules that will maximise the Matthews correlation coefficient.

It is possible to retain the rules that yield a minimum number of predicted pixels corresponding to the pixels observed, by adjusting the Min. positive parameter. As above, the aim is to maximise the Matthews correlation coefficient.

It is possible to display the decision tree (the *View tree* command), which appears in a new window. A right click selects a display mode: centred on the first node, fit to screen, automatic. Because the decision tree can be extremely complex, it may be necessary to zoom in on all or part of the tree. This is done by using the Maj key and drawing a frame around the chosen sector with the left mouse button held down. Analysing the decision tree is a particularly complex task, but may provide a better understanding of the rules generated, because it displays all of the conditions used as well as the order in which they were used. In other words, the higher up the decision tree, the more the conditions play a decisive role in structuring the rules generated by the decision tree.

To use the software in "automatic" mode, it is not absolutely necessary to check the content of the rules, especially as this can soon become very complex. There are, however, some points to watch:

- Because of the way it operates, the decision tree may be led to generate rules that change land use categories unnecessarily. This is because the analysis is performed pixel by pixel, and the software may interpret a very slight shift (in form or alignment, etc.) as a change in land use that has to be reproduced;
- The rules can potentially involve all of the land use categories, setting conditions for each of them that may prompt certain pixels to change. This might not make any sense whatsoever in terms of spatial organisation.

It is difficult to judge the quality of the rules generated, but users should bear in mind that, ultimately, the decision tree is only trying to reproduce all of the variations that occurred in a type of land use between the two reference layers, based on the neighbourhood parameters set by the user *(Neighborhood Radius).* Moreover, it is important to point out that the rules defined by a decision tree will use only one type of function: *bCellCir* or *nbCellSq*. In this respect, users have far greater latitude if they establish the rules by hand.

2.2.2.7 Output data

Once the transition rules have been established, the software is ready to model the changes that have occurred in the area.

There is no notion of time in the software. Each step of the modelling process comes to an end for one of the following reasons:

- At a given point, all of the pixels in an image can be changed;
- Not all of the rules can continue to be applied (i.e. the required geographical conditions are no longer fulfilled for the entire area);
- The user has placed certain restrictions on the amount of changes possible, for example by using Markov chains.

2.2.2.7.1 Creating simulations

The simulation bar can be used intuitively like an MP3 player to launch the simulation. From left to right, the controls are:

- Continuous (until the modelling comes to an end)
- Step by step
- Stop
- Reset to zero
- Simulation progress bar for each step (a counter moves in increments to the right)
- Synchronous mode
- Asynchronous mode

The automata has two operating modes: synchronous and asynchronous. In synchronous mode, all of the cells are run at the same time. "If the layer freezes, run all of the cells separately, then reassemble them." There cannot be any interaction between cells in the same simulation step. In the asynchronous mode, on the other hand, the cells run one at a time, so they can influence one other as the process unfolds.

Note that, to use the Markov chains constraint, the software must be run in synchronous mode. Otherwise, in asynchronous mode, the software may exceed the thresholds the user wants to reach with the Markov chains.

2.2.2.7.2 Simulations

The simulation maps can be exported in two formats - *.png and *.svg - in the *Export view* tab.

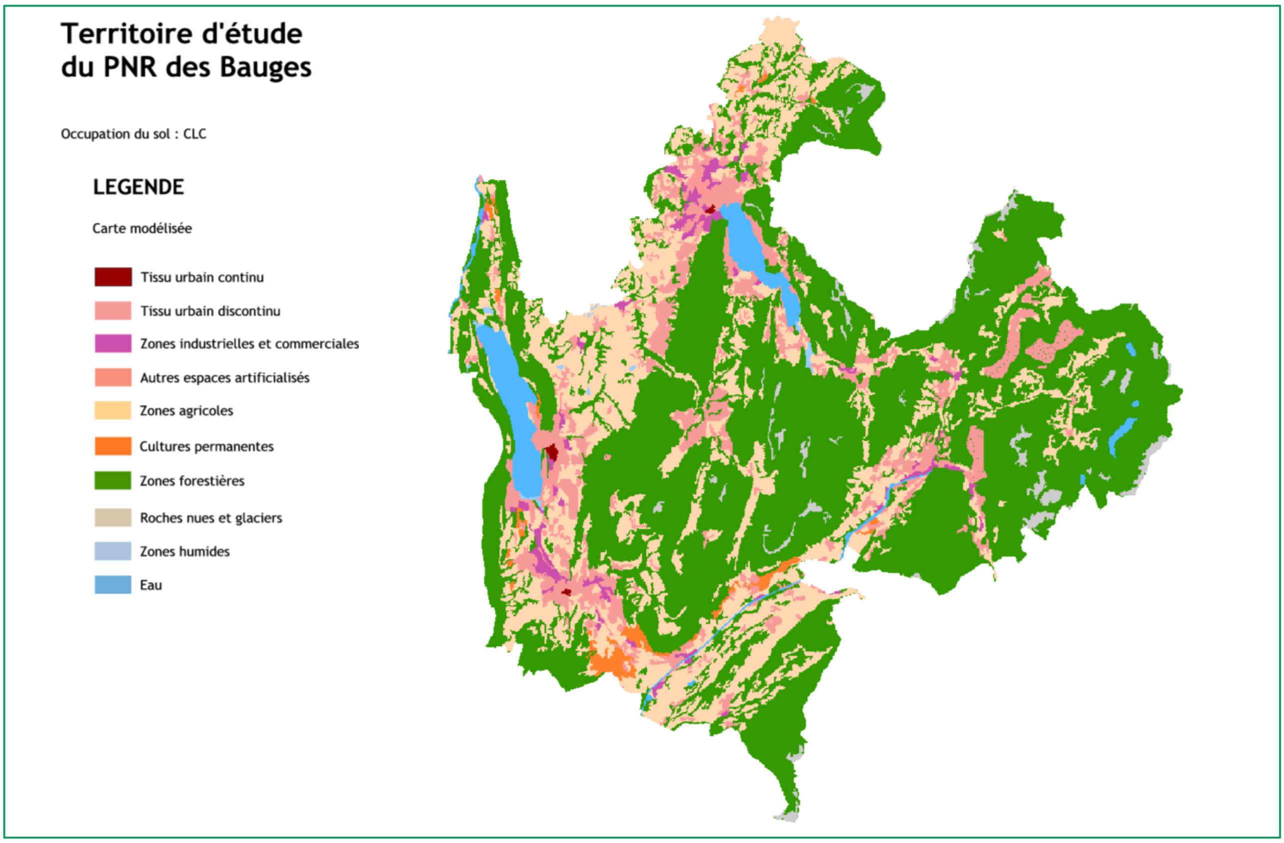


Figure 23: Map modelled with simplified CLC data and factoring in the strategic attractiveness map, for a rule search within a 30 pixel radius

3 Study programme – "What we planned to do"

3.1 Programme framework

In view of the ASTUS project's objectives and the way the two software programs to be tested operate, three key points guided the structure of the study programme assigned to the pilot sites:

- Use of different databases:
  - To ensure that the tests carried out on our test areas are transferable, at least one of the modelling tests must use European data sources, particularly for land use.
  - For more in-depth work on each of the areas, it is suggested that users use other data sources when possible.
- Modelling of different scenarios:
  - To be able to analyse the results obtained, it is important to remember that the purpose of these tools is not so much to represent a future image of the area as to enable a comparative spatial analysis of different simulations. The results are definitely not a prediction, which would be largely subject to debate as to why a particular sector was or was not developed.
  - For this reason, the objective should be to put forward different scenarios for the future development of the area, for use in each of the two software programs. Based on a handful of indicators, including land take and population growth, two scenarios should be drawn up: a baseline scenario and a strategic scenario. The scenario design should be informed by an analysis of the area's main strategic planning documents.
- Macroscopic modelling:
  - The FORESIGHT and LUCSIM software programs that will be used in the ASTUS project are both cellular automata tools. Because of the way they are built, these models cannot be used to model changes in land use on a very small scale (such as a land parcel or town district), since they do not reflect, in concrete terms, actual behaviour in the area. On the other hand, they are recognised for their ability to model land use changes on macroscopic scales.

3.2 Test strategy

3.2.1 Using FORESIGHT

We propose the following strategy for testing the FORESIGHT software:

- Use one or two land use databases:
  - CORINE Land Cover (CLC), 2012 edition
    - Land use database available with a spatial resolution of 100 m at European level. It has a detailed nomenclature broken down into three levels (five items for the first level, up to 44 items for the third level), available for the 1990, 2000, 2006 and 2012 editions.
  - OSCOM, 2013 edition (available only in the Auvergne-Rhône-Alpes Region)
    - Land use database developed by the DREAL Auvergne-Rhône-Alpes. Only one edition available to date. It has a detailed nomenclature, broken down into two levels (five items for the first level, 15 for the second level), and a spatial resolution of 10 m.
- Model a baseline scenario and a strategic scenario. These two types of scenario will be differentiated by the following features:
  - Land take:
    - For the baseline scenario, land take will be established by continuing the trends observed between 2000 and 2012 in CLC.
    - For the strategic scenario, land take will be determined on the basis of data from the area's various strategy documents.
  - The area's attractiveness map:
    - Two attractiveness maps, one for each of the two scenarios, will be drawn up without using the dedicated module available in the FORESIGHT software.
    - For the baseline scenario, the attractiveness of the various urbanised areas will be adjusted on the basis of recent demographic trends. The attractiveness of the natural, agricultural and forest areas (implying a notion that is the opposite of friction) will be adjusted on the basis of the pace of land take between 2000 and 2012.
    - For the strategic scenario, the attractiveness of the various urbanised areas will be adjusted on the basis of the regional framework defined in the area's strategy documents. The attractiveness of the natural, agricultural and forest areas will be adjusted on the basis of an appraisal of the level of protection of these various areas included in the strategic documents used.
  - The urban sprawl patterns:
    - For the baseline scenario, the formula for distribution among the various *Patterns* will be based on an analysis of the changes in the area between 2000 and 2012, using CLC.
    - For the strategic scenario, the formula for distribution among the various *Patterns* will be adjusted on the basis of an appraisal of the information contained in the area's strategy documents.

3.2.2 Using LUCSIM

We propose the following strategy for testing the LUCSIM software:

- Writing the rules "by hand" would require extensive discussion between the area's authorities and Cerema to agree on the rules to be modelled. Given that this task cannot be undertaken as part of the ASTUS project, the software will be used in "automatic" mode.
- Different land use databases will be used, bearing in mind that using the software in "automatic" mode requires access to at least two editions:
  - CORINE Land Cover, 2000 and 2012 editions, spatial resolution of 100 m;
  - High Resolution Layers (HRL), spatial resolution of 20 m;
    - This data is produced by the EU-funded Copernicus programme (which supplied the CLC data). The data is available in four raster layers (sealed areas, forests, pastures, wetlands and water bodies), with a level of definition significantly higher than that of CLC. There are several editions for sealed areas (2006, 2009, 2012 and 2015), but only one edition for all of the other types of land use (2015)[7].
  - Theïa
    - Theïa publishes a land use database for France as a whole, using Landsat 5 satellite data for the 2009, 2010, 2011 and 2014 editions, and Landsat 8 and Sentinelle 2 data for the 2016 and 2017 editions. The latter has a 10 m level of resolution (30 m for Landsat 5) and the layer is based on a 17-category nomenclature[8].
- The test program will model a "baseline" scenario calculated by the software and a "strategic" scenario adapted from the strategic scenario developed for FORESIGHT:
  - The "baseline" scenario will use the LUCSIM software in "automatic" mode. Modelling will be constrained by a Markov chain, sized to match the extent of urban sprawl in the baseline scenario drawn up for FORESIGHT.

The "strategic" scenario will also use the LUCSIM software in "automatic" mode. Modelling will be constrained by a Markov chain, sized to match the extent of urban sprawl in the strategic scenario drawn up for FORESIGHT, and by the attractiveness map drawn up for the same scenario. Since we are unsure of whether the model will take our attractiveness map into account, the *Potential model* will also be used to adapt our strategic scenario to the LUCSIM software.

[7] The 2012 edition has since been published on the Copernicus website on 21 June 2018.
[8] The 2016 and 2017 data has been available since June 2018 in dual raster and vector format.

4 Feedback – "What we managed to do"

4.1 Create an attractiveness map

4.1.1 Prepare the attractiveness map

To produce the attractiveness map without using the dedicated module available in the FORESIGHT software, we began by designing a "friction" map. Using the friction map, we assigned a value to each pixel in the area, based on our estimate of its permeability to urban development: the existing urban development pixels had the lowest friction values, while other areas (in particular water bodies) had a maximum value.

To design our friction map, we used the 2012 edition of CLC. With the latter, we were able to select objects according to the nomenclature and thereby distinguish between different areas, to which we were then able to assign different friction values.

Next steps

Insofar as these layers are handled in a GIS, it would have been possible to use other information to make an even clearer distinction between the friction values of the different areas comprising the study area:

- If we have the urban planning documents, we could then assign very low friction factors to the areas to be developed in these documents;
- If we have a layer showing the public easements, the different zoning classifications for the protection of natural areas, or an area's biodiversity protection areas (*Trames vertes et bleues* in French), we can then assign specific friction factors to these areas.

Note that, in principle, the level of definition of the input data for calculating the attractiveness map has little impact on the end result. Given the operation of the cost-distance algorithm used to draw up the attractiveness map, either with our own resources or using the dedicated FORESIGHT module, we obtain a certain data smoothing effect.

We wanted to be able to assign different values to the sealed areas in each commune in the area, so we began by dividing up the CLC database according to the boundaries of the communes.

For the baseline scenario, the attractiveness of the various urbanised areas was adjusted according to the recent demographic trends (between 2009 and 2014). For non-sealed areas, the friction factors were adjusted in accordance with the pace of land take observed between 2000 and 2012. The vast majority of land take occurs in agricultural and forest areas, with a smaller share in wetlands. Accordingly, we assigned a coefficient to each of these areas, while at the same time processing certain more specific land uses separately. Examples include permanent crops, which display virtually no change in our study areas, or bare rocks and glaciers, which do not change over time.

For the strategic scenario, the attractiveness of the various urbanised areas was adjusted in accordance with the regional framework defined in the area's strategy documents. At the same time, we agreed on a way to harmonise the level of the various main centres across our study area. For the friction factor of non-sealed areas, we adopted the following principles, while harmonising the information collected in our study areas:

- Major reinforcement of the protection of wet lands;
- Reinforcement of the protection of agricultural areas;

- No change in forest areas' friction factor: these areas are naturally better protected from the phenomenon of land take and there is little discussion of them in strategy documents, apart from biodiversity protection issues *(Trames vertes et bleues)*, so we decided not to change the forest areas' friction factor.

Next steps

After reading the strategy documents, we would have liked to make the attractiveness map more precise by taking into account other important factors, such as the *Trames vertes et bleues* or the definition of agricultural areas, which are more strategic and better protected in these documents. However, we were not able to undertake this within the framework of our tests. For example, we had considered adjusting it according to the level of protection provided for these different areas in the strategy documents, but the principle proved difficult to apply:

- We would first have to agree on the value of the level of protection actually provided by our different strategy documents;
- We would then have to build a prior land use layer, including all of the areas defined by the documents (a definition of the components of the *Trames vertes et bleues*, etc.).

We applied these principles and then designed the attractiveness map, using the cost-distance calculator available in ArcGIS. Two attractiveness maps (baseline and strategic scenarios) were produced this way.

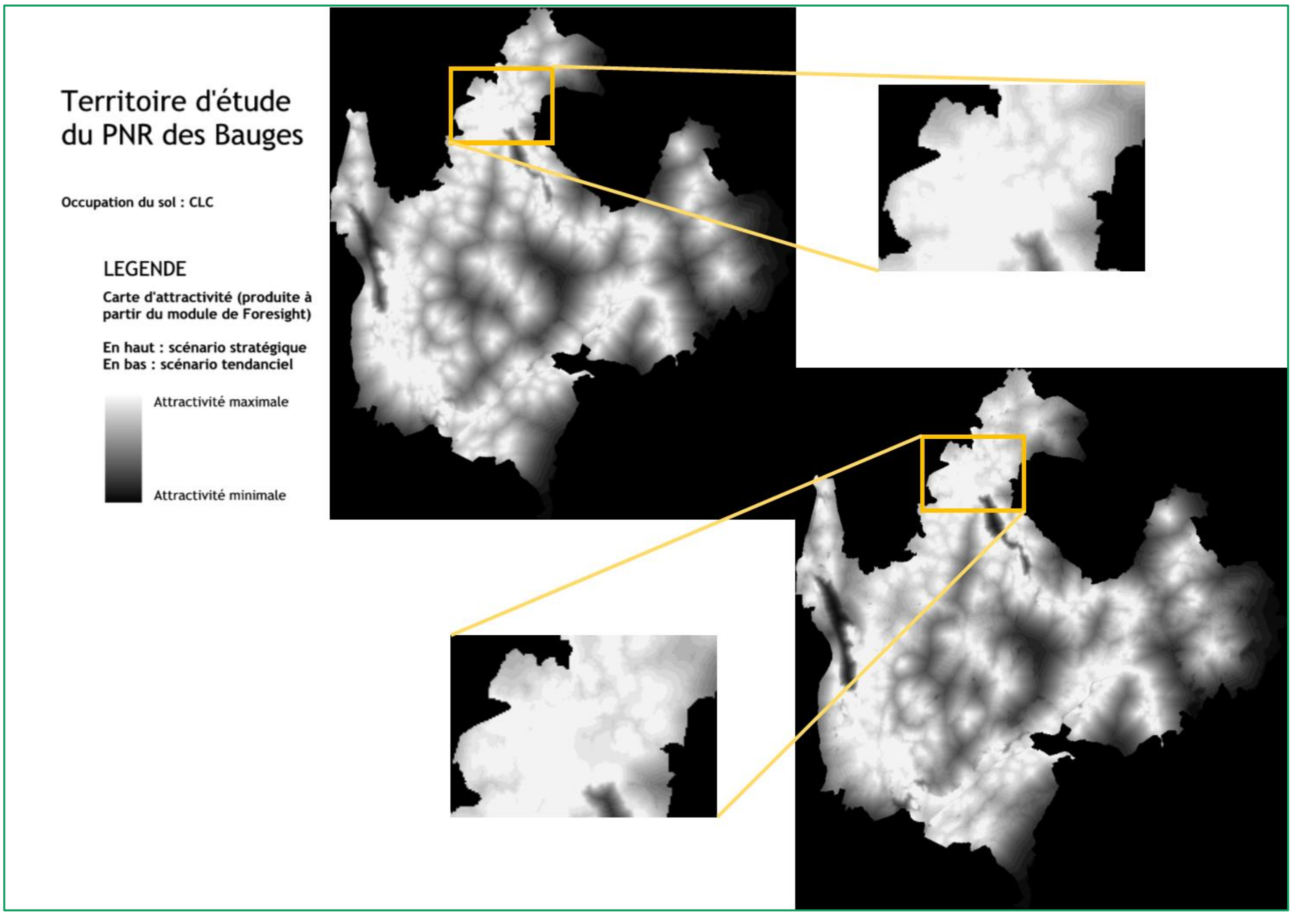


Figure 24: Attractiveness maps produced outside the dedicated FORESIGHT module for the Bauges Regional Nature Park study area and for the strategic and baseline scenarios (zoomed images)

4.1.2 Incorporating the attractiveness map into FORESIGHT

Once the various attractiveness maps had been prepared, we wanted to incorporate them into our input data in FORESIGHT. However, we were unable to do so, despite the existence in its pre-processing module *(Pre-processing* tab) of a tool for converting raster data into a source file that the software can use[9]. Since no solution was found within the WP2 time frame, we adapted our method and used the FORESIGHT software's dedicated module to produce our attractiveness maps.

This did not have any impact on non-sealed areas. However, it was no longer possible to distinguish between sealed areas, either by the demographic trends or the urban development framework. We therefore introduced a further distinction between our two "baseline" and "strategic" attractiveness maps by using two separate layers to define the points from which the cost-distance algorithm is developed and the "cost" of urban development is estimated throughout our area. To create the attractiveness map for the baseline scenario, we then placed our "points" on the communes with the strongest demographic trends and, for the strategic scenario, on the main communes in the urban development framework. While the result yielded by

[9] After further testing, we were able to incorporate one of our attractiveness maps in FORESIGHT. There is no way to do this through the *Pre-processing* tab. However, it was possible to generate the *.gif file for one of our attractiveness maps directly through ArcGIS. Users will need to ensure that the resolution and coverage are strictly identical to those in the files generated by FORESIGHT in the *Input* folder for the *Initial Urban Map* data, for example. The file was then copied into a model's *Input* folder, where it was recognised by the software *(Preview* in the *Initial settings* tab), then used to generate a model based on our own attractiveness map.

this processing was not absolutely identical to the one we obtained ourselves, we had something similar that enabled us to continue our testing of the FORESIGHT software.

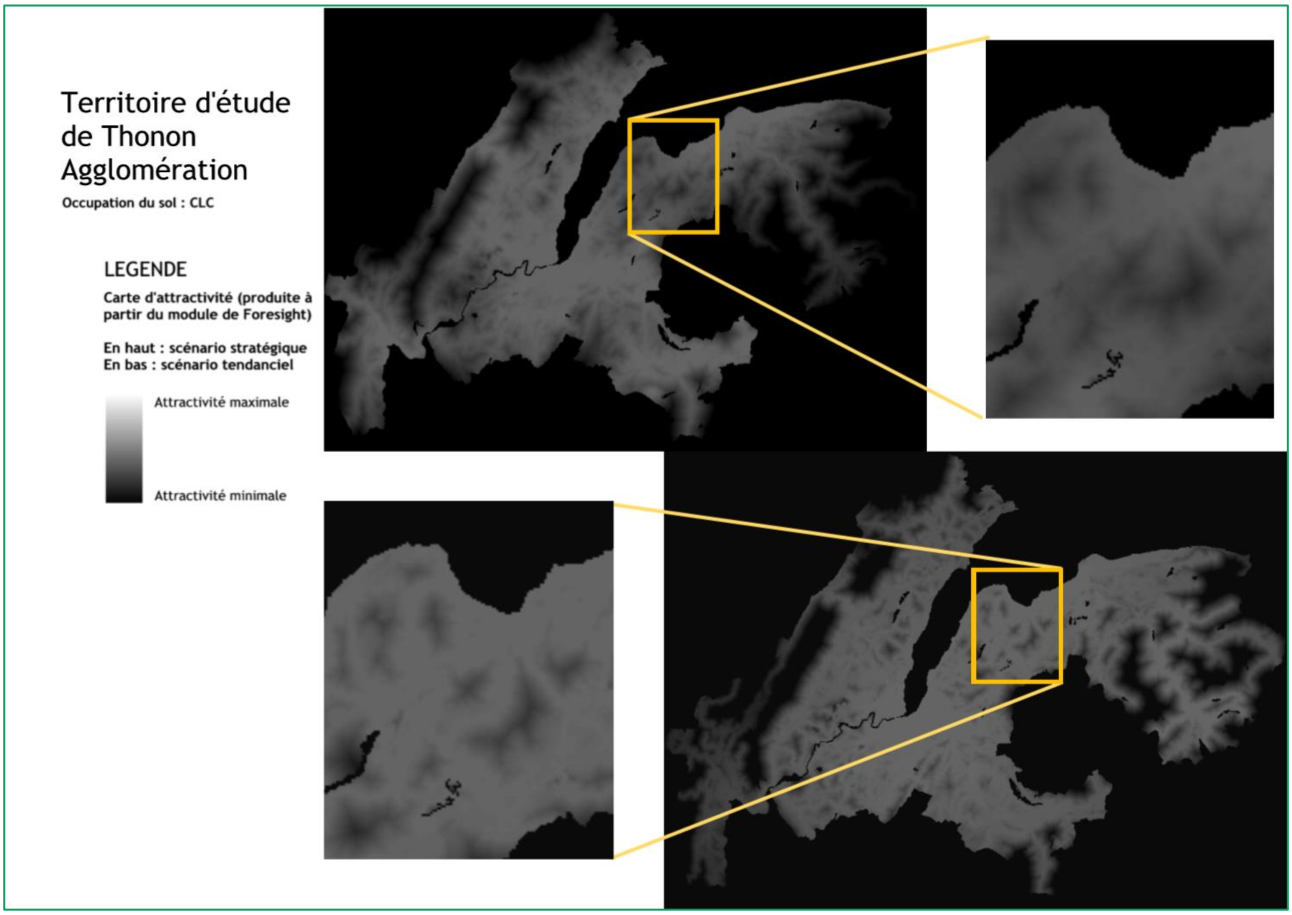


Figure 25: Attractiveness maps produced with FORESIGHT for the Thonon Conurbation study area and for the strategic and baseline scenarios (zoomed images)

4.1.3 Incorporating the attractiveness map into LUCSIM

We incorporated the strategic scenario's attractiveness map into the software in order to influence the modelling. Our attractiveness map was recognised by the LUCSIM software and used for the following two purposes:

- As an additional information layer for defining rules in the decision tree;
- As a constraint for simulation (since it was for the potential model provided by the software).

Since we were unsure of the map's impact on our models, we also used the potential model available in LUCSIM.

Our attractiveness map sometimes had an identifiable impact on the decision tree's expression of the rules, but the main impact was on the model itself.

4.2 Testing with FORESIGHT

4.2.1 Data preparation

4.2.1.1 The level of definition of the input data

The tests were carried out with a pixel size of 100 metres square for creating the input data. This is the default value in FORESIGHT, mainly because of the level of definition of the source data for the *Hillshade* and *Slope Map* maps.

The OSCOM data on the Bauges Regional Nature Park study area uses a smaller resolution, so when it was incorporated, the question arose of using the same minimum mapping unit as in FORESIGHT. For the reason given above, we retained a pixel size of 100 metres square, which had the direct result of making the OSCOM data less precise.

4.2.1.2 Initial Urban Map based on CORINE Land Cover and OSCOM

For our tests of FORESIGHT, there was, in principle, no particular difficulty in using CLC. We accordingly used the 2012 edition of the CLC data to obtain the *Initial Urban Map*, selecting all of the sealed areas according to the CLC nomenclature.

For the OSCOM database, we used the FORESIGHT option that allows users to incorporate specific data. We therefore prepared the vector layer beforehand by creating a numerical field in which we assigned the value 1 to all of the areas classified as sealed in the OSCOM nomenclature and 0 to all other types of land use. Another detail is that OSCOM includes the entire road network in its definition of sealed areas. We accordingly decided to restrict the sealed areas in the OSCOM database to the urban footprint, which we established on the basis of the IGN topographic database.

We were able to use OSCOM for the Bauges Regional Nature Park study area. On the other hand, because OSCOM is not available for the Swiss section of the Thonon Conurbation study area and we did not have a land use layer equivalent to that of OSCOM for the Swiss side, we were only able to perform CLC-based modelling for this pilot site.

We noted that, even when the OSCOM data had been reprocessed by FORESIGHT (increased pixel size), some of the precision of the original data was intact and the data actually proved to be more detailed than the CLC data.

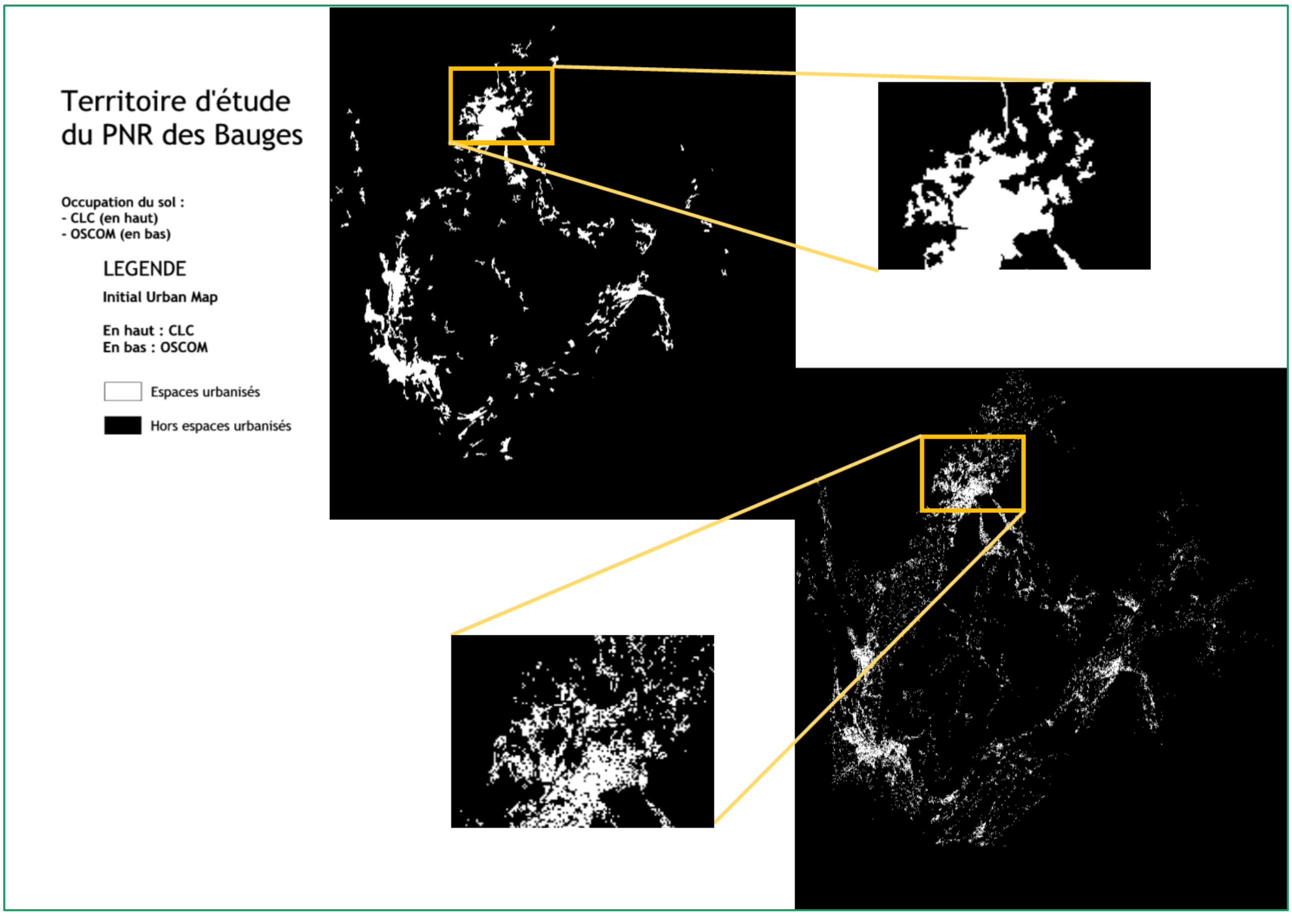


Figure 26: Initial Urban Maps generated for the Bauges Regional Nature Park study area using CLC (top) and OSCOM (bottom)

4.2.1.3 Slope Map and Hillshade using EU-DEM

To create these two maps, we used EU-DEM data supplied by the EU-funded Copernicus programme. It is a digital terrain model (DTM) available for all European Union countries. This hybrid product is based on SRTM and ASTER-GDEM data, and is available in the form of raster tiles in *.hgt format. Because our study area straddled two tiles, we had to begin by combining them in a GIS, either by merging them or by creating a single, virtual raster from the two tiles, which was the simpler solution. Because the tiles were in an identical format to that of the SRTM data, they could be incorporated directly into FORESIGHT.

4.2.2 The minimum mapping unit for modelling

The default minimum mapping unit in FORESIGHT is 100 metres square, so we retained this parameter, whatever the modelling process carried out. We made this decision to maintain consistency with the input data, created with an identical level of precision. As mentioned in the detailed description of the FORESIGHT software's operation, the program's inventor advises users to confine modelling processes to minimum mapping units of between 50 and 200 metres square.

4.2.3 Land take

Land take was defined for the baseline scenario by simply using the same figure as recorded in CLC between the 2000 and 2012 editions.

The use of the 2006 edition raises questions insofar as, for both of our study areas, very few changes were made to sealed areas between 2006 and 2012, especially by comparison with the 2000-2006 period. However, in similar periods, the change we can see in the urban footprint, which we calculated on the French section of our study areas, does not corroborate this sudden decrease in land take.

Compared with the rate of population growth in similar periods, which has remained steady over time, it raises another question, which also prompts us not to use the results from the 2006 edition. This sudden decrease in land take between 2006 and 2012, by comparison with the previous period, is only possible if there is a real change in the way land is developed and used. Yet this would not appear to be the case.

For the strategic scenario, we identified the rate of land take predicted by each of our strategy documents for their own period of action. We then arbitrarily used the same figure to calculate land take for the strategic scenario[10].

4.2.4 Estimating the Patterns for the baseline and strategic scenarios

To be able to distinguish between our two "baseline" and "strategic" scenarios, we use different *Patterns* to distinguish the forms of urban sprawl generated by FORESIGHT.

On the advice of the University of Rennes, we tried to establish a relatively simple method for determining the baseline scenario's *Patterns* by drawing on the analysis of past trends. We applied our method using not CLC data but an urban footprint we calculated on the basis of the topographical database. We have two editions of this database, 2008 and 2017.

Our method can be broken down as follows:

- Based on the 2008 urban footprint:
    - Create a buffer zone around the main entities in the urban footprint:

The buffer zone is intended to help us distinguish between urban scattering (the *Spontaneous* and *New spread center* patterns) and other phenomena that occur near already developed areas (the *Edge-growth* and *Road-influenced* patterns).

Even so, urban scattering cannot really occur in the middle of nowhere. If we had placed a buffer zone around the entire 2008 urban footprint, all of the urban development created between 2008 and 2017 would lie in that buffer zone. However, we thought it was important to make a distinction between a type of urban scattering that occurs not in the middle of nowhere but around tiny urban entities (such as farms, small groups of houses, small hamlets, etc.) and other, far more frequent phenomena that occur around the main urban entities. This is what prompted us to select only areas larger than a hectare to conduct our testing. We then applied a fixed, 100-metre-wide buffer zone.

[10] The detailed values are given in the two reports on the specific areas.

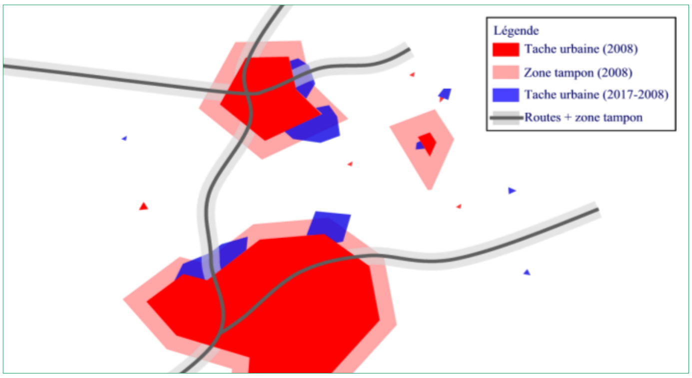


Figure 27: buffer zone around the urban footprint

- Based on the difference between the 2017 and 2008 urban footprints:
  - Outside the buffer zones described above:
    - Determine the *Spontaneous* pattern: select the smallest entities (the tests are run on entities of less than 5,000 m²);
    - Determine the *New spread center* pattern: select the largest entities;
  - Inside the buffer zones described above:
    - Determine the *Road-influenced* pattern: select entities inside a buffer zone around the road network. It is important to take a critical look at the *Road-influenced* pattern of linear development, which we have set out to describe. If we take the road network as a whole, virtually all of the urban development created between 2017 and 2018 lies near that network. To avoid overvaluing this pattern, we confined ourselves to the primary network, to which we applied a 500-metre-wide buffer zone.
    - Determining the *Edge-growth* pattern: by deduction, we assigned all of the remaining entities (inside the buffer zones calculated on the basis of the 2008 urban footprint) to this pattern.

We then calculated the ratio of the areas of each of our entities to the total area of the entities created between 2008 and 2017, giving us an indication of the suitability of our patterns.

To set the patterns of our strategic scenario, we significantly reduced all of the urban scattering by increasing *Edge growth.* The strategy documents all reflected a move to curb linear development and urban scattering, and tended to expand the main existing urban entities[11].

4.2.5 The models

For each of our two areas, for each of the two databases we used to define the *Initial Urban Map* (CLC and OSCOM) and for each of our two "Baseline" and "Strategic" scenarios, we produced the following:

- A single model of the area as a whole;
- A series of 20 models based on the option available in the *Output* tab;
- A map of the probability of urban development for each scenario, based on the *Future uncertainty* option and using the 20 models produced earlier.

[11] The detailed values are given in the two reports on the specific areas.

4.2.6 The results

Our testing of the software yielded two types of output from FORESIGHT:

- The map of the new urban development;
- The probability map.

4.2.6.1 Examples for the Thonon Conurbation study area

The urban development created by FORESIGHT covers 15,500 ha for the baseline scenario and 9,000 ha for the strategic scenario[12].

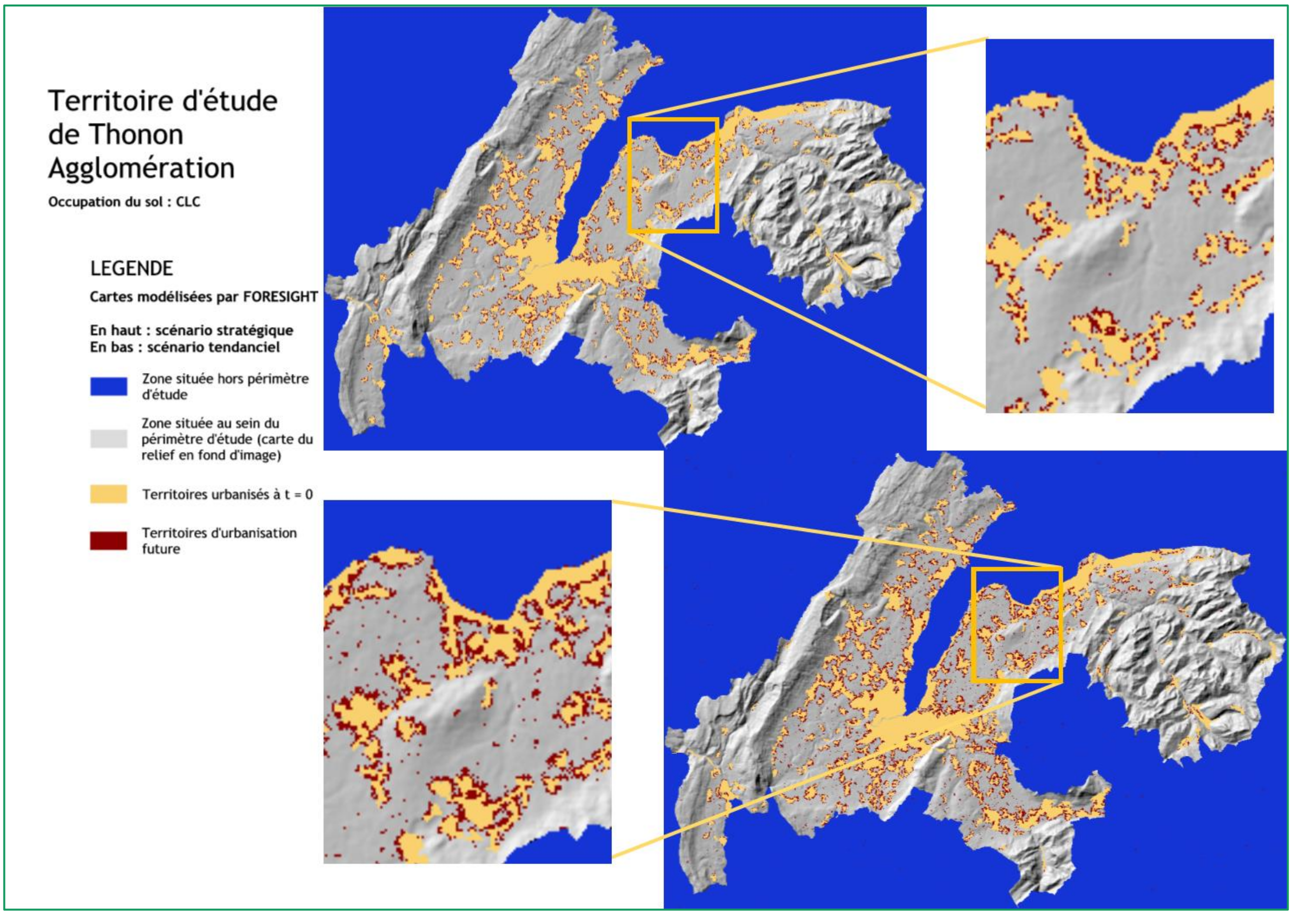


Figure 28: Models obtained for the Thonon Conurbation study area using CLC data and for the strategic and baseline scenarios

[12] An explanation of the values used can be found in the two reports on the study areas.

The probability maps were obtained from 20 models produced for each of our scenarios.

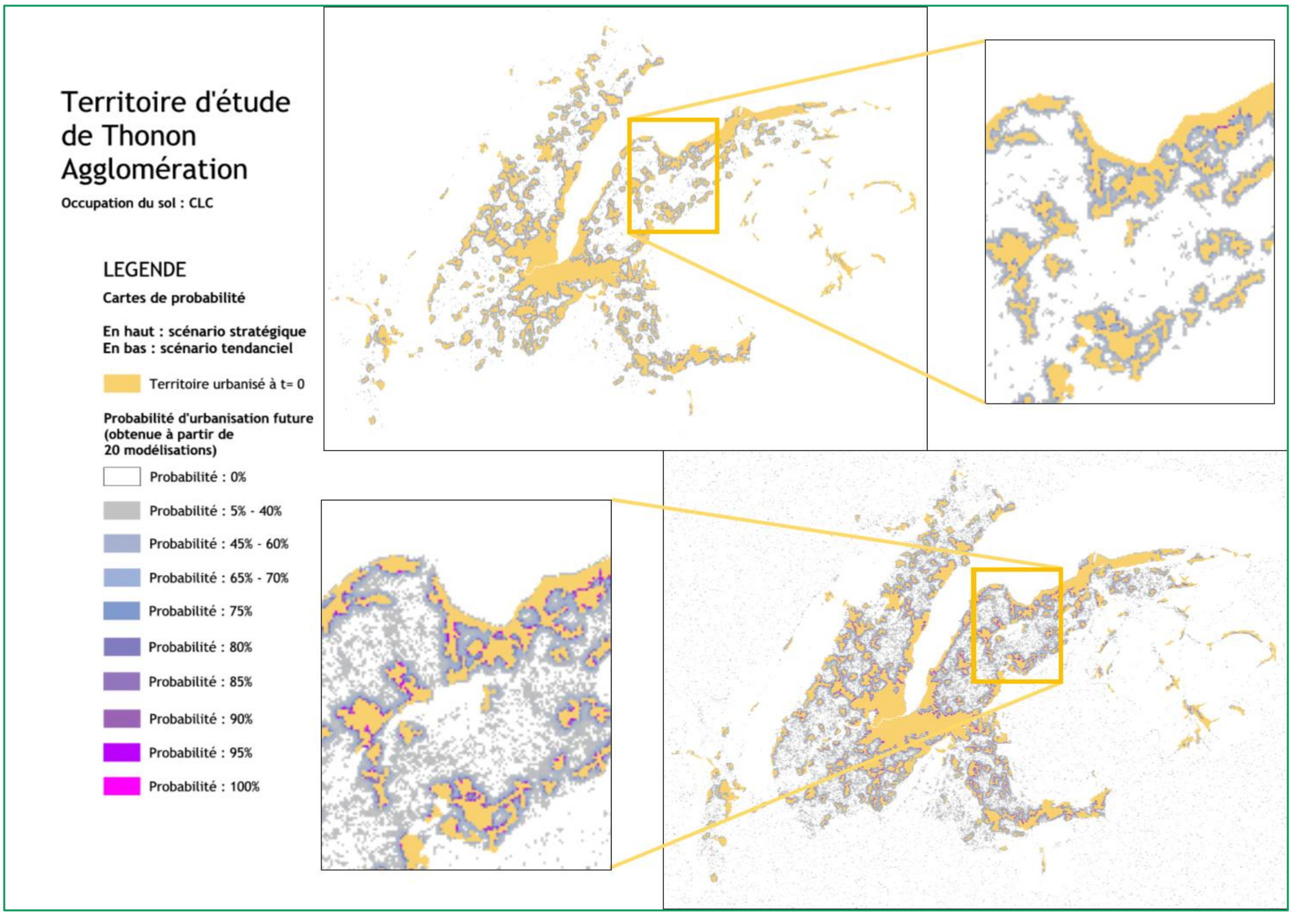


Figure 29: Urban development probabilities based on 20 models of the Thonon Conurbation study area using CLC data and for the strategic and baseline scenarios

4.2.6.2 Illustrations for the Bauges Regional Nature Park study area

The urban development created by FORESIGHT covers 7,000 ha for the baseline scenario and 8,000 ha for the strategic scenario[13].

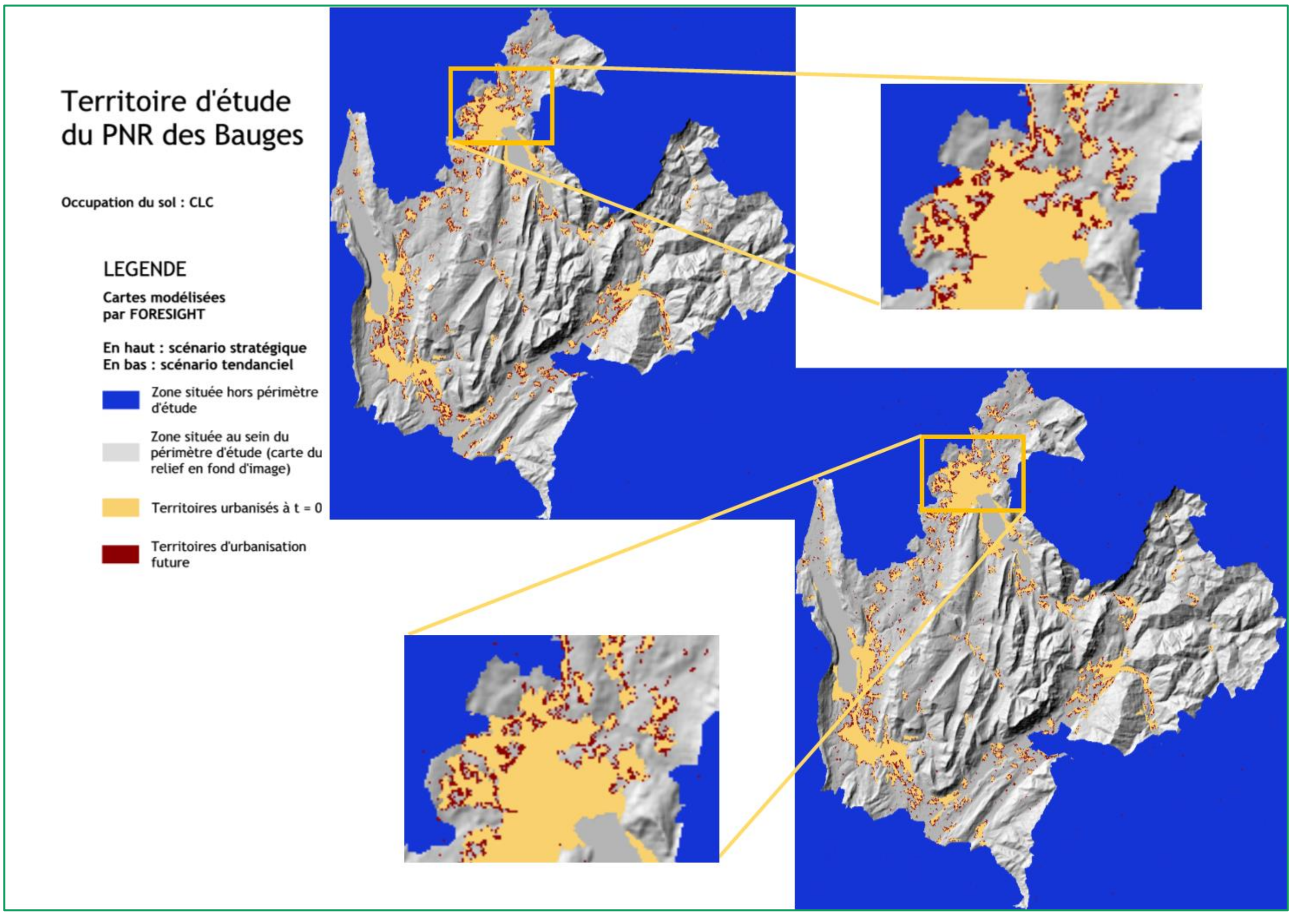


Figure 30: Models obtained for the Bauges Regional Nature Park study area using CLC data and for the strategic and baseline scenarios

[13] An explanation of the values used can be found in the two reports on the study areas.

The probability maps were obtained from 20 models produced for each of our scenarios.

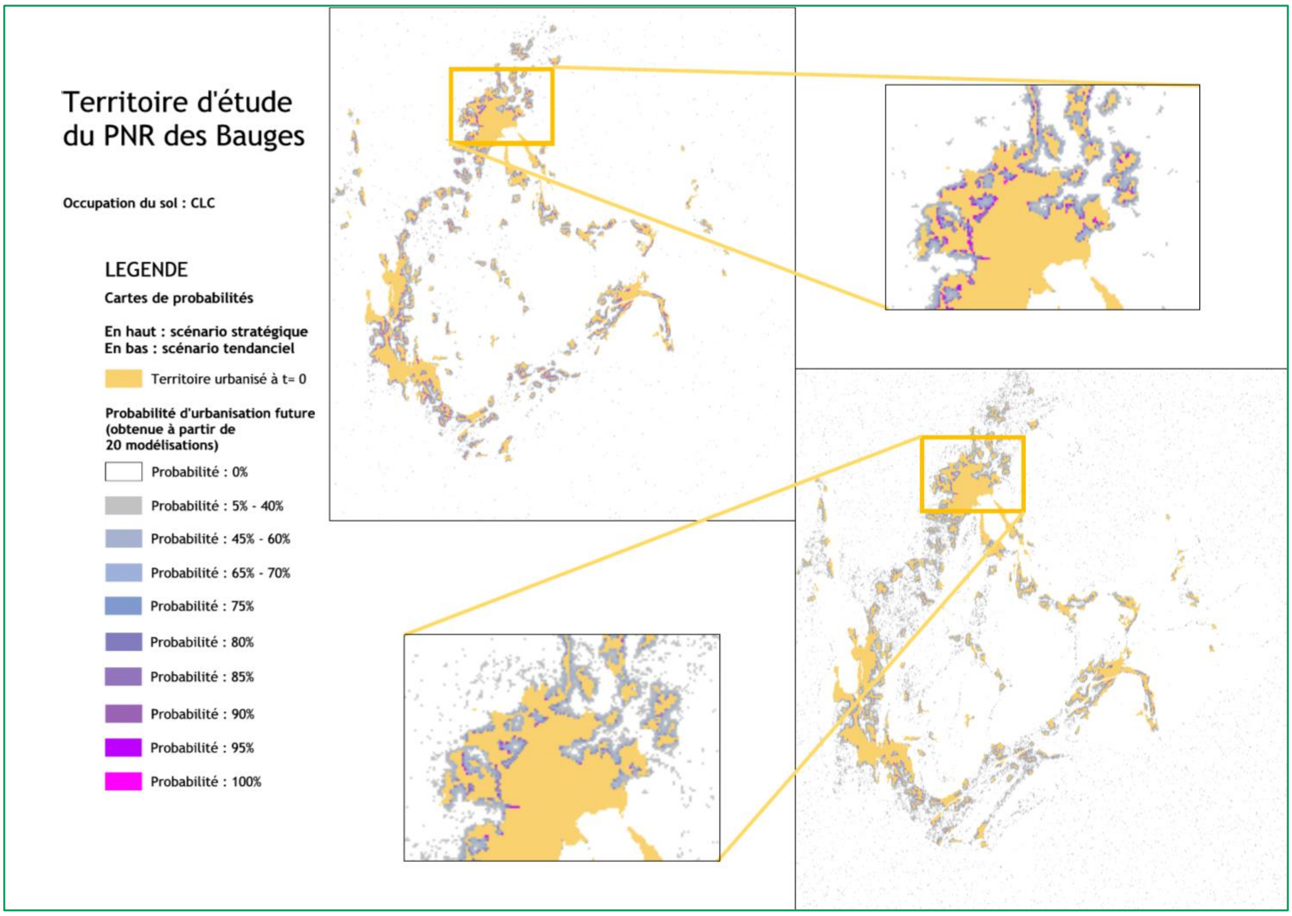


Figure 31: Urban development probabilities based on 20 models of the Bauges Regional Nature Park study area using CLC data and for the strategic and baseline scenarios

4.2.7 Analysis of the results

4.2.7.1 Analysis of the modelled maps

Because each of the maps produced with FORESIGHT is in *.gif format, they must first be geolocated so that they can be analysed in a GIS.

Our analytical strategy is then to count the different types of pixel, based on various regional viewpoints.

To be able to compare each scenario's impact on the models produced by FORESIGHT, our analytical strategy also entails analysing the maps from two regional viewpoints:

- Based on the breakdown into groups of communes;
- Based on the area's development framework, defined by analysing the area's strategy documents.

Our objectives are to confirm the differences between our scenarios with regard to land take and the way the area is developed, proof that changing the *Patterns* does indeed have an impact on the models.

In the Bauges Regional Nature Park study area, the maps are modelled with CLC and OSCOM data. Our objective is to analyse the results obtained in order to see how the data source influences the models.

The analysis performed according to the area's development framework is aimed at confirming the impact of the attractiveness maps on the preferred location of the new urban development pixels in the area.

Any other regional breakdown could have been used to carry out this analysis. While there would be little point in analysing the exact location (a specific land parcel, for example) of the pixels generated by FORESIGHT, an analysis by geographical sector is far more meaningful.

4.2.7.2 Analysis of the probability maps

The probability maps provide additional information not contained in the modelled maps. By enabling us to identify sectors in which the probabilities of urban development are very high for a given scenario, they provide a far more macroscopic approach than with a map simply modelled by FORESIGHT. Probability maps also make it easier to see urbanisation trends, whether linear development or the agglomeration of several developed areas (gradually filling in certain vacant spaces).

For all of these reasons, these maps would be worth analysing in greater detail than we were able to do, because we used the same type of processing for these maps as for the modelled maps.

This analysis nevertheless confirms certain outcomes:

- The influence of the *Patterns*: the *Spontaneous* and *New spread center* parameters are typically random phenomena whose effects on urban development are not systematic. In the probability map, this is reflected in a proliferation of pixels with relatively low probabilities of urban development. The *Edge growth* parameter, on the other hand, will systematically develop the perimeter of all of the entities initially developed, in all of the models. In fact, we will be able to see very high probabilities of development in the immediately adjacent areas.
- The influence of the databases: in the Bauges Regional Nature Park study area, we can see that the initial urban development defined with OSCOM differs to that defined with CLC, particularly in the way the area is divided up. Where CLC yielded extensive areas of urban development, we have far more precisely defined developed areas with OSCOM (despite using a 100 m square minimum mapping unit). Moreover, in some sectors with sparse development, OSCOM identifies developed areas that CLC failed to register. Even if no detailed analysis of these differences was undertaken during the current testing, we can suppose that this degree of precision in the information had an impact on the modelling, since the *Edge growth* parameter is directly influenced by these differences.

4.3 Testing with LUCSIM

4.3.1 Preparation of the land use layers

4.3.1.1 For CORINE Land Cover, the number of land use categories was reduced

For our tests of the LUCSIM software, we used the CLC database, which distinguishes up to 44 types of land use at the third level of its nomenclature.

The level of detail in CLC raises the question of the type of rules that the tool could create, based on the decision tree. From our viewpoint, the great variety of the CLC nomenclature makes each rule constructed by the decision tree extremely specific and detailed (for a cell to change, it will take so many cells from each land use). Accordingly, there is a considerable risk of making the rule inapplicable, or of making it applicable to only part of the area, for no reason, in principle.

On the advice of the Université de Bourgogne-Franceh-Comté, we accordingly deliberately limited the number of land use categories for each of the various databases we intended to use. To do so, we created a numerical field in which we simplified the CLC nomenclature, reducing it to 10 items.

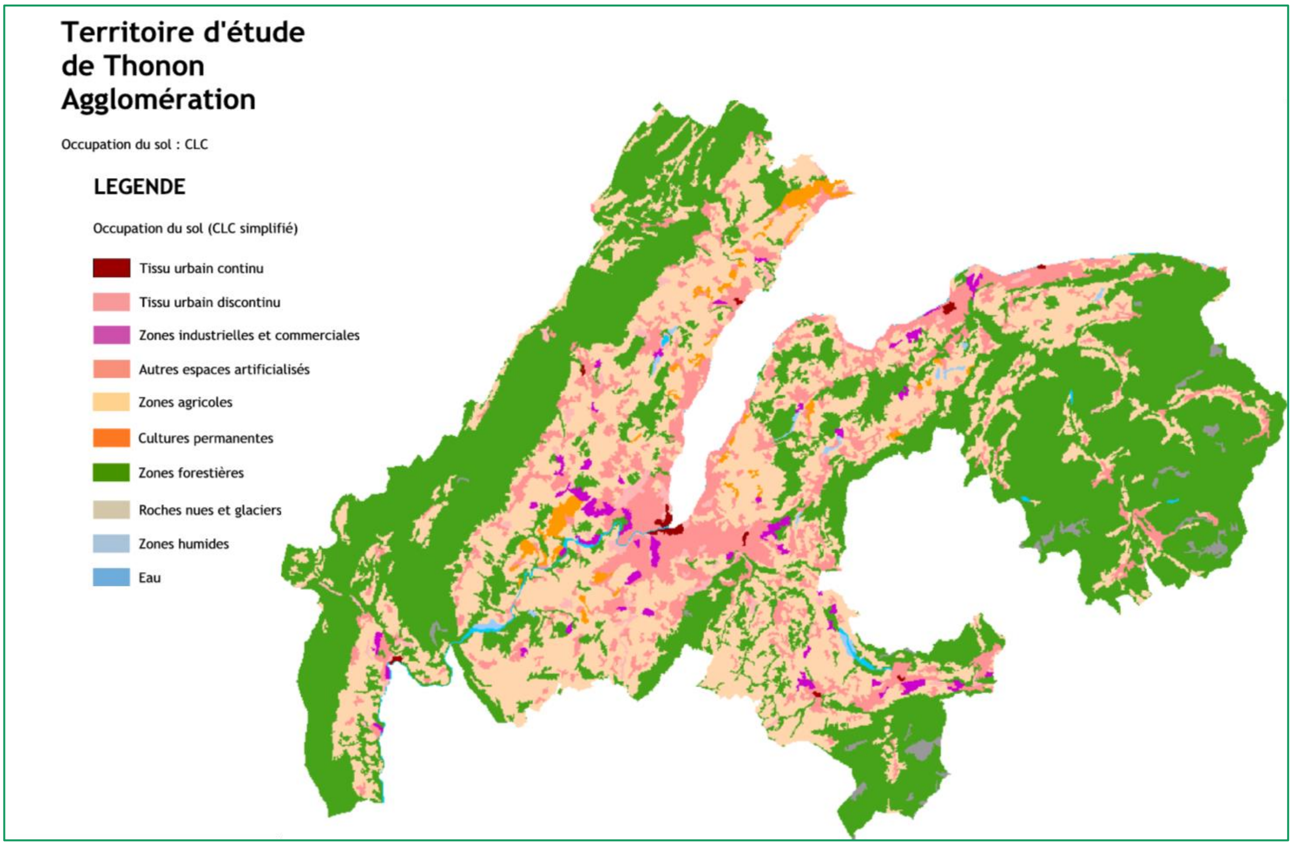


Figure 32: Land use (simplified CLC) in the Thonon Conurbation study area in 2012

4.3.1.2 For the High Resolution Layers database, how complete is the information?

To test LUCSIM, we used the High Resolution Layers (HRL) database published by the EU-funded Copernicus programme. This database offers four raster layers: "Imperviousness", "Forest", "Grassland" and "Water and Wetness". There are several editions for sealed areas (2006, 2009, 2012 and 2015), but only one edition for all of the other types of land use (2015)[14].

Various types of information are available in the different raster layers:

- *Imperviousness:* the value of the pixels indicates a level of soil imperviousness between 1% and 100%;
- *Forests:* when over 10% of the land examined is covered by forest, the value of the pixels indicates whether the forest is primarily coniferous or deciduous;
- *Grassland:* the pixels all have the same value;
- *Water and Wetness:* the value of the pixels indicates different categories: (1) permanent water, (2) temporary water, (3) permanent wetness and (4) temporary wetness.

For 2015, we have the four layers of HRL data, but assembling the information still does not produce a land use layer that covers the whole area. This is because the layers do not provide information about areas that are either too high or in the shade. (The information is obtained by automatic analysis of satellite images, but some areas were situated in the shade, so were not analysed.) For 2006, we have only the *Imperviousness* layer for the HRL data.

We therefore decided to work with only the *Imperviousness* data for 2006 and 2015 (plus the 2015 *Water and Wetness* layer, to avoid unwanted developments in these areas). We also simplified the *Imperviousness* data by creating four land use categories determined by the level of soil imperviousness: 1-49%, 50-69%, 70-89% and 90-100%.

However, working with only the *Imperviousness* layer means that the rules generated by the decision tree depend solely on that one layer, with no other geographical factor taken into consideration.

[14] The 2012 edition has since been published on the Copernicus website at the end of the first half of 2018.

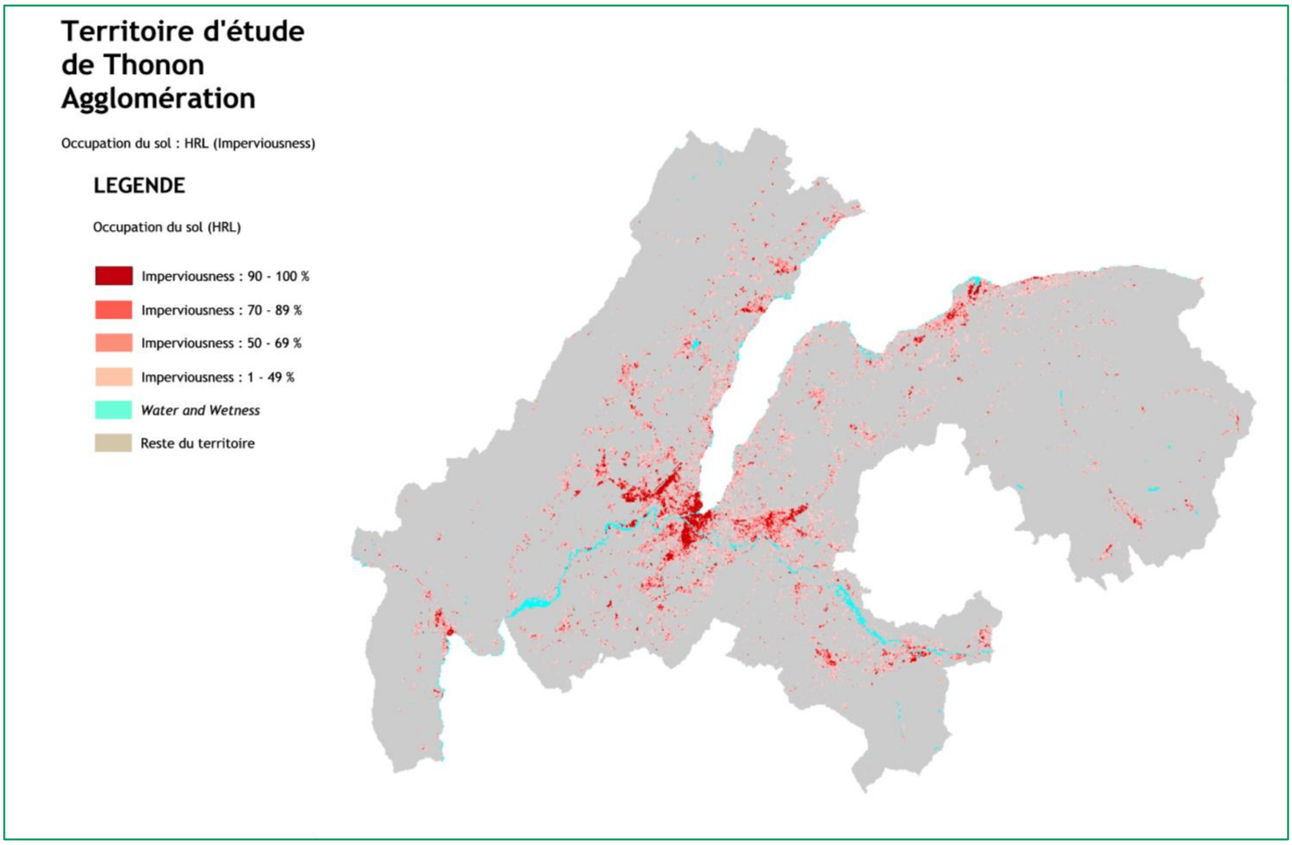


Figure 33: HRL Imperviousness layer (2015) for the Thonon Conurbation study area

4.3.1.3 For Theïa, the choice of two editions

For the Theïa data, which has a nomenclature consisting of 17 items, we did not pre-process the initial raster data.

However, the question arose of the choice of editions.

The technical changes due to the use of a new satellite from the 2016 edition onwards made it impossible to compare previous editions with those released after that date (due to a difference in pixel size). As the only two available editions, released in 2016 and 2017, were very close, it was deemed unnecessary to use both layers.

We therefore decided to use the editions from 2009 to 2014, but it was impossible to download the 2014 edition from the website that distributes the geographical data. We accordingly worked with the 2009 and 2011 editions.

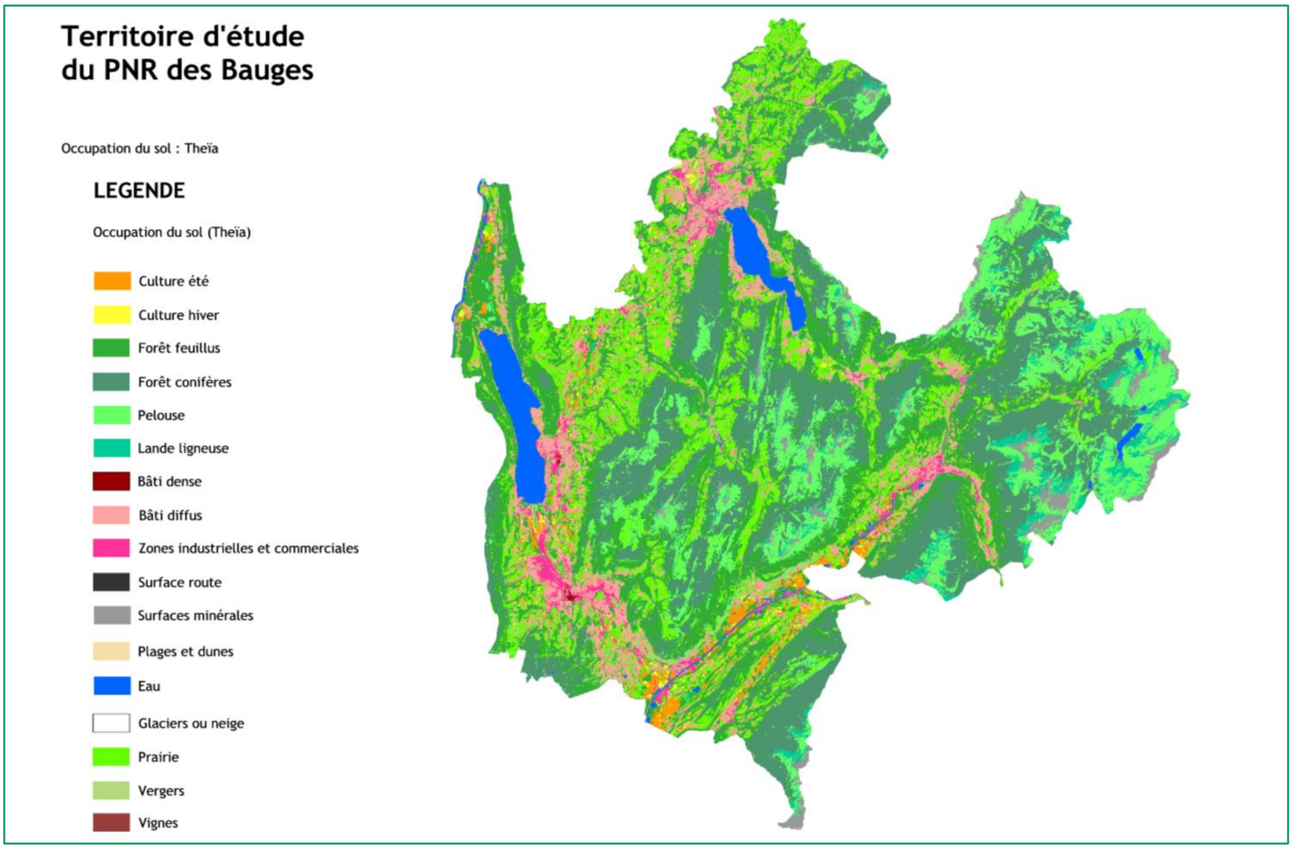


Figure 34: Theïa land use data for the Bauges Regional Nature Park study area

4.3.2 Modelling with LUCSIM

4.3.2.1 Using the CLC database

For each of our study areas, we were able to incorporate the 2000 and 2012 editions of CLC using the simplified nomenclature described above.

The *Continuous urban fabric* land use category is very uncommon in our study areas, so we decided not to include it in the model. Instead, we focused solely on the land use category *Discontinuous urban fabric.*

In accordance with the planned study programme, we carried out three series of modelling processes:

- With neither an attractiveness map nor a potential model;
- With a potential model calculated by LUCSIM, incorporated into both the decision tree and the constraints;
- With the attractiveness map we produced for the strategic scenario, incorporated into both the decision tree and the constraints.

For each model, we also applied a *Markov chain* constraint:

- A *Markov chain* takes the form of a transition matrix, which we can subsequently raise to the desired power. By multiplying this transition matrix with the initial land use (in matrix form), we obtain a theoretical quantity of each final land use. This calculation applies the following principle: if we repeated the past trends X times (X being the value of the transition matrix power), what would be the new quantities of the various types of land use in our study area?
- Focusing solely on the *Discontinuous urban fabric*, we used the power values that would produce a quantity of created pixels (≈ land take) equivalent to the land take calculated for our strategic and baseline scenarios. For the series of modelling processes with neither an attractiveness map nor a potential model, we used the land take data for the baseline scenario. For the other series of modelling processes, we used the data for the strategic scenario.

For each situation, we went through the decision tree three or four times, varying the size of the learning sector in the tree (*Neighborhood radius* parameters).

From the set of transition rules obtained, we retained for the modelling process only those that did not explicitly change the following types of land use: *Water, Wetlands, Continuous urban fabrics, Industrial and commercial zones, Other sealed areas*, because we believed that these rules would produce unwanted effects.

4.3.2.2 Dramatic increase in the computational time for analysing the HRL and Theïa databases

When we were testing LUCSIM with the other databases, we ran into a problem: the time taken to run the decision tree using Theïa and HRL data increased dramatically. When we used CLC data, the run time was about half an hour. For HRL and Theïa alike, we stopped the process after two days without having obtained any results.

This situation is largely due to the level of precision of the initial data (size of the minimum mapping unit and precision of the nomenclature). The minimum mapping unit is much smaller for the latter two databases than for the CLC data, making it more complicated to run the decision tree's algorithm.

Next steps

Given the longer run times, which compromise the use of these databases for modelling, and to reduce computational time, possible solutions could be to:

- Simplify learning within the decision tree:
    - Reduce the share of learning: within the decision tree, the Training ratio, set by default at 75/100, indicates the proportion of pixels with which the decision tree determines the transition rules. Lowering this ratio should make it possible to reduce the computational time, though we cannot be sure of this.
    - Reduce the learning sector: by adjusting the Neighborhood Radius parameters within the decision tree, i.e. the parameter used to set the sector around each pixel studied by the decision tree to determine the transition rules, it is possible to reduce the learning sector and, in so doing, shorten the computational time. For example, reducing the radius from 10 to 8 will approximately halve the quantity of pixels in each sector analysed by the algorithm.
    - Do not use the Weight option: this option can be used to overvalue the changes in order to make the rules more accurate. However, it also makes the decision tree algorithm more complex, which may affect computational time.
- Simplify the input data:
    - Reduce the number of classes.
    - Reduce the precision of the information: HRL is available with 20 metre square pixels and Theïa with 30 metre square pixels. One way of reducing processing time would be to use larger-sized pixels, up to 100 metre square, which would yield a level of precision comparable to that of CLC. However, it should be borne in mind that this has the direct result of reducing information precision.

In view of these considerations, we decided to apply some of these solutions to the HRL data and the Thonon Conurbation study area. With the HRL data, the changes observed between 2006 and 2015 are significantly greater in the Thonon Conurbation study area than in the Bauges Regional Nature Park study area (negligible changes). We therefore decided to confine our last processing operations to the Thonon Conurbation study area.

4.3.2.3 Modelling the Thonon Conurbation study area with HRL data

To simplify the input data, we reduced the level of precision of the HRL data by increasing the pixel size (to 100 metres square). Based on our two layers, 2006 and 2015, we were then able to obtain concrete results in defining rules with the decision tree. The decision tree was applied to all of the different types of land use so that we could generate changes for each of them.

For want of time, we did not run modelling operations constrained by the potential model or an attractiveness map. Given that we intended to change different types of land use concomitantly, it was effectually impossible to apply a constraint such as a Markov chain. Even if the Markov chain indicates changes for each type of land use, these changes will not necessarily be carried out through the application of our transition rules. As we saw from experience, during the modelling process it is quite possible to have a type of land use that increases significantly whereas the other types are stable or even diminish, since this depends directly on the rules applied.

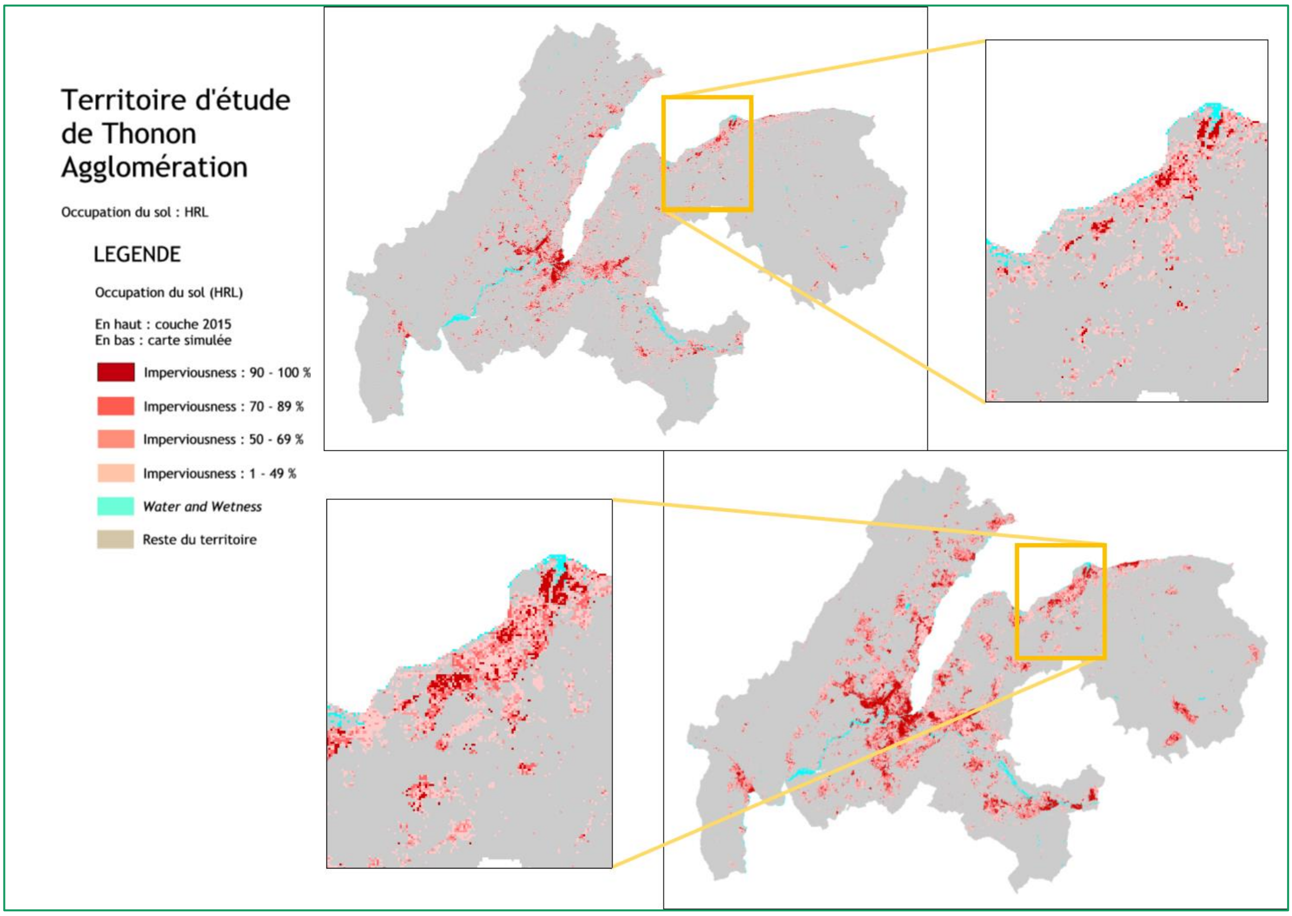


Figure 35: Model of the Thonon Conurbation study area obtained using LUCSIM and the HRL database

4.3.2.4 Analysis of the results and feedback

For want of time, we were unable to develop any real strategy for analysing the results obtained from LUSCIM. Though the theory behind its operation is very simple, LUCSIM proved to be quite complicated to use in practice and the models obtained merit a far more in-depth study than we were able to undertake.

4.3.2.4.1 Operation

Detailed feedback follows concerning the various steps conducted during our testing:

- Decision tree processing times can be long: many parameters will have an impact on the processing time necessary for the decision tree to identify the transition rules. These parameters concern the basic unit for defining the land use layers, the number of land use types and the way in which learning occurs (decision tree internal parameters). The impact was such that we were unable to perform modelling with our most detailed data (HRL or Theïa).
- It is difficult to verify the impact of a *Suitability map* (attractiveness map or potential model) with the decision tree: while there are sometimes conditions that specifically use the *Suitability map* or a change in the number of rules, many rules are strictly identical with or without the *Suitability map*.
- At the end of the decision tree, the transition rules merit a detailed analysis:

    - The number of rules obtained varies from a handful to hundreds, and for each rule, the number of conditions varies from about 10 to close to 40 for the most complex ones. Without a detailed analysis, it is impossible to understand the phenomena they are trying to reproduce.
    - Because the decision tree is a purely statistical process, it may bring up transition rules that have no "real" meaning and can be deemed inapplicable. To avoid unwanted phenomena during our CLC data-based testing of Discontinuous urban fabric land use, we eliminated all transition rules involving changes to water bodies and other developed or sealed areas in the Discontinuous urban fabric.

- During modelling, it is possible to incorporate a constraint such as *Potential/Suitability map*, which has a real impact on models. However, we were unable to analyse its impact in any detail:
    - This type of constraint adds a condition to all transition rules that means that changes only occur above a certain level of attractiveness. The results obtained are different with or without a Suitability map but, without more in-depth analysis, there is no way of describing its impact in detail.

- During modelling, it is possible to incorporate a Markov chain constraint (which constrains the amount of change). However, this does not imply that the software will be capable, given the transition rules assigned to it, to change the area sufficiently to attain the maximum set by the Markov chain.

- Processing times can be long during modelling. The processing time depends not only on the size of the study area, but also on the following three parameters:
    - The number of iterations: we ran continuous modelling, which means that the software applied all of the transition rules to the study area and kept reiterating this step until either it attained the level of the Markov chain constraint or no further changes were possible using the transition rules provided. The number of iterations varied from a few steps to over 100 and had a direct impact on the modelling time.
    - The number of transition rules: the rules are applied at each iteration of the modelling process, so the higher the number of rules, the longer the time taken for modelling. Note that not all of the rules may be applicable at each iteration (or only to a very small number of pixels) if the required geographical conditions are not met. However, the software is still obliged to check each rule at each step.
    - The complexity of the transition rules: this complexity is mainly determined by the number of conditions that have to be checked, but it also depends on the size of the geographical sector in which the condition applies. Note that this sector of analysis, i.e. the size of the sector in which the software will count the different types of land use to check whether they meet the different conditions of the rule, or not, is specific to each condition. It is necessarily lower than or equal to the sector set in the *Neighborhood Radius* parameter (the learning sector) of the decision tree.

4.3.2.4.2 The modelling processes

Over and above the software's operation, we ran a large number of modelling processes. Though, for want of time, no detailed analysis was performed of each one, the following results were observed:

- At each step or iteration of the modelling process, the transition rules are based on geographical conditions that may or may not be met for the study area:
    - A rule that cannot be applied will have no impact at all in the modelling process;
    - A rule that can only be applied on a very small portion of the area will have a very limited and very localised impact in the modelling process;
    - A rule that, on the contrary, can be applied to large sectors of the area (unless it is a Markov chain constraint) will have a marked impact in the modelling process, even if its impact will, in principle, remain localised.
- It would appear that the majority of rules with some degree of complexity have geographical conditions that stop them having anything more than a very limited and localised impact at each iteration. They generally end up not being able to be applied at all.
- Moreover, we have only ever modelled changes in a single type of area: *discontinuous urban fabric*. All other types of land use can only diminish. For certain types of land use that are clearly predominant in our study areas (i.e. forests and agricultural land), this has only a minor impact on the applicability of the rules, in principle. For others (in particular all of the other developed areas), it will rapidly make many of the rules' conditions inapplicable. We tried to limit this effect by removing all rules that resulted in the conversion of these other developed areas. Despite this, because the Discontinuous urban fabric had expanded, these areas had become more distant and many of the conditions (and hence the rules) were no longer applicable.
- It is not always possible to attain the threshold levels laid down with the Markov chain.

Visual analysis of the models produced by LUCSIM in fully-automatic mode reveals certain phenomena:

- Examination of the models produced shows that some sectors never change: not all of the rules are applicable in certain areas, such as natural areas but also the outskirts of urban areas. This outcome raises questions insofar as it is ungrounded: a collateral effect of the way in which we wrote the rules, based on the decision tree.
- Some areas, on the contrary, change, and sometimes the changes observed resemble rather conventional urban densification, more or less linear development or spontaneous urban development:
    - In some areas, we can see urban densification, or more or less linear development, without necessarily having an "oil stain" effect.
    - We can see spontaneous or transferred urban development: new urban development sectors can spring up and grow in the middle of almost exclusively agricultural areas. Sometimes extensive areas will suddenly be developed because a way is found to apply one or more rules on a large scale.

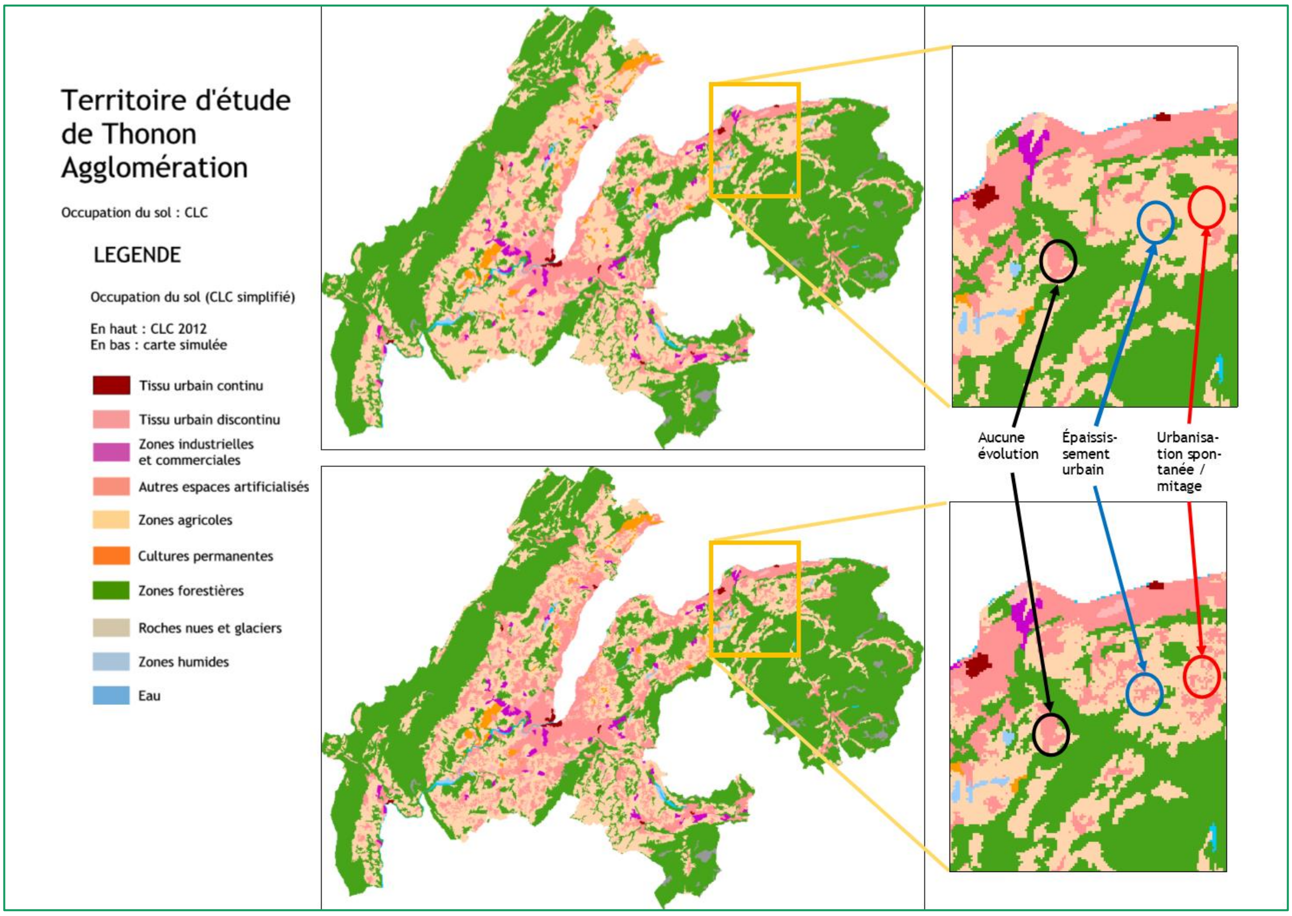


Figure 36: Illustration in the Thonon Conurbation study area of phenomena modelled by LUCSIM (no change, urban densification and spontaneous urban development)

- Then there is another type of phenomenon that is impossible to link to a real-life phenomenon. This phenomenon is the result of the reiterated effect of a combination of rules that, at each step of modelling, create the sufficient conditions for them to be applied again. This tends to appear with time as the number of iterations in the simulation increases. The result is the appearance of development pockets in the shape of "snakes" or "oil stains".

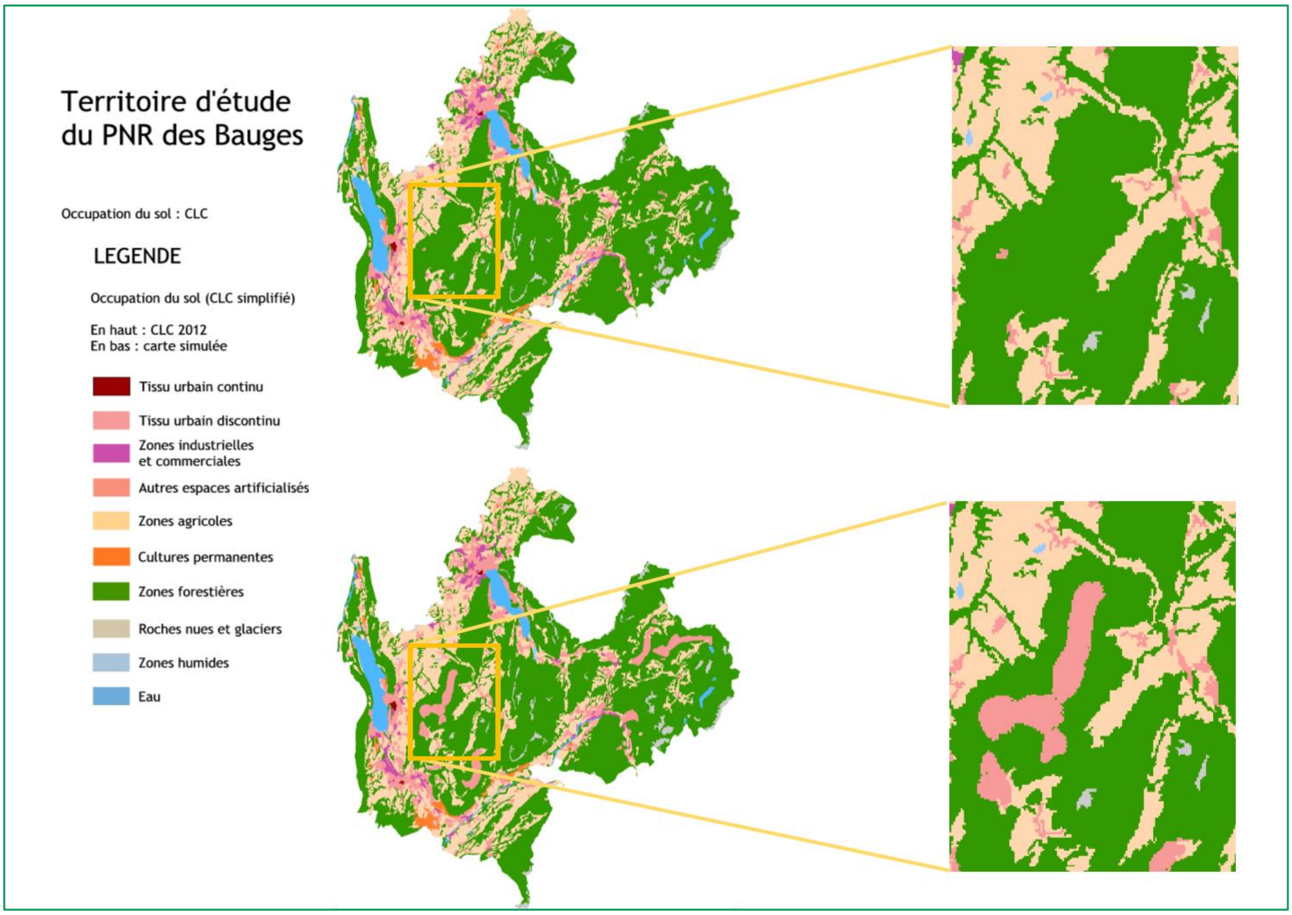


Figure 37: Illustration in the Bauges Regional Nature Park study area of the "snake" phenomenon modelled by LUCSIM

The models obtained in our two study areas are quite dissimilar in this respect.

Though our two study areas are similar in size (≈ 2,600 km²), the developed areas are markedly more extensive in the Thonon Conurbation (36,694 ha) than in the Bauges Regional Nature Park study area (23,037 ha), and the urban sprawl is also more marked there between 2000 and 2012. It is very clear that our two study areas were not processed in exactly the same way by LUCSIM:

- The changes observed between 2000 and 2012 have a direct impact on the definition of the change rules by decision tree within LUCSIM. They are much easier to determine and there are more of them in the Thonon Conurbation study area than in the Bauges Regional Nature Park study area. For all of our tests, the decision tree yielded between 18 and 30 transition rules for the Bauges Regional Nature Park study area and between 89 and 169 rules for the Thonon Conurbation study area (with a strictly identical modelling strategy);
- In the modelling processes, the "developed" quantities increased by between 1,500 ha and 8,400 ha (the maximum threshold used with the Markov chain) in the Bauges Regional Nature Park study area, and by between 7,800 ha and 14,800 ha (the maximum threshold used with the Markov chain) in the Thonon Conurbation study area.
- With regard to the number of iterations before modelling stopped, certain modelling processes required up to around 100 steps in the Bauges Regional Nature Park study area, but never exceeded 30 stages in the Thonon Conurbation study area. There is no obvious connection between the quantities developed and the number of iterations. On the other hand, the higher the number of iterations, the more likely we are to see "impossible" phenomena such as "snakes".
- The thresholds set by a Markov chain, which differ for baseline and strategic models, were attained for 4 out of 9 models in the Bauges Regional Nature Park study area, and for 7 out of 14 models in the Thonon Conurbation study area.

We were able to run a final series of modelling processes for the Thonon Conurbation study area using the HRL *Imperviousness* data. We artificially decreased the precision of the information in order to have a pixel size of one hectare (the same as for CLC). For these modelling processes, we applied the decision tree for each type of land use in an endeavour to have rules that, theoretically, allowed us to change all of the types of land use concomitantly.

Since the results were obtained at the very end of the assignment, we were unable to analyse the results yielded by this data.

5 Results and feedback for the modelling programs

5.1 Modelling with FORESIGHT

To recap, we had a number of objectives for our testing programme:

- Model development in the area using separate databases;
- Model different scenarios, differentiating the scenarios on the basis of the attractiveness map, land take and the urban sprawl patterns.

Learning to use the software

The FORESIGHT software, in itself, is easy to learn to use and has a detailed user guide that rapidly equips users to begin modelling the area's development.

The software's operating principle

The modelling principle is clear and quite easy to grasp because the phenomena described can be directly linked to actual observations on the ground (urban scattering and linear development, for example, or the land take, etc.).

Preparing data and incorporating it into the software

The pre-processing module makes data preparation simple. Preparing an attractiveness map with the dedicated module is just as straightforward.
We had no difficulty incorporating our two databases, which differed in the precision of the information. However, it should be noted that the modelling processes used a minimum mapping unit of one hectare, which led to a "simplification" of the most precise data. In the cross-border Thonon Conurbation study area, we did not have a uniform layer for the entire land use area that was more precise than CLC. Moreover, given that there are no updates planned for the software, all future updating of the databases could prove problematic[15].
The attractiveness map makes it possible to include a great deal of detail in the modelling process (tailoring attractiveness to the area's level of protection, the desired direction and focus of urban development, etc.). However, it was not possible for us to test the incorporation of our own attractiveness maps[16]. We managed to find a workaround solution using the software's module for creating the attractiveness map. We were able to define land take and urban sprawl patterns, which are modelling input parameters, for each of our scenarios.

[15] The data pre-processing module could become inoperative if the databases were changed in future editions. To avoid this difficulty, it is theoretically possible to incorporate our data directly into the modelling module (without going through the pre-processing module), but we did not test this option.
[16] After further testing, we were able to incorporate one of our attractiveness maps into FORESIGHT. There is no way to do this through the *Pre-processing* tab. However, it was possible to generate the *.gif file for one of our attractiveness maps directly through ArcGIS. Users will need to ensure that the resolution and coverage are strictly identical to those in the files generated by FORESIGHT in the *Input* folder for the *Initial Urban Map* data, for example. The file was then copied into a model's *Input* folder, where it was recognised by the software *(Preview* in the *Initial settings* tab*)*, then used to generate a model based on our own attractiveness map.

Analysis of the results obtained

The software generates urban sprawl maps, which might be considered an overly simple output (given that there is absolutely no characterisation of the densification phenomena), but it has the advantage of being easy to analyse. Given the "automatic" nature of the software and a substantial degree of randomness in its workings, there is no point trying to use this tool for other than its intended conditions of use. More specifically, there is no point trying to:

- Analyse the results in too great a detail (for example on the scale of a commune or part of a commune);
- Model urban sprawl with overly small pixels;
- Model scenarios that differ in overly subtle or microscopic ways.

The tool should also be seen as part of a long-term forecasting approach and used to represent different scenarios for the area's future development, since the software's sole utility is to compare various possible developments of the area.

However, the software's ability to rapidly generate a number of models and combine them into a probabilistic outlook (which is more useful than the random phenomena introduced by the software) is a real asset.

The visual analysis of the results, in particular, highlights the impact of land take and the patterns of urban sprawl. It was also possible to identify the impact of the distinctions added to our attractiveness maps, even if, on the other hand, it was more complicated to measure the latter in practice.

In the Bauges Regional Nature Park study area, for which we were able to test two data sources (CLC and OSCOM) for defining the initial urban map, the results are not identical. The data source used has an impact on the results. In spite of recalibrating the OSCOM information to a one-hectare minimum mapping unit, its level of precision affected the modelling process. Whatever scenario we use, with OSCOM, urban development is more dispersed over the whole area, and the quantities of pixels created vary with the minimum mapping units studied.

Conclusions

In the light of the testing carried out in the ASTUS project, we can state that the FORESIGHT software can be a useful tool for modelling urban sprawl in an area. This easy-to-use software has useful features (urban sprawl patterns, link between transport infrastructures and urban development, attractiveness map) that make it possible to adapt the concrete specifics of the modelling processes to the areas. Apart from the software's numerical parameters, which have to be calibrated beforehand, the attractiveness map theoretically offers real possibilities for factoring in some or all of the area's geographical information. This makes it quite simple for users to link it to the area's planning documents. Even so, within the framework of the tests we conducted, we did not manage to incorporate our own attractiveness maps into the software.

In any case, given that the tool was capable of producing an image of what our study area might look like in the future, it delivers information that may be of direct interest to an area's decision makers: it shows the impact that urban sprawl might have, the impact of certain policy directions taken in urban planning documents and the combined, large-scale effects of all of these urban planning documents.

With certain precautions, such as generating multiple models and combining the results, it is also possible to carry out a fine-scale analysis of the maps. This approach would provide a means of studying the impact of new urban development on aspects such as networks, mobility, infrastructure and facilities, and/or urban development patterns.

5.2 Modelling with LUCSIM

To recap, we had a number of objectives for our testing programme:

- Model development in the area using separate databases;
- Model different scenarios, making each one different by using an attractiveness map (and the potential model included in LUCSIM) and a constraint such as a Markov chain.

Learning to use the software

From a functional point of view, LUCSIM is easy to start using in "fully automatic" mode, as we saw from our own experience in the ASTUS project, even if the user guide is at best succinct.

The software's operating principle

The rules generated by LUCSIM are exclusively geographical. The development of a given cell will depend on whether, in a specified neighbourhood, there are certain quantities of the various types of land within an initial land use layer. The effects of this type of rule are difficult to represent or to relate to traditional phenomena such as linear development or urban scattering, for example. LUCSIM's "fully automatic" operating mode is not explicit about this aspect.

Preparing data and incorporating it into the software

We had very few problems modelling with CLC in the *Discontinuous urban fabric* mode. However, because of the processing times per decision tree, no results were obtained with the HRL and Theïa databases. For the HRL database, after simplifying the data by changing the minimum mapping unit from 20 metres square to 100 metres square, we were able to model the Thonon Conurbation study area.

To be able to use the decision tree to automatically define transition rules, and to use a Markov chain constraint to cap the quantity of land use change allowed during modelling, it is necessary to have two land use layers at two separate dates. Moreover, because the decision tree works on changes that occurred between these two dates, the more changes there are, the more the results will be precise and useful.

Using the decision tree makes it possible to quite quickly obtain the first rules for changes in the area, in order to be able to execute our first models. Note that the precision of the initial information has a major impact on the processing time necessary for the decision tree to yield its results[17]. Note also that the number of rules generated by the decision tree has an impact on the subsequent model.

[17] Going from a matter of minutes when using CLC data to several days for the more detailed databases we studied, such as HRL or Theïa, with no guarantee of a result.

Analysis of the results obtained

After simply studying the transition rules obtained through the decision tree and through visual analysis of the models obtained, we can put forward a few remarks about LUCSIM's operation.

While the Markov chain can be used to cap growth in certain types of land use, it is not always possible to attain these thresholds and often the modelling process stops before this because the transition rules can no longer be applied in the area.

The attractiveness map, *Suitability map* or potential model used by LUCSIM have an impact on the definition of rules by the decision tree and on the models themselves.

The changes do not include any notion of time, so are not regular in time or space. The first iterations of the model generally generate extensive change, while the last ones generate only very minor change.

Lastly, the changes modelled for the area are far from uniform:

- Some sectors display no change at all, while other show significant change;
- Changes to the Discontinuous urban fabric sometimes resemble traditional urbanisation phenomena, such as urban densification, linear development or urban scattering;
- Sometimes fictional phenomena known as "oil stains" or "snakes" appear.

Conclusions

Given these results, we are not in a position to draw a conclusion about LUCSIM's suitability for use as a decision aid.

In some respects, though, LUCSIM delivers very real benefits:

- LUCSIM can change different types of land use concomitantly, so would be capable of modelling certain densification phenomena;
- LUCSIM can also identify and model the impact of certain interactions between the urban environment and the natural environment, and, if the initial land use layer is sufficiently detailed, between the different types of land use in an urban environment.

However, we were not in a position to test these possibilities in practice.