

CELLULAR AUTOMATA FOR THE ANALYSIS OF SPATIAL EVOLUTION IN ALBANIA:

A decision tree approach

Giancarlo CARRAI⁽¹⁾, Louisa J.M. JANSEN⁽²⁾, Serena PECORI⁽³⁾ and Massimiliano PETRI⁽³⁾

⁽¹⁾ Managing director, Svaltec s.r.l., Florence, Italy

⁽²⁾ Land/natural resources consultant, Rome, Italy

⁽³⁾ L.I.S.T.A. – Laboratorio di Ingegneria dei Sistemi Territoriali e Ambientali
University of Pisa, Department of Civil Engineering

The aim of this ongoing research is to investigate the land use transformations in a large territorial area including the common of Preza, placed in Albania. It is an attempt to analyze the connections existing between the deep political and socio-economical changes that took place in the last decade in Albania and land use transformations.

The tools adopted derive from the field of the KDS (Knowledge Discovery Systems) and their aim is the identification of the relations and the "rules" that structure the evolution of land use dynamics that took place in the period between 1991- 2003. Among the many tools referring to the field of KDS, we used in particular classification methods based on decision trees induction, that have been applied to multitemporal spatial data in order to produce IF/THEN type rules. The extracted rules will be implemented, in the next phase of this research, inside a Cellular Automata.

KEYWORDS

G.I.S., Decision Trees, Cellular Automata, Rules, Land Use

INTRODUCTION

With the recent technological developments in the field of remote sensing and survey methods (as GPS, laser scanner and so on) it has become very important to verify and test new methodologies for the construction and extraction of knowledge from data. In the last few years researches were attempting to answer this demand by applying techniques belonging to the field of Artificial Intelligence.

In context such territorial planning process, the quality of the evaluation procedure increases with the knowledge of the dynamic behaviour of the spatial system for which decisions are requested. Factors and relations, which influence such behaviour, have been theorized, formalized and modeled since the birth of urban and territorial studies in many disciplines. In the last few years, such theories and models (including the "dynamic revolution" of the eighties) were based on the explicit formulation of the rules, the relations and the forces which trigger the actors choices and then determine both city's and territory's evolution. But, if we suppose that nowadays some of such rules and relations could have changed, there is the need to develop a strongly aimed theoretic and operational field of study: knowledge building and management in territorial planning process. In this case it is proposed an area where the deep political socio- economic change occurred, make necessary to verify the sustainability of future land use development.

For this purpose it is, therefore, necessary to select and implement innovative tools able to handle the huge amount of available data concerning territorial systems (statistical, economic, cartographic data and so on) in order to extract useful information from them.

In this context, the building of cognitive systems (intelligent systems) is proposed, which, observing a given reality, be able to extract knowledge on the role played by urban/territorial factors and by their relations in urban/territorial evolution and change.

On the basis of a survey on the state of art about the presently available tools, some of them, derived from Artificial Intelligence and from the field of G.I.S, are selected. Though the elaboration of various kind of available data (thematic maps, population density, transport infrastructures, productive settlements, etc.) it is possible to extract and build knowledge directly from experimental data and it is also possible to represent the extracted knowledge in a very efficacious and communicable way, in the form of sets of spatial transformation rules. Afterwards, these rules can be analyzed and compared with traditional rules we adopted, until now , to explicate urban/ territorial dynamics, in order to change/ correct some theoretical issues.

This methodology of extracting knowledge directly from data is very suitable in territorial contexts such as Albania where new and unknown territorial dynamics are emerging, so that it is impossible to use formalized theories, because we really don't know if they can be considered valid: the only way is to 'observe' what is happening and try to formalize extracted information in a structured and understandable form, that is to build knowledge about the dynamics of this particular context. For this purpose is necessary to dispose of large amounts of data: however, the value of such data mainly depends on our ability to exploit knowledge from them. In the first chapter tools and methodologies adopted for this work are described in detail. In second chapter the application to the case study of Albania is presented, for what concern input data and definition of variables. In third chapter data analysis and the final results are described.

TOOLS AND METHODOLOGY

The study of the described phenomena has been performed by using and integrating each other two kind of tools: G.I.S. and automatic learning tools.

During the first phase, by mean of a G.I.S. software, has been built the geographical georeferenced land uses database related to two different time sections. By mean of spatial elaborations performed within the G.I.S. software, the information coming from the various maps are put in the form of relational tables, where in each record are reported the elementary cells through which the territory has been represented and in the columns the attributes that give information about the characterization of each single cell.

During the second phase the table is given as input data of the automatic learning tool, able to perform the extraction of a set of land use transformation rules. During the third phase, still ongoing, the extracted rules are implemented into a Cellular Automaton in order to perfect the calibration of the model.

Geographical Information Systems - G.I.S.

G.I.S. tools have been used during the first phase, that was related to data consolidation and preprocessing; this is because G.I.S. systems make possible both to store a remarkable amount of digital cartographic data by georeferencing them with respect to a same reference system, both to perform on them any kind of spatial analysis.

First of all the study area was modelled as a vectorial grid containing over 10.000 square cells (50X 50 m), able to memorize many attributes for each cell used to subdivide the whole territory of Preza municipality. Beginning from three land use maps and some vectorial spatial data, thematic maps have been built using Overlay Mapping and Geoprocessing tools.

Infact, the information coming from land use maps has been enhanced and homogenized with other information related to the presence of anthropic infrastructures, with social and demographic data and morphologic variables.

These information has been assigned to each cell. At this point it is therefore possible to select the local elements that are in the neighbourhood of every elementary cell.

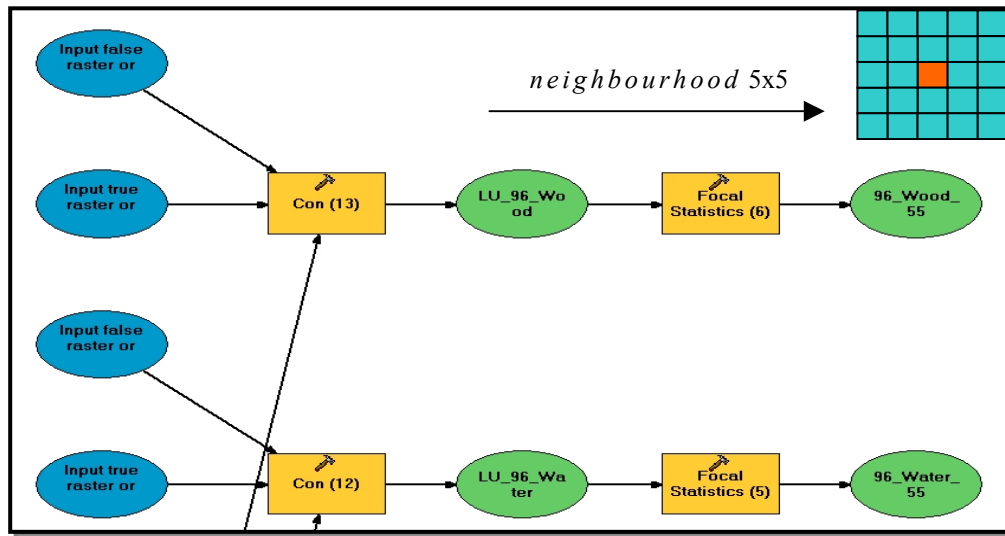


Figure 1 Section of model used for neighbourhood analysis

The characterization of “neighbourhood ” has been performed by mean of Model Builder within ArcGis 9 Software, able to create a very efficient data analysis computational model, even in presence of large amount of input data: in Figure 1 you can see the model that has been built for neighbourhood analysis related to single land uses, that make possible to automatically extract, for each cell, the number of cells of each land uses that are located within its neighbourhood, consisting in 5X5 (25) cells,namely a rectangle with each side measuring 250 meters (see Figure 2).

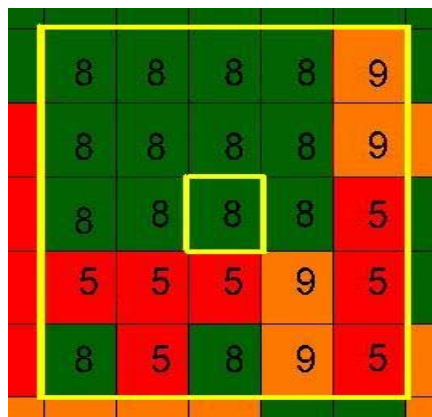


Figure 2 Characterization of neighbourhood of each cell

In order to better explain what Figure 2 represents, considering the “core cell” (characterized by land use number 8), it presents a neighbourhood consisting in 14 cells of Land use number 8, 7 cells of Land use number 5 and 4 cells of land use number 9.

The results of these elaborations are put, by using some automated procedures, in the form of relational tables, in which, for each spatial entity there is a description (defined by mean of alphanumeric attributes) of the state of the cell and a description of the states of the neighbourhood entities, previously characterized. Such table represents the input data for the automatic learning tools that make possible the explanation of the relationships among the elements of the studied area by generating, therefore, a set of transformation rules.

Knowledge Discovery Systems

In the last few years the problems related to handle increasing amount of available data (in particular, territorial data) has produced a growing interest with regard to tools able to automate learning processes. In this context, a new technology emerged, called Knowledge Discovery in Databases" (or KDD in short) consisting in a number of systems and tools oriented to discover useful and interesting knowledge from data in large databases. For this work an algorithm based on a simple classification method is used, able to perform learning from example, namely classification based on decision trees induction.

Decision trees belong to supervised classification techniques: initial data consist in a certain number of records each of them presenting multiple attributes (i.e: land use at the first time , activity density, slopes) and a class label, that is the classification target attribute (in our case land use at the second time).

In classification, input data (called training set) are analyzed in order to produce an accurate description or model for each class: this model can be later applied to classify further and/or future cases whose classes are unknown. In our case, records are related to each single cell within " vectorial Grid" and contain a series of attributes. In a specific way the used tool is the Open-Source Software Weka 3-4, from University of Waikato, in New Zealand [18]. This algorithm, named "C4.5 – version 8", works splitting attributes beginning from the root.

In the first step input data are partitioned in groups (i.e. two groups) on the basis of the values that a particularly explicative variable can assume; afterwards each of these group is partitioned again on the basis of another variable (less explicative than the first one); the partition goes on until one of the following conditions holds:

- all samples for a given node belong to the same class;
- the number of samples in the node is less than a threshold, which is given to the algorithm as input parameter;
- there are no remaining attributes on which the sample may be further partitioned.

In the last two cases a majority voting is employed: the node becomes a leaf and it is labeled with the class which is in majority in the samples of the node. Alternatively, the class distribution of the node sample may be stored.

The criterion used to chose the split attribute is the main characteristics which differences a decision tree induction algorithm from the other. The information gain measure used by C4.5 is an information-theoretic measure which selects the attribute which minimize the information needed to classify the samples in the resulting partitions (maximize entropy reduction), and reflects the least randomness or impurity in these partitions [5]. The classification results can be visualised by using a tree structure like a flow chart in which we can find a root, edges, nodes and leaves (see Figure 3).



Figure 3 An example of Decision Tree

The root represents the attribute on which the partition has been performed, the nodes are labelled with names of the remaining attributes, the edges with the possible values that attributes can assume and the leaves are labelled with the different classes. Such a framework is easy to be interpreted, as in it each path going from the root to one leaf, through the edges corresponding to the attribute values, represents a classification rule. It is therefore possible to classify cases for which label class is unknown, using decision trees like a provisional technique. Each rule presents a value of “support” and “confidence”, able to measure frequency and strength of the rule. Rules that are valid for many cells present a high value of support, while very precise rules present a high value of confidence.

Various methods can be applied for rules’ validation: in the so called Holdout method [3] the dataset is split in two parts, namely a 66 percent for the training and a 33 percent for the validation of the model. In this application, in order to avoid the problem of working with a not so much representative training set, the k-fold Cross-Validation method was applied. It consists in the partition of the whole dataset in k parts (folds) and in the construction of a number of k trees; in each of them a single fold is used for the validation and the remain folds for the model training (figure 4). From the k steps the algorithm computes the mean of the classification errors. As argued in previous research, a k-value equal to ten can be accepted.

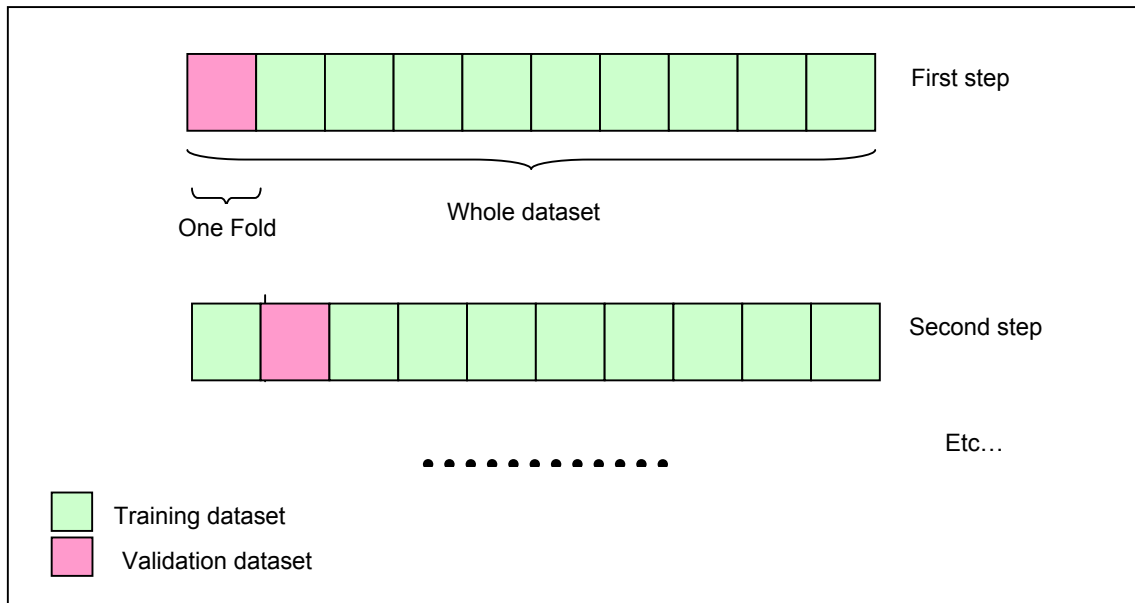


Figure 4 The ten-folds Cross Validation method

APPLICATION

The case study of the present application is related to a territory of Municipality of Preza, in Albania. The main aim is to investigate the land use changes that took place in the periods between 1991- 1996 and 1996- 2003.

The general assumption of our experimental work, derived from some applications developed within L.I.S.T.A. both in urban and in territorial areas[4,5] is that the phenomena lying at the basis of territorial changes are:

- the land use and the presence of various kind of infrastructures or morphologic characteristics in each elementary territorial cell;
- the spatial relations of each cell with other cells, measured in term of ‘enlarged adjacency’.

In other words, for each application has been defined a set of land use classes, obtained by aggregating and homogenizing the land use values corresponding to each map used for that application. The studied area has therefore been modeled in elementary square cells and land use values referred to two different time slices has been assigned to each cell, in order to study the evolution of the cell itself.

For each spatial entity there is a description (defined by means of alphanumeric attributes) of the state of the entity itself and a description of the states of the neighbourhood entities: all the results have been put in the form of relational tables. Moreover, before running the decision tree, some preliminary statistical analysis have been performed, using OLAP (On-Line Analytical Process) Cube, in order to have some previous information useful to better set parameters during the mining phase.

Input Data

The input data have been provided by the EU Phare Land Use Policy II project funded by the European Union Phare Program in Albania [1,2]. The LUP II project provided GIS-oriented instruments and innovative methodologies to support decision-making for land-use policy and planning to the Ministry of Agriculture and Food in Albania. (Giancarlo Carrai worked as GIS/IT Manager and Project Coordinator and Louisa J.M. Jansen as Land Use/Land Evaluation Expert in the LUP II project).

Land Use Classes

Actually, for what concern land-uses definition, an international agreement does not exist. For this reason there are many different opinions about what the word land use constitutes (UNEP/FAO, 1994). This is because it has very different meanings within the different disciplines and, as a result, it implies a set of mostly unidentified characteristics. In the context of the LUP II Project land-use is defined as *'the type of human activity taking place at or near the surface'* [2].

The LUP II project developed the Land Use Information System of Albania (LUISA) [1,2] to harmonise the different land-use legends existing in Albania and these were subsequently included in the developed geodatabase that uses Case Tools (Computer Aided Software Environment) and Visio software in order to structure and organise the data.

LUISA applies certain rules and criteria in a systematic manner. The main distinctive criteria adopted for the description and classification of the various land-uses are therefore: (1) function and subsequently (2) activity. Function refers to the economic purpose of the land-use and can group many different land-use types in a single category; activity refers to a process resulting in a similar type of products and is used at the lower levels of the hierarchy. Activity is defined as "the combinations of actions that result in a certain type of product" (UN, 1989) and refers to a process. On the basis of these criteria the resulting land-use classes have been systematically structured in a hierarchical order, able to regroup them at different levels [2].

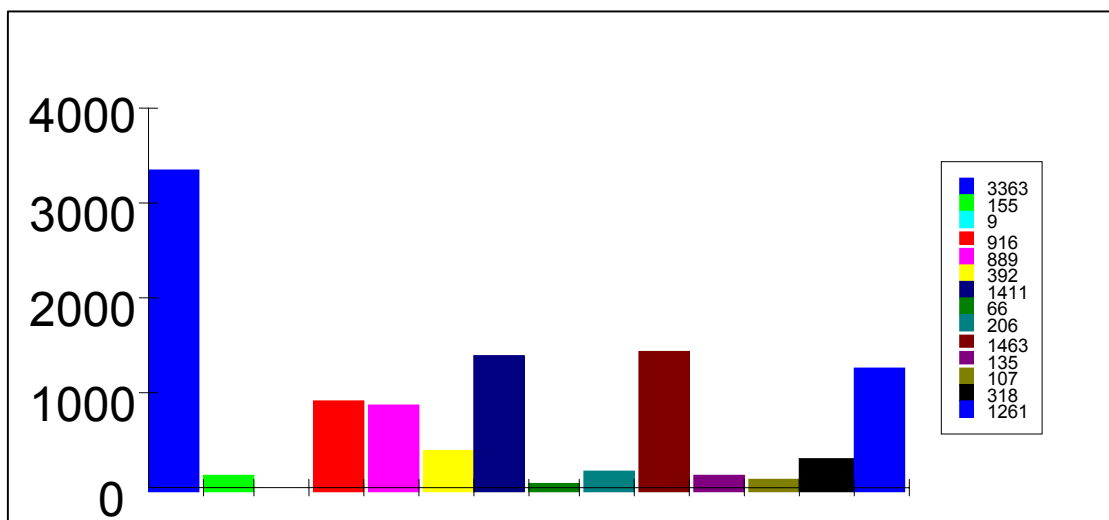
The results of the first classification produced 14 land use classes, presented in Table 1.

Crop	Service
Fruit Tree	Transport
Industrial area	Uncultivated
Meadow/pasture	Unknown
Olive-grove	Vineyard
Recreation	Water
Residence	Wood

Table 1: Land use classes

We perform experiments using these land use classes; the results are presented in a previous work [11].

These classes are representative of different land uses in Albania in the period comprises between 1991 and 2003 and the rules we extracted by using this disaggregation are rather significant; however we realized that these classes are very unhomogeneous. with respect to the criterion of frequency, (in Graph 1 you can see the frequency distribution of this first classification). For this reason the tool was able to extract almost only rules giving information about the most populated classes (crop, wood, etc).



Graph 1: Frequency distribution related to classes in Table 1 at the 2003

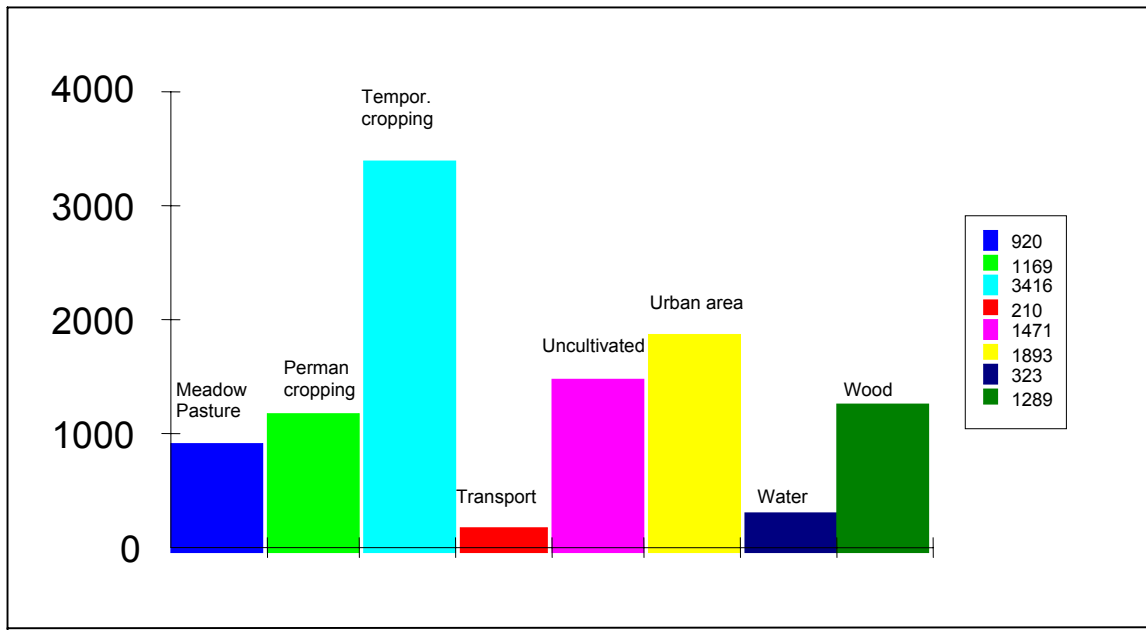
Therefore we tried to perfect land use classes splitting in order to obtain a more homogeneous a morphological criterion and we obtain a more homogeneous classification, presented in Table 2.

AGGREGATED CLASS	INCLUDED CLASSES
1) <i>Temporary Cropping</i>	<i>Vineyard Olive-grove Fruit Tree</i>
2) <i>Permanent Cropping</i>	<i>Crop</i>
3) <i>Meadow/pasture</i>	<i>Meadow/pasture</i>
4) <i>Transport</i>	<i>Transport</i>
5) <i>Uncultivated</i>	<i>Uncultivated</i>
6) <i>Water</i>	<i>Water</i>
7) <i>Wood</i>	<i>Wood</i>
8) <i>Urban area</i>	<i>Service Recreation Residence Industrial area</i>
9) <i>Unknown</i>	<i>Unknown</i>

Table 2: Land use aggregated classes

The frequency distribution of each class is shown in Graph 2.

The 'unknown' cells will be therefore analyzed in order to assign each of them to a known aggregated class.



Graph 2: Frequency distribution related to classes in Table 2 at the 2003

Definition of Significant Attributes Able to Describe Each Cell

The second phase was related to find some good variables able to describe land use changes occurred in that area during the considered time periods. About this, Land-use is determined by natural, economic, institutional, cultural and legal factors. In general, possible land-uses are limited by biophysical constraints that include climate, topography, soils and the geological substrate, presence or availability of water and the type of vegetation, and the type of vegetation. The location of an area with respect to other types of land-uses, such as nearby residential and industrial areas, is also an important factor (e.g., the location of Preza commune close to main urban centres like Tirana and Durrës as well as its proximity to Rinas airport) and economic incentives as part of policy has also a great influence on the land-use patterns. In this context we select a series of variable able, in our opinion, to affect the land use change occurred in that period, considering both functional that physical aspects.

The choice of variables also depend from initial data availability and from the observation of some land use dynamics. In fact, preliminary analysis showed that the dimension of parcels strongly influences their transformations, so that was necessary to include as initial variable the original area of the parcel in which examined cell is included. For the same reason the very different behaviour between edge and internal regions of the same parcel, suggested to introduce the distance of examined cell from the border of the parcel.

Attribute	Description	Domain values
<i>SLOPE</i>	<i>Slope class</i>	(1-4)
<i>WATER_NAT</i>	<i>Distance from natural water sources</i>	(N; B; MB; M; MA; A) ¹
<i>IF_EDIFIC</i>	<i>Presence of building</i>	(0,1)
<i>NUM_EDIF</i>	<i>Density of building</i>	(M, N, P, T) ²
<i>AREA_PART</i>	<i>Area of original (1996) cadastral parcel</i>	(N; B; MB; M; MA; A) ¹
<i>EDIF_R500</i>	<i>Density of buildings within 500 meters</i>	(N; B; MB; M; MA; A) ¹
<i>ROAD_DIST</i>	<i>Distance of nearest road</i>	(1-7)
<i>ROAD_SURF</i>	<i>Surface type of nearest road</i>	(0-3) ⁴
<i>ROAD_COND</i>	<i>Maintenance condition of nearest road</i>	(0-3) ⁵
<i>ROAD_TYPE</i>	<i>Classification of nearest road</i>	(2-5) ⁶
<i>WATER_ART</i>	<i>Distance from artificial watering canal</i>	(N; B; MB; M; MA; A) ¹
<i>EROS_RISK</i>	<i>Erosion risk</i>	(1-4)
<i>DRAIN_VAL</i>	<i>Value of land drainage class</i>	(1-4)
<i>LANDUSE_96</i> <i>LANDUSE_03</i>	<i>Prevalent aggregated land use class</i>	(TC, PC, Wa, Wo, Uc, T, Ua, M) ³
<i>DISTBOUN96</i>	<i>Distance from the edge of cells in 1996</i>	(1-7)
<i><LandUse>_55</i>	<i>Number of cells of each land use located within a neighbourhood with sides of 250 meters, centred on examined cell</i>	(N; B; MB; M; MA; A) ¹
LEGEND:		
1: N= nothing; B= little; MB= medium-little; M=medium; MA= medium-high; A=high		
2: M= medium; N= nothing; P= little; T= high		
3: TC= Temporary Cropping, PC= Permanent Cropping, Wa= Water, Wo= Wood, Uc= Uncultivated , T= Transport, Ua= Urban area, M= Meadow/pasture		
4: 0=N.A.; 1=Paved; 2=Unpaved, 3=gravel loose		
5: 0=N.A.; 1=Good; 2=Fair, 3=Poor		
6: 2=Main road, 3=Secondary road, 4=Urban main road, 5=Urban secondary road		

Table 3: Tables of attributes for each cell

The Problem of 'Unknown' Land Uses

Another problem was due to the fact that in some periods (1996) there are some relevant lack of knowledge in land use maps, due to the fact that there are many cells for which land use value is 'unknown' (see Figure 5).

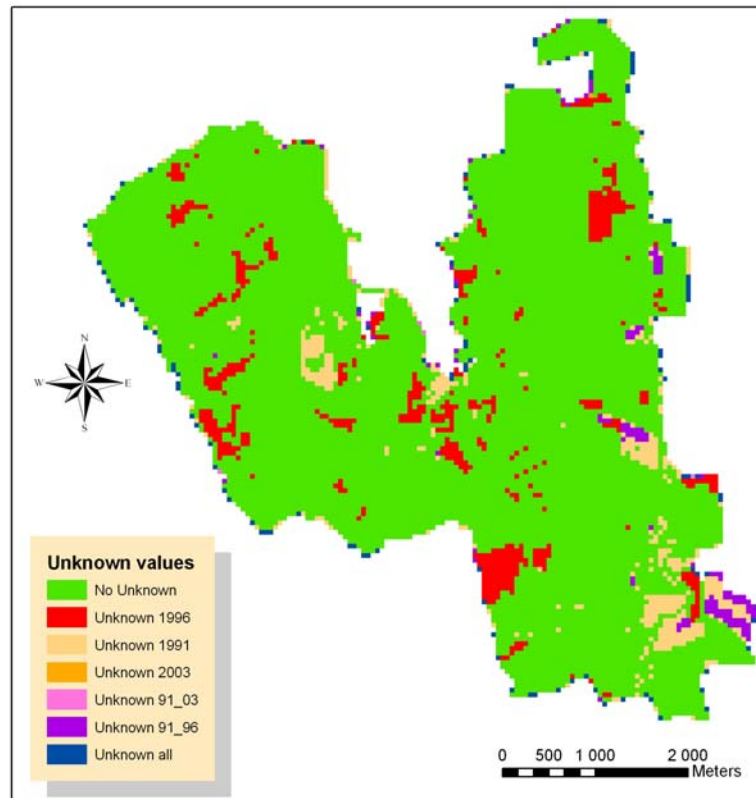


Figure 5 Distribution of the Unknown cells

To resolve these problems, the missing land uses values have been obtained by mean of classification rules. In other words, by observing the whole area it is then possible to extract some regularities between the land use value assumed by each cell and the values of the attributes used to describe the cell itself at the same data (while for what concern transformation rules presented in the following paragraph, the extraction has been performed relating, for each cell, the values of attributes used to describe the state of the cell itself at a certain data with the land use value of the same cell at another data).

Some of these rules at 2003 are presented below:

SLOPE = 1 AND
 IF_EDIFIC = 1 AND
 EROS_RISK = 1 AND
 Transp_55 = 3: Urban_area
 Each cell labeled 'Unknown' presenting the above characteristics becomes 'Urban Area'

SLOPE = 1 AND
 AREA_PART = B AND
 IF_EDIFIC = 1 AND
 WATER_ART = N AND
 ROAD_TYPE = 2 AND
 EDIF_R500 = MB: Transport

Each cell labeled 'Unknown' presenting the above characteristics becomes 'Trasport'

SLOPE = 1 AND
 AREA_PART = B AND
 IF_EDIFIC = 0 AND

ROAD_COND = 1 AND
Transp_55 = 1:

Temp_cropping

Each cell labeled 'Unknown' presenting the above characteristics becomes 'Temporary cropping'

In Figure 6 you can see one of the homogenized land use maps that have been used for building some attributes of the relational tables, as previous mentioned.

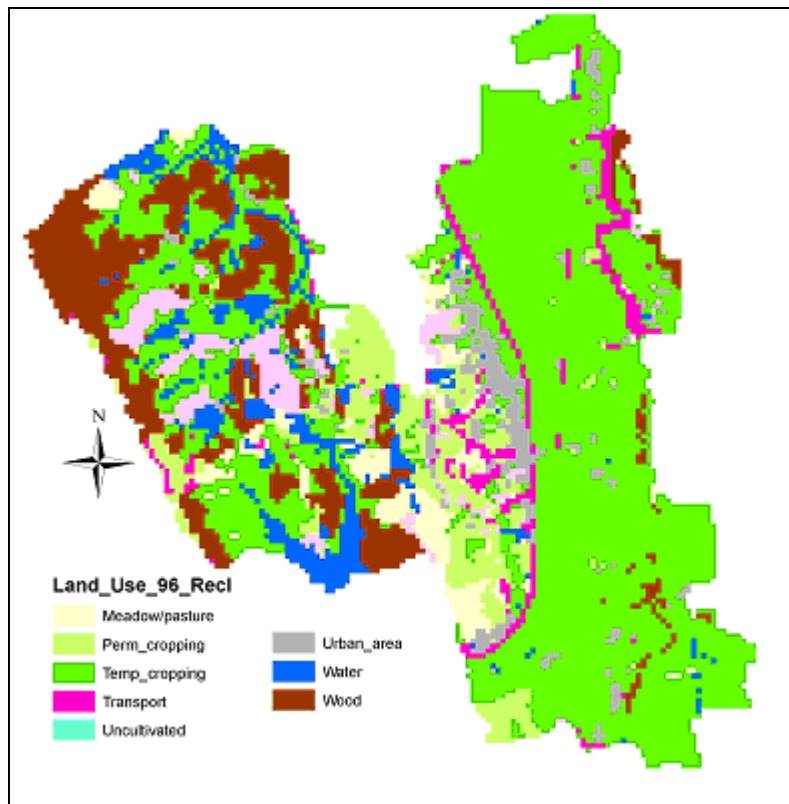


Figure 6 Land use map without unknown cell values related to 1996

DATA ANALYSIS AND RESULTS

Before running the decision tree, the dataset has been analyzed using multitemporal analysis (see following paragraph): beginning from these results we realize that decision trees is not able, in the specific case of Preza, to analyze at the same time tree different land use maps, because the two couple of maps (1991-96 and 1996-2003) are representative of two periods presenting very different dynamics.

Infact, while the first idea was to use the extracted rules in relation to 1991-1996 period as training set for the algorithm and then validate these rules with the maps related to the second period (1996-2003), results of multitemporal analysis showed that phenomenon of land fragmentation that occurred in 1992 (but effective only from 1996) makes really very different the land use dynamic of the two periods, so that we retained the rules related to 1991-1996 not able to explain the transformations occurred in the second time slot.

An historical analysis related to landed property in Albania in this period, allows to better explain the reasons of these difference.

In 1992, consequently to deep political changes occurred in Albania, there was a total reallocation of lands that caused a fragmentation into small parts of original cadastral parcels consequent to transformation of landed property from public to private: the State decided that a portion of 0,7 hectares had to be equitably distributed to each person, as the documents certifying the old land property were not found. However, until 1996, free transaction of lands was forbidden: it is therefore evident that until 1996 land use dynamic has been paralyzed, while after this date we assist to a deep transformation of general structure of territory.

Beginning from the results of multitemporal analysis is, therefore, possible to compare the two period. Moreover, a more detailed study concerning the comparison between dynamic territorial rules related to period 1991- 1996 and period 1996- 2003 rules (emerged after land fragmentation phenomenon) has been performed by mean of decision tree.

In the following paragraph we present the results of these tests.

Multitemporal Data analysis

For the preliminary multitemporal statistical analysis we used OLAP (On-Line Analytical Process) Cube, a Datawarehouse technique, considering only four initial variables, namely

- Land use change class
- Time section of analysis (1991/96, 1996/03)
- Land slope class
- Land suitability

In this context the result of this first OLAP analysis show that:

- In the period 1991/1996 we have more stability for what concern land uses, with about 39% of correspondence and medium level land use changes homogeneously distributed within territory as regards various slope and "Land Suitability" classes.
- In the same period transformations are uniformly distributed between the different land use classes and slope categories. Moreover there aren't significant conversions of land use but only some medium level modifications.
- In the period between 1996/2003, differently from what happened before, portion of land presenting high values of slope have been abandoned (15%); it was probably because during regime, steep land management by mean of terracing was imposed.
- Moreover, private land management caused wood abandonment, such as agricultural lands, at the border of wood, little by little became bigger encroaching on wood areas: in this period transformation from wood to meadow and pasture is about 9,9%. Is furthermore interesting to note that there is a strong correspondence between slope classes and land use classes: steep lands are always related to less profitable land use classes, such as wood and meadow and pasture, and vice versa.

Mining Phase: Extraction of Rules

Afterwards, in order to better analysing time slots 1991- 1996 and 1996-2003, a decision tree technique was implemented. In this case, as input data, much more variables than the four considered for multitemporal analysis have been used, namely the 29 variables in Table 3.

The prediction model provided rather good results.

- In the period 1991-1996 the model presents a land use prediction accuracy of 87.5403 %, with 8136 Correctly Classified Instances and only 1158 Incorrectly Classified ones. The Mean absolute error is 0.0391.

- In the period 1996-2003 the model presents a land use prediction accuracy of 74.8387 %, with 7308 Correctly Classified Instances and 2457 Incorrectly Classified ones. The Mean absolute error is 0.0771.

The bigger error in the second period is due to the fact that, as described in next Paragraph, in this period land fragmentation generates more 'confuse' transformation dynamics , so that there are less regularities (the same conditions produced different actions).

Results: emerging territorial dynamics in Albania

From the comparison between rules related to 1991- 1996 period and rules related to 1996- 2003, presented in Tables below, it is possible to point out the difference between the two period and outline the emerging dynamics.

For what concern the period 1991- 1996 366 rules have been extracted, while almost double (719) is the number of those related to period 1996- 2003.

Some of the most representative ones are described. A more aggregated explanatory level was built, to express the extracted knowledge in a more synthetic and communicable way: for the two periods extracted rules have been divided in two groups, namely transformation and inertial rules, giving for each group the conditions that influenced the evolution or the preservation of the original land use.

Transformation rules are those having an impact in change dynamics; these are owed to planning activities or to a land use policy – should be considered in a Decision Support System the most critical aspect because it involves the prediction of the next future based on facts and events measured in time series. Inertial rules are those referred to land use stability.

In the following tables some extracted rules and the support (number of cells involved) of each one are reported:

'91	Principal Conditions	'96	N°
TC	High level of Olive grove in neighbourhood;	PC	28
	Low number of buildings in 500 m.; High level of residence in neighbourhood;	UA	68
	Medium level of uncultivated in neighbourhood; Original parcel medium low size; No water in neighbourhood;	UC	15
WA	No transport in neighbourhood; No water in neighbourhood;	TC	166
	Medium level of transport in neighbourhood;	T	90
PC	High level of uncultivated in neighbourhood;	U	167
LEGEND: TC= Temporary Cropping, PC= Permanent Cropping, WA= Water, WO= Wood, UC= Uncultivated, T= Transport, UA= Urban area, M= Meadow/pasture			

Table 4: Transformation rules in 1991-1996

'96 '03	Principal Conditions		N°	
WO	Low-medium number of buildings in 500 m.;	High level of wood in neighbourhood;	464	
		Poor road condition;	501	
	Medium level of wood in neighbourhood;		118	
	Medium low level of wood in neighbourhood;		63	
TC	High level of crop in neighbourhood;	Low level of Fruit tree in neighbourhood;	137	
		No Fruit tree in neighbourhood;	125	
		Low number of buildings in 500 m.;	223	
	Medium level of crop in neighbourhood; No buildings within 500 m.;	No Olive- grove in neighbourhood; No uncultivated in neighbourhood; No Water in neighbourhood; No Fruit trees in neighbourhood; No meadows/pasture in neighbourhood;	224 5	
		Medium level of crop in neighbourhood; No buildings within 500 m.;	264	
		Medium high level of crop in neighbourhood; No buildings within 500 m.;	No water in neighbourhood;	283
			No residence in neighbourhood; Low distance from the edge of cell;	93
	Poor road condition;	No distance from the edge of cell No Olive- grove in neighbourhood; No uncultivated in neighbourhood;	103	
		Medium high level of crop in neighbourhood;	No residence in neighbourhood; Fair road condition;	96
	Medium high level of crop in neighbourhood;		No buildings within 500 m.;	150
Low erosion risk; No Olive- grove in neighbourhood; No wood in neighbourhood; No Fruit trees in neighbourhood; Original parcel Low size; Low distance from artificial watering canal;				
PC	Near to Urban secondary road No meadow/pasture in neighbourhood; No residence in neighbourhood; No Water in neighbourhood; No uncultivated in neighbourhood; No Transport in neighbourhood;		173	
	Low level of uncultivated in neighbourhood; Medium level of Olive grove in neighbourhood; No wood in neighbourhood;		40	
	Medium high level of Olive grove in neighbourhood;		54	
T	Original parcel medium-low size No water in neighbourhood;		58	
M	Medium low level of meadow/pasture in neighbourhood; No vineyards in neighbourhood;		67	
LEGEND: TC= Temporary Cropping, PC= Permanent Cropping, WA= Water, WO= Wood, UC= Uncultivated, T= Transport, UA= Urban area, M= Meadow/pasture				

Table 5: Inertial rules in 1991-1996

'96	Principal Conditions		'03	N°	
TC	Slope < 25%; Presence of buildings;	Low level of fruit in neighbourhood;	UA	48	
		No olive grove in neighbourhood; Original parcel Low size No residence in neighbourhood; Gravel loose road in neighbouring		25	
		Medium high distance from artificial watering canal; Medium number of buildings in 500 m.;		14	
		Original parcel medium-low size		19	
	Low erosion risk Low distance from artificial watering canal; No transport in neighbourhood;	Original parcel Medium high size No natural watering canal; Unpaved road in neighbouring No wood in neighbourhood;		32	
		Original parcel Medium high size Natural watering canal inside; Gravel loose road in neighbouring		26	
TC	100%>Slope>75% Low number of buildings in 500 m. ; No transport in neighbourhood; Medium low erosion risk;	Original parcel High size; No meadow in neighbourhood;	U	36	
		Original parcel Low size; Low distance from nearest road;		45	
	100%>Slope>75%; Gravel loose road in neighbouring; Medium low erosion risk	Low number of buildings in 500 m. ; No transport in neighbourhood; Original parcel Medium size No residence in neighbourhood;		71	
		Very low drainage value; Medium low distance from natural watering canal;		19	
	Gravel loose road in neighbouring; No residence in neighbourhood;	Low number of buildings in 500 m ; No transport in neighbourhood; Original parcel Low size; No meadow in neighbourhood; Low level of olive in neighbourhood;		48	
		75%>Slope >50%; Original parcel Medium size;		26	
	Slope < 25% Low number of buildings in 500 m. ; No wood in neighbourhood; Near to urban secondary road No transport in neighbourhood; No residence in neighbourhood; Low drainage value No water in neighbourhood;				53
	Medium High distance from artificial watering canal; Medium distance from nearest road				21
	TC	100%>Slope>75% Near to Urban main road No meadow in neighbourhood;		PC	21
	WO	High distance from artificial watering canal; No transport in neighbourhood; No uncultivated in neighbourhood;		M	49
Very low drainage value		Medium High distance from artificial watering canal; No water in neighbourhood;	31		
		Original parcel Low size	23		

WO	Gravel loose road in neighbouring Poor road condition Original parcel Medium high size		M	93
PC	Medium low erosion risk Low drainage value No meadow in neighbourhood; No uncultivated in neighbourhood; Medium high number of buildings in 500 m.;		M	21
	Medium high level of fruit tree in neighbourhood;	Original parcel Medium size No additional condition condition	U	48 23
M	Paved road in in neighbourhood Original parcel medium-low size Medium High distance from artificial watering canal;		UA	38
U	Paved road in neighbouring Original parcel Medium size	No meadow in neighbourhood; No wood in neighbourhood No crop in neighbourhood;	PC	28 49
	Medium low distance from natural watering canal; Paved road in neighbourhood; Low distance from nearest road			UA
WA	Original parcel medium-low size	Medium distance from natural watering canal; Unpaved road in neighbourhood; High distance from artificial watering canal; No uncultivated in neighbourhood;	UA	51 25
	No olive grove in neighbourhood; Natural watering canal inside; Original parcel medium-low size			PC
LEGEND: TC= Temporary Cropping, PC= Permanent Cropping, WA= Water, WO= Wood, UC= Uncultivated, T= Transport, UA= Urban area, M= Meadow/pasture				

Table 6: Transformation rules in 1996-2003

'96 '03	Principal Conditions		N°
UA	Low distance from artificial watering canal; No water in neighbourhood;	No olive grove in neighbourhood; Original parcel Low size; No meadow in neighbourhood; Very low distance from nearest road; Low level of crop in neighbourhood;	124 49
	No artificial watering canal; No wood in neighbourhood; Near to urban secondary road Low number of buildings in 500 m.;		68
U	Gravel loose road in neighbouring Original parcel Low size Presence of buildings		51
	Unpaved road in neighbouring Low number of buildings in 500 m. ; Medium distance from nearest road		64
WO	Original parcel Medium high size High distance from artificial watering canal; Low number of buildings in 500 m. ;		47
	Original parcel High size Unpaved road in neighbouring	No residence in neighbourhood; No transport in neighbourhood;	394
		Medium High distance from artificial watering canal;	57

WO	Original parcel medium-low size Low distance from the edge of cell		89
	Medium low number of buildings in 500 m.; No residence in neighbourhood;		116
PC	No fruit in neighbourhood; No transport in neighbourhood; No water in neighbourhood; No wood in neighbourhood; No crop in neighbourhood; Medium distance from artificial watering canal;		54
	No fruit in neighbourhood; Paved road in neighbouring Medium High distance from artificial watering canal; Medium low number of buildings in 500 m.;		87
TC	Slope < 25%; Original parcel Low size; No buildings inside	Low distance from nearest road No water in neighbourhood; Low number of buildings in 500 m.; No residence in neighbourhood; Unpaved road in neighbouring	133
		Paved road in neighbouring Medium distance from artificial watering canal;	59
M	Gravel loose road in neighbouring Medium low erosion risk Original parcel medium-low size		98
LEGEND: TC= Temporary Cropping, PC= Permanent Cropping, WA= Water, WO= Wood, UC= Uncultivated, T= Transport, UA= Urban area, M= Meadow/pasture			

Table 7: Inertial rules in 1996-2003

The status of neighbourhood cells doesn't influence at all the transformation in the period 1996-2003 while, on the contrary, has a strong impact in the transformation of the previous period, i.e. 1992-1996; In the period 1991- 1996 we can extract almost exclusively inertial rules; the few dynamic rules are related to few cells (see supports of dynamic and inertial rules in Table 4 and 5), while in the period 1996- 2003 dynamic rules are much more than inertial rules

The above mentioned facts seem to lead to the following conclusions:

1. In the period 1991-1996, if transformations occurred, they were still influenced by a central planning policy, most likely due to the persisting influence of officials, technicians and experts working under the previous regime and still considered as a reference in the use of the land. Decisions were related to zones influencing clusters of parcels with a close connection between them. With the collapse of the central system and in the absence of any planning authority and without any improvement in the land market, the use of land was mainly preservation where environmental conditions were more favourable or a degradation where conditions where environmental conditions were less favourable.
2. With the beginning of the land market and corresponding lack of regulation and legislation (1996-2003), changes in the use of land were more dynamic but almost on an individual land-user basis without any planning or policy. Transformations are prevailing on inertial development, but without any plan.

The combination of inertial and transformation rules could be notably critical when both trends indicate negative developments at national level such as the trend confirming that individuals tend to exploit better environmental conditions for their own benefit while a planning policy should distribute resources and exploitations over the area in a well-balanced manner.

Indirectly these results should stimulate the Albanian government to develop a land use policy and strongly invest in land-use planning to prevent the permanent deterioration of the environment with non-reversible transformations.

FUTURE DEVELOPMENTS

The future development regards two main points:

1. Application of calibration and validation techniques;
2. Implementation of new methods to calculate the more likelihood land use values for unknown cells.

The first development is due to the consciousness that calibration in all these models is a central need that requires special attention. It is possible to use, for the calibration, deterministic optimization methods or, and perhaps better, probabilistic algorithms as the evolutionary ones [15]. It is important to verify the sensitivity of the system according to two main parameters: the cell dimension and the neighbourhood type and dimension.

The calibration will be carried out applying the Cellular Automata starting from the 1996 land use map and analysing the classification errors at the 2003.

The second point rises from the problems encountered in the unknown cells classification using the Decision Tree Induction tools. As previous mentioned, not all the cells verify an IF statement or a logical reasoning in order to allow their classification. It is then necessary to find out a method useful for the whole set of cells. From the review of the techniques applied in recent researches two main directions can be deduced:

- the construction, for each cell, of a potential function [17] indicating the likelihood that in a cell there will be each land use;
- the use of statistical regression in order to calculate each land use future development, the relative portion of territory required and the consequent disaggregation [16].

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the EU Phare Land Use Policy II project (www.lup2.org) of the Ministry of Agriculture and Food (MoAF) in Tirana, Albania, and AGROTEC S.p.A. (www.agrotec-spa.net), Rome, Italy, which was responsible for the project execution. Permission to publish was granted by the EU Phare Programme Management Unit and MoAF.

REFERENCES

1. **Agrotec S.p.A. Consortium**, LUP geo-database model - version 1.0. Technical report No. 1. EU Phare Project 98-0502 Land Use Policy II, 2003 (author: G. Carrai)
2. **Agrotec S.p.A. Consortium**, A standard methodology for land-use description and harmonisation between data sets. Technical report No. 2. EU Phare Project 98-0502 Land Use Policy II, 2003 (author: L.J.M. Jansen)
3. **Amato G.**, Valutazione dei risultati della classificazione, notes from the university course in Computational Economy, University of Chieti-Pescara, 2004
4. **Bonchi F., Lombardo S., Pecori S., Santucci A.**, Learning about Land Use change in Rome and in Pisa Urban areas. In L.Diappi (ed), *Evolving Cities: Geocomputation in Territorial Planning*. Ashgate, England, 2004
5. **Bonchi F., Pecori S.**, Knowledge discovery and data mining to investigate urban and territorial evolution: tools and methodologies. In L.Diappi (ed), *Evolving Cities: Geocomputation in Territorial Planning*. Ashgate, England, 2004
6. **Burrough P.A., McDonnell R.A.**, *Principles of Geographical Information System*. Oxford University press, Oxford, 1997

7. **Fayyad U., Piatetsky- Shapiro G., Smith P., Uturusamy R.**, Advances in knowledge Discovery and Data Mining. AAAI/MIT Press, 1997
8. **Han J., Kamber M.**, Data mining: concepts and techniques. Morgan Kaufmann, S.Francisco, 2000.
9. **Holsheimer M., Siebes A.**, Data mining: the search for knowledge in databases. CWI Technical Report CS-R9406, Amsterdam, 1994
10. **Lombardo S., Pecori S.**, Investigating Urban structure and evolution through Artificial Intelligence tools. Studies in Regional & Urban Planning, 2004
11. **Lombardo S., Pecori S., Petri M.**, Investigating territorial dynamic using Decision Tree, paper presented at the 9th International Conference of Computers in Urban Planning and Urban Management (C.U.P.U.M.), London the 29th june-1st july 2005
12. **Pecori S., Santucci A.**, The competition for land between urban and rural systems investigated by G.I.S. and Data Mining techniques. Proceedings of 3rd International Conference of Decision Making in Urban and Civil Engineering, London, November 6 - 8, 2002
13. **Piatetsky-Shapiro G., Frawley W. J.**, Knowledge Discovery in Databases. AAAI/MIT Press, 1991
14. **Quinlan J. R.**, Induction of decision trees. Machine Learning, 1986, pp. 81-106
15. **Santillana M., Serrano F.**, Calibration and validation of a CA based model for urban development simulation using an evolutionary algorithm. A case study in Mexico City, paper presented at the 9th International Conference of Computers in Urban Planning and Urban Management (C.U.P.U.M.), London the 29th june-1st july 2005
16. **Singh A.K.**, Modeling land use cover changes using cellular automata in a geo-spatial environment, thesis submitted to the International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands, 2003
17. **White R. & Engelen G.**, High-resolution integrated modelling of the spatial dynamics of urban and regional systems, Pergamon press, Computers, Environment and Urban Systems 24, pp. 383-400, 2000
18. **Witten I.H., Frank E.**, WEKA-Machine Learning Algorithms in Java, Morgan Kaufmann, S. Francisco, 2000.

AUTHORS INFORMATION

Giancarlo Carrai

gc.carrai@svaltec.it
Svaltec s.r.l., via del Campofiore 106, 50136,
Florence, Italy

Louisa J.M. Jansen

louisa.jansen@tin.it,
Via Girolamo Dandini 21, 00154 Rome, Italy

Serena Pecori

s.pecori@ing.unipi.it
L.I.S.T.A. – Laboratorio di Ingegneria dei Sistemi
Territoriali e Ambientali,
University of Pisa, Department of Civil Engineering

Massimiliano Petri

m.petri@ing.unipi.it
L.I.S.T.A. – Laboratorio di Ingegneria dei Sistemi
Territoriali e Ambientali,
University of Pisa, Department of Civil Engineering