CoastGIS 2009
1–12

A Bottom Up Approach to the Modelling of Coastal and Land Use Evolution through GIS

Joana Simões†‡, Jorge Rocha•, José Carlos Ferreira⋆, António José Tenedório‡ and Paulo Morgado•

†e-GEO, Faculdade de Ciências Sociais e Humanas, UNL Avenida de Berna, 26-C 1069-061 Lisboa Portugal
‡ CASA, University College London, 1 - 19 Torrington Place, London WC1E 7HB, UK
⋆ CENSE - Center for Environmental and Sustainability Research, DCEA, FCT, Universidade Nova de Lisboa, Campus da Caparica, 2829-516 Caparica, Portugal
• IGOT/Universidade de Lisboa Alameda da Universidade, 1600-214 Lisboa, Portugal

(*v1.0 released September 2009*)

Coastal areas stand in the intersection of human and physical factors, some of them interacting in a non linear fashion and presenting feedbacks, which typically are characteristics of complex systems. This kind of systems are very difficult to model in a top-down approach, due to the high number of variables in question and the nature of relations between them. The spatial dimension itself, also introduces some complexity which is difficult to capture in a deterministic approach.
In this study, we propose a bottom-up approach for modelling the coastline evolution, integrating land use cover. The system is based in a probabilistic Cellular Automata (CA) model, which divides the study area in a grid. The hermetic structure of CA is overcomed by using transition rules whose weights have been calibrated by a Artificial Neural Network (ANN). In this way, the knowledge of past events is incorporated into the model and projected into the future, by means of using "intelligence"; this contrasts with techniques such as linear regression, whose efficacy in the case study was evaluated to be much inferior.
Finally, is important to mention that the use of a Geographical information Systems (GIS) environment, enabled to assemble together different types of data and overlap them in an efficient way, which would be very hard or impossible to do otherwise; therefore, we believe that the use of GIS is crucial in spatial based simulations. The case study for this model was an area of the Municipality of Almada (Portugal) that has an extensive coastal line, both Atlantic and estuary. The modelling of spatially dynamic and naturally complex phenomena occurring in these coastal areas, is important for the definition of an innovative strategy for their physical planning and also their environmental management.

Keywords: GIS, Artificial Neural Networks, Planning, Cellular Automata, Coastal Areas;

1 Motivation for this Study

As landscape became humanized, there were many cases in which the intervenients in this process were less careful about using a more passive approach regarding to nature. Therefore human action, side by side with natural agents, is leaving its strong marks in the landscape (Rocha *et al.* 2007).
Nowadays, the idea of a sustainable development of the landscape has gain some importance; it relies in the maintenance, preservation and recovery of the spaces that are vital to the ecological balance of the land, and for that purpose methods of sustainable land use and occupation are created and used.
This study establishes and implements a methodology for evaluating land use and occupation of developed coastal areas, subjected to human pressure. The motivation for this task was, in one hand to provide answers for assuring a balanced organization and management of the occupation patterns of coastal areas, and on the other hand to allow the safeguarding and value improvement of the existing natural resources.
Human-environment systems, such as the ones occurring in the coastal areas, are characterized by heterogeneity, non-linear relationships, and hierarchical structures that give rise to difficulties in understanding system behavior in response to exogenous factors (An *et al.* 2005). This kind of systems, that we can

*Corresponding author. Email: j.simoes@ucl.ac.uk

define broadly as **complex systems**, are normally composed of many parts which are coupled together in a non-linear fashion and typically present **self organization**, a behavior in which order may arise from low level interactions without any supervision from higher-order structures (Nowak and Lewenstein 1996). The bottom-up approach seems the natural way of studying them, since it considers that spatial extended systems are capable of non-trivial collective behavior - unexpected behavior which is observed in macroscopic quantities. Since this kind of emergent properties cannot be trivially derived from the properties of individual elements, it is difficult to predict them. The use of computer simulations allows the precise study of the dynamical consequences of models which cannot be solved by analytical methods, and therefore the importance of computation in this kind of approach. Multi-agent systems, which include agent-based models (ABM) and **Cellular Automata** (CA) - that were used in the present study - are examples of Artificial Life (ALife) techniques that try to mimic life processes to understand the appearance of single phenomena (Simoes 2007). On figure 1, page 2, we can see how this three methods relate as subsets of each other.

As ABM, CA models typically consist of an environment in which the interactions occur, and individuals which are defined in terms of their behaviors (procedural rules) and characteristic parameters. Unlike ABM, CA are characterized as being spatially-explicit.

These models may and should, be verifiable so as to enable validation with the use of collected data and

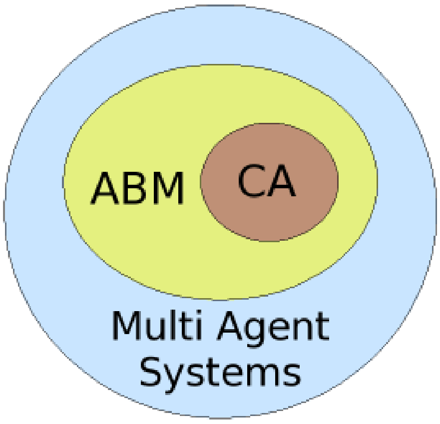


Figure 1. Relation between multi-agent systems, ABM and CA;

various statistical techniques. Thus, the study of the spatial distribution patterns (represented by dots, lines, areas or matrixes to which events are translated) constitutes the basis for the spatial-quantitative studies.

Our case study is located in the Municipality of Almada, which provides a good example of a coastal area with an extensive coastal line - both Atlantic and estuary - that has been subject of intensive growth and development; we can see an image of this area on figure 2, page 3; unfortunately, the growth of a region is often mistaken as economic growth. This innacuracy has tremendous consequences, especially when taking into account the principles for defining methods of landscape management and planning, with the repercussions of such errors being translated into spatial asymmetries that mirror the production of functional, social and landscape segregation.

Furthermore, we believe that the modelling of this kind of spatially dynamic and naturally complex phenomena, should be relevant for the definition of innovative strategies of physical planning and environmental management of the coastal areas.

Finally, is important to emphasize that in this work we tried to escape from the static representation of space that is usually tied to tradional cartography and also to GIS; the temporal component, that is so essential for the purposes we established here, can only be added by means of simulation.

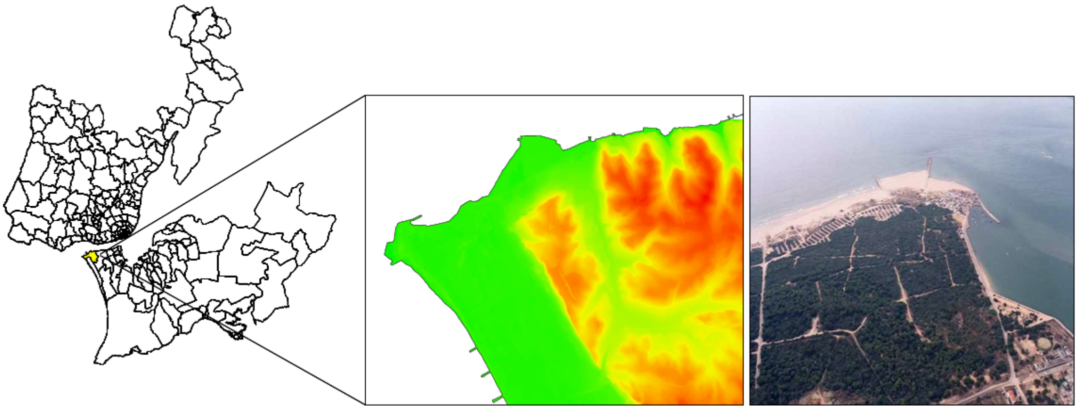


Figure 2. Case Study: Cova do Vapor region;

2 Data and Methods

The simulation technique followed in this study - CA - presents a set of cells (pixels) in interaction, each one of them being a computer (automata); an example of this can be seen on figure 3, page 3. CA can be broadly defined as discrete dynamical systems (in space and time) whose behavior is completely specified in terms of a local relation (Margolus 1987). These local relations - **transition rules** - are of vital importance to the model and there are different ways of defining them, that result in substancially different models. The determination of the influence of each variable (in terms of weights) can be an extremely complex task, if we relax a binary state representation (like the one used in urban growth models, from the perspective rural-urban) and use continuous values; such is the case of scenarios with different land use/cover, in competition between them for the territory.

In tour case study, we have identified five groups of factors that are in the basis for the allocation of land

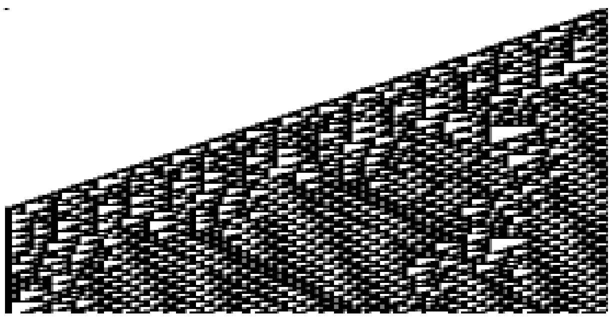


Figure 3. Example of a CA, generated by *Wolfram*'s rule 110;

use:

- Environmental characteristics.
- Neighborhood characteristics at a local scale.
- Spatial characteristics of the urban areas (e.g. accessibility).
- Regional and urban planning policies.
- Social-economic and psychological factors: related with individual preferences, level of economic development, socio-economic and political systems.

As mentioned before, the determination of the weights for each factor (i.e. balance factor that is associated to it) is of great importance to the resulting patterns of the simulation, and there are many methods to achieve it. In this study we used an Artificial Intelligence (AI) technique - Artificial Neural Networks (ANN) - that proved to be very efficient in this particular case. Neural networks are non-linear statistical data modeling tools that can be used to model complex relationships between mapped inputs and outputs

or to find patterns in data. The Multilayer Perceptron (MLP) is perhaps one of the most widely used ANN, and it can be trained by an algorithm called **backpropagation** (W.K.T. Hau 1998) (BP); this is a gradient-descent algorithm that minimizes the average squared error between the network outputs and the desired outputs.
The most common MLP architecture is a three layer network, involving:

- Input layer.
- Hidden layer.
- Output layer.

The input layer is responsible for feeding data into the network; the hidden layer contains the processing elements (PES) that process this information and generate the weights of the connections on this layer; the output layer feeds the desired output back into the network.
This process of feeding forward the signal and back propagating the errors is iterative and stops only when the error stabilizes at a low level.
On table 1 we can see the common specifications of the employed ANN.

Table 1. Common specifications of the employed ANN.

Attribute	*Value*
Neural model	Multilayer perceptrons (MLP)
Learning algorithm	Back propagation
Input PEs	Base land use and prediction variables
Output PEs	Land use/cover

3 Description of the Model

Following the theoretical approach described on section 2, this model was implemented following five sequential stages. Figure 4, page 4 presents an activity diagram, with a step-by-step workflow of components in the system.

Figure 4. UML representation of the model;

The first phase concerns the processing and codification of information, in order to create spatial layers of information, based on the forecast variables; these layers are the input of the model. The second phase focuses in the use of spatial rules that relate the forecast variables with the land use changes; the third phase integrates all layers of information using one of three techniques (logistic regression, multi-criteria analysis and ANN); the fourth phase sorts out the data in order to create a time series, that ultimately allows the forecasting of land use; maybe it is important to note that the third and fourth stages are relatively independent, and therefore the order in which they are applied is not really relevant for the model.
In the following subsections we are going to look at each one of these phases with a bit more detail.

3.1 *Pixel-value Differencing*

As privileged tools for processing geographic information within an integrated environment, and manipulating geographical information in digital form (Longley *et al.* 1999), GIS seemed like the best option for generating the input data for this model. From a collection of base layers in different formats (raster and vector) it was created a database with the intention of supplying the basic spatial information for the simulation. The information layers represent diverse thematic maps such as the land use/cover multi-temporal dataset (1815, 1940, 1991 and 1995) (figure 5, page 5) and the 2004 base map (figure 6, page 6), the topography, landscape elements (e.g. roads, shore line), Agriculture National Reserve (RAN), Ecological National Reserve (REN) municipality master plan and erosion susceptibility map. As CA are grid based (or raster based), the information in vector format was discretized into a 10x10 pixel grid. Finally the pixels were codified either to represent binary states (constraints), or continuous realizations of a variable (probabilities of occurrence).

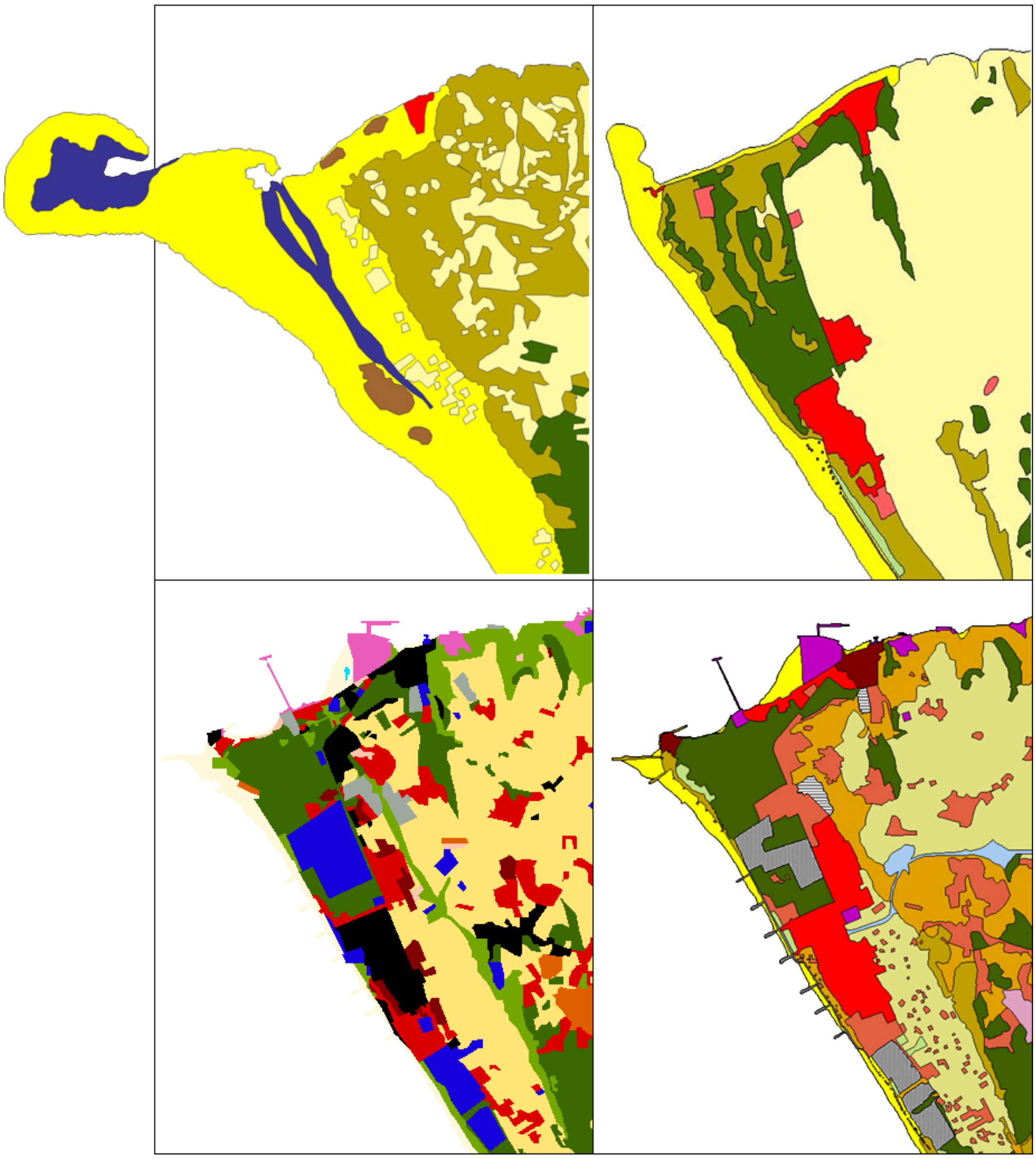


Figure 5. Multi-temporal Land Use/Cover dataset: from left to right 1815, 1940, 1991 and 1995;

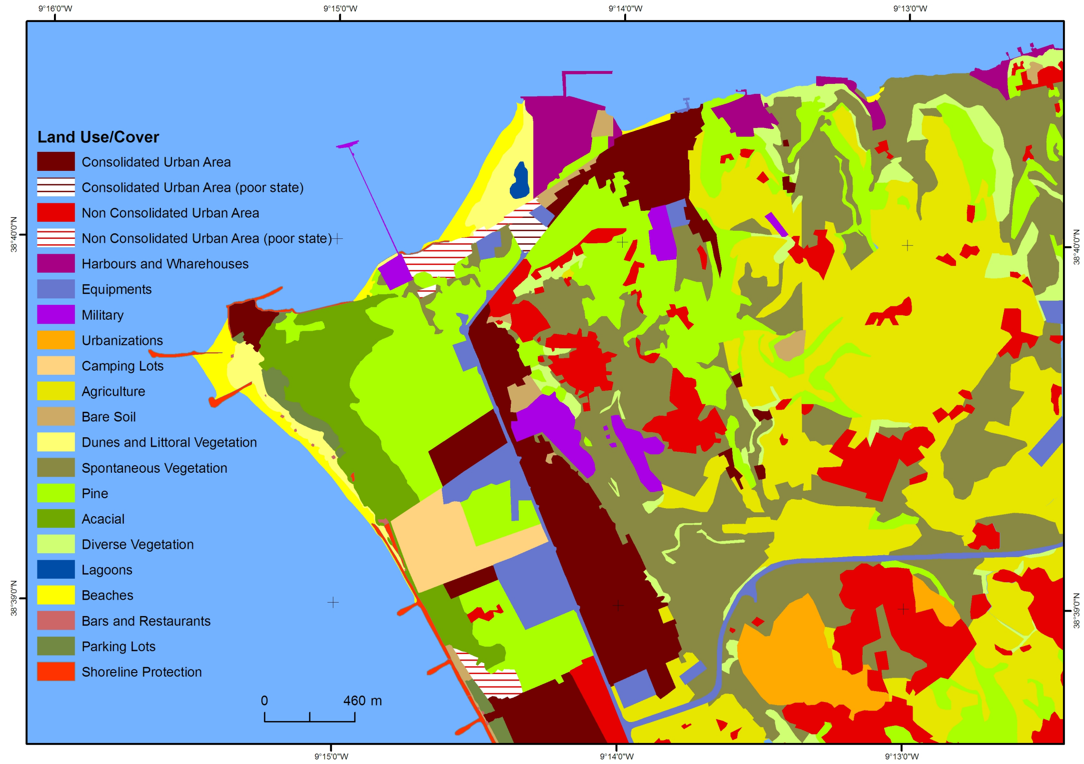


Figure 6. Land use/cover map for Cova do Vapor in 2004;

3.2 *Generating Transition Rules*

As mentioned in section 2 transition rules are the core of the CA model as they are responsible for the state of the cell, at each time step; they quantify the spatial effect that the forecast cells contain in the land use/cover changes (Pijanowski 2002) and we can group them into two classes:

- Neighborhoods or densities.
- Distance to the forecast cells.

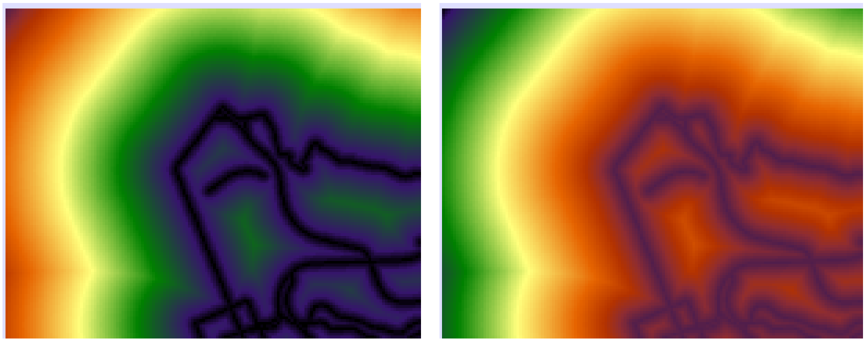
The neighborhood effect is based on the premise that the update of a given cell requires the knowledge of the state of the cells in its vicinity (Chopard and Droz 1998). The region in which a cell needs to search is called *neighborhood*. For 2D CA, two neighborhoods are often considered: the *Von Neumann*, which considers of the four geographical neighbors: North, South, East and West, and the *Moore* neighborhood, which adds the second nearest neighbors: North-east, North-west, South-east and South-west. In this study, we used a combination of the two, that results in 12 adjacent cells, that we see represented on figure 7, page 6, in red.

0	0	1	0	0
0	1	1	1	0
1	1	1	1	1
0	1	1	1	0
0	0	1	0	0

Figure 7. Structure of the neighborhood used in this model;

Rules of spatial transition based on distance, relate the Euclidean distance between each cell and the nearest forecast variable.

It is important to make a parenthesis here, to emphasize that many real world phenomena sometimes cannot be captured by crisply defined rules but rather through vague concepts that accommodate certain positions and/or boundaries that change and and/or move over time (Palancioglu 2003). This kind of *vagueness* is called **fuziness** and it can be approached at the light of *Fuzzy Set Theory*; this theory captures the nature of many spatio-temporal phenomena which do not have sharp boundaries or exact positions or whose boundaries or positions cannot be precisely determined, such as the case of variables in our model. The fuzzy approach uses membership functions to distibute values between sets: in this case, different functions were used to normalize the distance values, attributing memberships characteristics to cells. For example, to predict the shoreline change, the distance to shoreline was normalized with a linear function because as we move away from the coast the probability of the terrain being affected by the sea is lower. Alternatively, the distance to the road network, to predict urban proliferation, was normalized with the sigmoid function, for the reason that in some surveys carried out using the local population, people state that they prefer to live 500 m from a road and absolutely dislike being as far away as 1500 m. On figure 8, page 7, we see how the GIS displays the Euclidean and fuzzy distances.



(a) (b)

Figure 8. Distance representation; In a) we see a the Euclidian distance and in b) we see the distance normalized through the fuzzy approach;

Finally, we had to create a constraint mechanism to prevent changes in particular locations. This becomes necessary in areas inside of which a certain land use expansion (e.g. urban) is forbidden (e.g. National Ecological Reserve - REN). In these cases we used a binary approach, attributing a logical value of *0* to all the cells where the change cannot occur, and a logical value of *1* to all the remainders. Later, the product of all these levels was calculated, engendering a single level corresponding to the *exclusion zones*.

3.3 *Variable Integration*

The integration of the forecast variables (weights evaluation) can typically be carried out through logistic regression, multi-criteria analysis (Pijanowski 2000) or neural networks (Pijanowski 2002), having in mind that each of these processes requires a different form of data normalization. As mentioned in 2, page 3 in this study we opted for a ANN with a MLP.

The choice of the number of hidden layers was constrained by speed, since the simulation covers a large time span; although some studies point out that difficult learning tasks can be simplified by increasing the number of hidden layers, a three-layer network can form any decision boundaries, through varying only the number of hidden neurons. According to *Kolmogorov*'s theorem, the use of $2n + 1$ hidden neurons can guarantee a perfect fit of any continuous functions and reducing the number of neurons may lead to lesser accuracy. However, when dealing with real world data, $2n + 1$ hidden neurons may be a extremely large number. A solution of $2n/3$ hidden neurons can generate results of similar accuracy with much less training time. In this model, nine hidden neurons were used, which is a compromise between accuracy and simulation speed.

Thus, we have an input layer with n neurons corresponding to base land use (1995) and all the prediction variables, a hidden layer and an output layer with 2004 land use/cover. The next step was to determine the input weights using the BP algorithm, which randomly selects initial weights and then compares the calculated output for a given observation with the expected output for that observation. The differences between the expected and the calculated output values across all the observations is summarized using mean square error. After all observations are presented to the network, the weights are modified so that the total error is distributed among the various nodes in the network. The final result is a transition probability map (for each cell and for each land use/cover class), obtained through the weight sum of all the transition values derived from the prediction variables. Figure 9, page 8, shows the structure of the particular ANN used in this study.

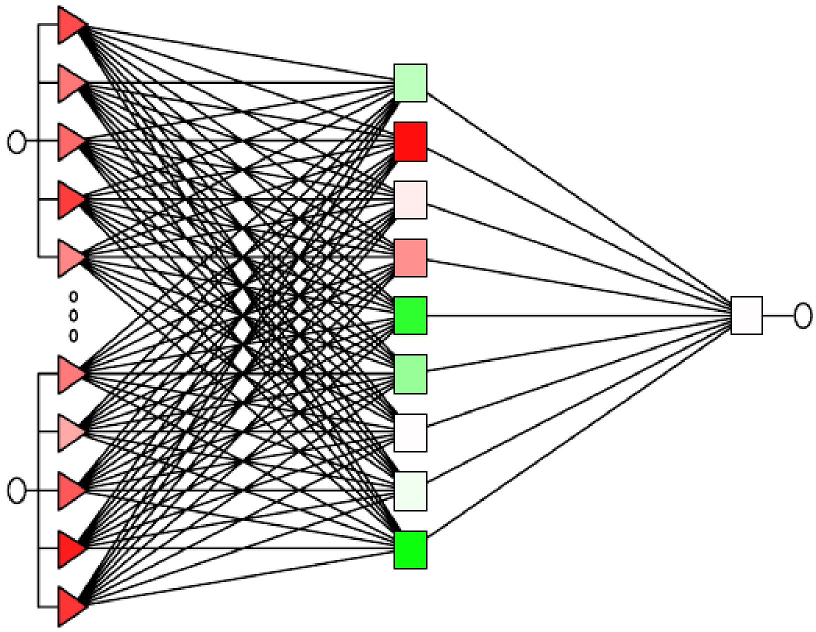


Figure 9. Structure of ANN used in this model;

3.4 *Temporal Indexing*

The decision of which datasets to use was constrained by the data itself. The most recent dataset available was 2004, but the former one could be: 1815, 1940, 1991 or 1995. The oldest one (1815) was impossible to use due to lack of accuracy (100 metre positioning error) and the 1991 one was unusable because it has an artificial shoreline, extracted from the 1:25000 Portuguese topographic chart. Finally, we opted for the 1995 dataset because it is the most recent and most accurate. This solution gives the most reliable results, but it narrows the forecast to 2013, since training the algorithm with nine years implies a forecast span of nine years. Having decided which datasets to use (1995 and 2004) it was necessary to carry out a semantic generalization, shown in figure 10, page 9, in order to make the land use/cover classes compatible in the two charts. The options lay between six super classes: agriculture, beach, vegetation, bare soil, urban and water.

The amount of cells (land) that are due to change to a different land use can be calculated with methods such as the Principal Components Analysis (Li and Yeh 2002) or the **Markov Chains**. The *Markov* approach - which was the one adopted in this model - establishes that for a variable with the *Markov* property the only information about the past needed to predict the future is its current state; knowledge of the values of earlier states **do not change the transition probability** (Walsh 2004). In this case the transition probability matrix is the result of crossing two images - initial and final land use/cover - adjusted by proportional error. The transition area matrix is a result of the product of each column and

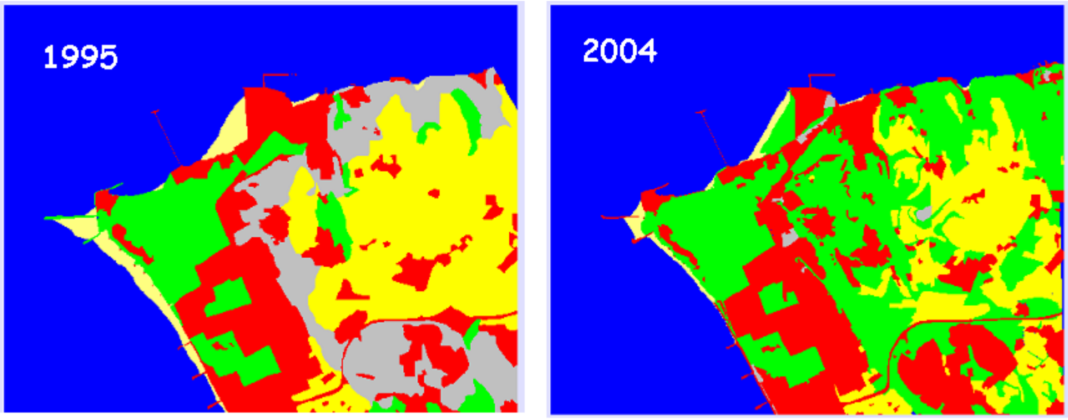


Figure 10. Semantic generalization of the land use/cover datasets from Cova do Vapor;

the number of corresponding land use/cover cells in the oldest image. The proportional error expresses the probability of the land use/cover classes in the input maps being inaccurate. The conditional probabilities obtained on exit are multiplied σ(1), where ς is the proportional error.

$$\sigma = 1 - \varsigma \tag{1}$$

Table 2 shows the final conditional probabilities, used in this model.

Table 2. Final Conditional Probabilities.

Use	*Agriculture*	*Beach*	*Vegetation*	*BareSoil*	*Urban*	*Water*
Agriculture	0.5330	0.0000	0.4132	0.0045	0.0406	0.0086
Beach	0.0000	0.2582	0.2334	0.0000	0.0134	0.4950
Vegetation	0.0131	0.0077	0.8984	0.0001	0.0777	0.0030
Bare soil	0.1832	0.0005	0.6138	0.0087	0.1901	0.0036
Urban	0.0214	0.0006	0.1094	0.0131	0.8501	0.0055
Water	0.0000	0.0024	0.0047	0.0005	0.0031	0.9893

4 Case Study

As mentioned in section , our case study for applying this methodology is a region in the Lisbon Metropolitan Area - more precisely in the Almada municipality - called *Cova do Vapor* (*Steam Hollow*). The choice of this municipality is based on an important range of factors, that turn this region into a unique example of strong human intervention in a particularly fragile natural environment; namely:

- Its proximity to the country's capital (Lisbon).
- The increased accessibility provided by the construction of the first Tagus bridge;
- The possibility of urban expansion, and the rare natural potential.

Due to all these factors, between 1967 and 1986, the territory had experienced strong urban pressure, resulting in an irreversible ground occupation in areas with important physical restrictions. Between 1960 and 1991 its population did not stop growing, despite it happened at different rhythms. Job dependency from the captial is a reality that the *pendular* movements between Lisbon and Almada corroborate. Furthermore, its beaches had emerged as an alternative to the ones to be found alongside the Lisbon-Cascais train line and therefore, second residence also took on great importance as another form of territorial occupation.

On the other hand, Cova do Vapor remains still a relatively untouched area - alongside with the rest of the Almada coastline - with a major natural hazard. In fact, if in the remaining territory of the municipality this hazard comes from both natural and human factors, in Cova do Vapor the natural phenomena are still pre-eminent and the few fishermen that live there are more victims then enhancers of the adverse situations. The main risk in this area is related to shoreline retreat, as its stated by a spetacular 410 m retreat between 1940 and 1995, that we can see represented on figure 11, page 10.

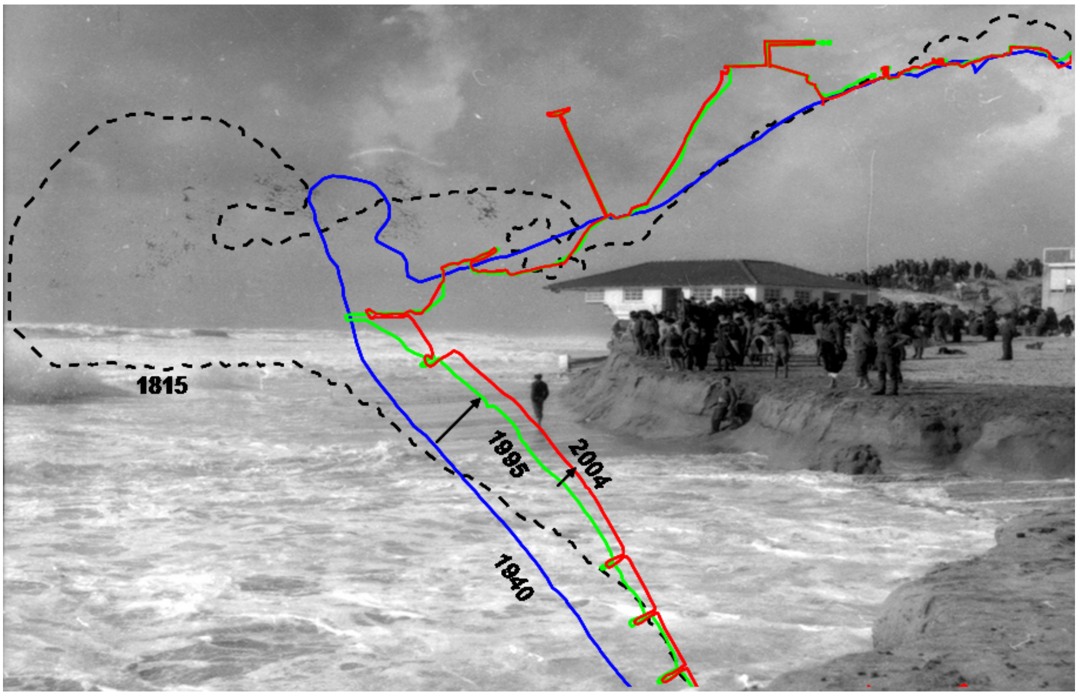


Figure 11. Schematic representation of shoreline retreat, between 1815 and 2004;

In terms of land use, the charts from 1986 show a highly differentiate territory (figure 5, page 5). Within a 7156.51 ha area, 35% indicates urban occupation and within this, 15,7% refers exclusively to residential areas. In fact, this occupies the third place on the land use ranking (1125 ha), just after agriculture (2115.67 ha) and forest (1548.95 ha). This trend was kept between 1981 and 1991, although at a slower pace. There was an increase in *artificial* land use, as opposed to *natural* land use; this has expression on a lost of importance of agriculture land use and a gain of importance in residential land use, with emphasis on occupation by secondary residential units. Unfortunately this strong trend in building residences, was not accompanied by the desirable similar growth in infrastructures.
The model described on section 3 forecasts a major recession of the shoreline in only nine years, especially to the North (40 m) and the West (20 m); we can see this output on figure 12, page 11. Concerning land use the situation will be relatively stable, just with some agriculture left over with those spaces to be occupied by vegetation.

Finally, it should be noted that to validate this results, we applied exactly the same methodology to the 1940-1995 dataset, with the aim of predicting 1995 land use/cover map. Then the observed 1995 map was crossed with the predicted one, resulting in an overall accuracy of 86% (14% of error in 55 years). We also tested different approaches for weights evaluation (described on subsection 3.3, page 7) : multi-criteria and logistic regression, obtaining, respectively, 78% and 60% overall accuracy.

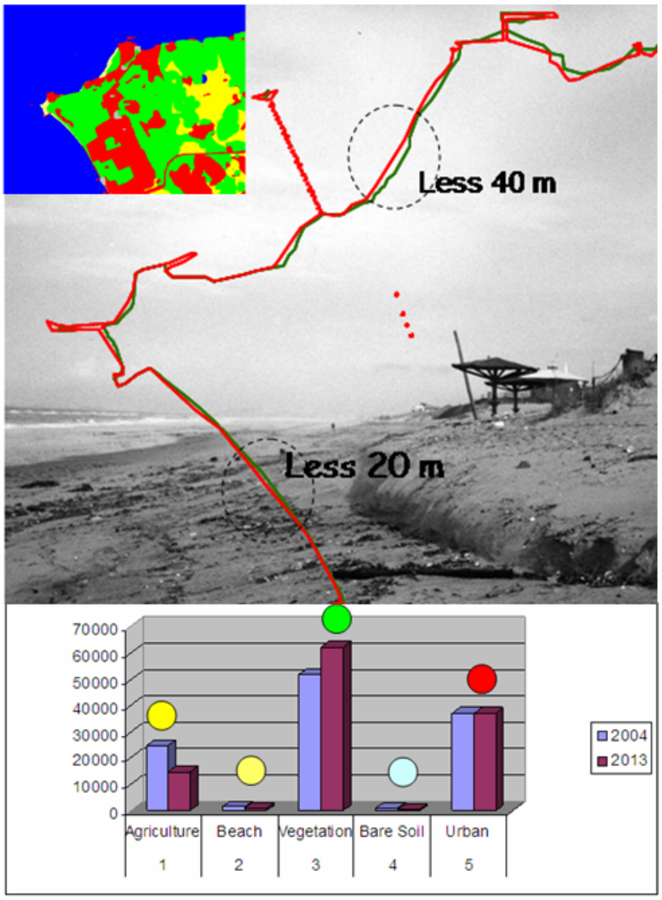


Figure 12. Shoreline predictable drawback between 2004 and 2013;

5 Conclusions and Final Remarks

This article shows an example of how a bottom-up approach is able to capture and model the complexity of a dynamic system such as a coastal area. This particular methodology, can overcome some of the difficulties traditionally experienced in CA models that are usually related to defining the variable weights, the transition rules and the model structures. Moreover, we hope to have demonstrated that *non-deterministic* approaches such as fuzzy logic and ANN can conveniently be incorporated within CA models, in order to simulate the evolution of multiple land uses/covers.
It is never enough to praise the relevance of GIS in this kind of spatial simulations as it allows the easy acquiring of data for *training* the model and obtain the parameters referred on section 3.3, page 7.
Finally, it is important to emphasize that this approach is able to cope with incomplete and erroneous input data, and that its non-linear generated forecast surface creates an obviously high number of probabilities in comparison to the ones produced by traditional linear regression models. it is also important to add that in a great deal of geographic phenomena, variables are correlated and therefore traditional methods (e.g. multi-criteria analysis) are inadequate to evaluate their weights; In a CA-ANN model, spatial variables **do not** necessarily need to be independent from each other which validates this methodology as a good approach, not only for land use modelling, but also for spatial modelling in general.

References

An, L., Brown, D., Rand, W. and Page, S., 2005, Can statistical methods on land-Use change patterns calibrate agent-based models?. In *Proceedings of the 8th International Conference on GeoComputation*, University of Michigan, Eastern Michigan University, USA, August http://www.cscs.umich.edu/research/projects/sluce/publications/An_geoco%mp_%20abs.pdf; accessed 23 August, 2006.

Chopard, B. and Droz, M., 1998, *Cellular automata modelling of physical systems*, Collection Alea-Saclay: Monographs and Texts in Statistical Physics (Cambridge, UK: Cambridge University Press).

Li, X. and Yeh, A., 2002, Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, pp. 323–343.

Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W., 1999, *Geographical information systems - volume 1: principles and technical issues* (New York, USA: John Wiley & Sons, Inc.).

Margolus, T.T.N., 1987, *Cellular automata Machines: a new environment for modeling* (Cambridge, MA, USA: MIT Press).

Nowak, A. and Lewenstein, M., 1996In *Computer simulations from the philosophy of science point of view* (Dodrecht, Netherlands: Kluwer Academic Publishers) http://cognitrn.psych.indiana.edu/rgoldsto/complex/nowak.pdf; accessed 23 August, 2006, pp. 249–285.

Palancioglu, H.M., 2003, Extracting Movement Patterns Using Fuzzy and Neuro-fuzzy Approaches. PhD thesis, The Graduate School, The University of Maine http://www.library.umaine.edu/theses/theses.asp?Cmd=abstract&ID=SIE2003%-002.

Pijanowski, B.C., 2000, A land transformation model: integrating policy, socioeconomics and environmental drivers using a geographic information system. *Harris, L. and Sanderson, J. (eds.), Landscape ecology: a top down approach*, pp. 183–198.

Pijanowski, B.C., 2002, Using neural networks and GIS to forecast land use changes: a Land Transformation Model. *Computers, Environment and Urban Systems*, pp. 553–575.

Rocha, J., Ferreira, J.C., Simoes, J. and Tenedorio, J.A., 2007, Modelling Coastal and Land Use Evolution Patterns through Neural Network and Cellular Automata Integration. In *Proceedings of the* , 50, Gold Coast, Australia (Journal of Coastal Research).

Simoes, J., 2007, An agent-based approach to spatial epidemics through GIS. PhD thesis, Department of Geography, University College of London, University College of London, Gower Street, London WC1E 6BT UK http://www.casa.ucl.ac.uk/JoanaMargarida/thesis_post_viva.pdf.

Walsh, B., 2004, Markov Chain, Monte Carlo and Gibbs Sampling. Technical report.

W.K.T. Hau, I.C. Bruce, L.S.E.C., 1998, Motor programs: an artificial neural network approach. In *Proceedings of the* , 3, Hong Kong, China, October (Engineering in Medicine and Biology Society).