MARKOV PROCESSES IN MODELING LAND USE AND LAND COVER CHANGES IN SINTRA-CASCAIS, PORTUGAL

PROCESOS DE MARKOV EN LA MODELIZACIÓN DE ALTERACIONES DEL USO E OCUPACIÓN DEL SUELO EN SINTRA-CASCAIS, PORTUGAL

PEDRO CABRAL

Instituto Superior de Estatística e Gestão de Informação, ISEGI, Universidade Nova de Lisboa, Portugal, pcabral@isegi.unl.pt

ALEXANDER ZAMYATIN

Tomsk Polytechnic University, Tomsk, Russia, zamyatin@tpu.ru

Recibido para revisar julio 10 de 2008, aceptado noviembre 13 de 2008, versión final noviembre 18 de 2008

ABSTRACT: In this article the process of land use and land cover change (LUCC) is investigated using remote sensing and Markov chains for the municipalities of Sintra and Cascais (Portugal) between years 1989 and 2000. The role of the Natural Park of Sintra-Cascais (PNSC) in LUCC dynamics is evaluated. Results show that, inside PNSC, present LUCC depends on the immediate past land use and land cover following a Markovian behavior. Outside PNSC, LUCC change is random and does not follow a Markovian process. Estimates of LUCC for year 2006 are presented for the area inside the PNSC. These results reinforce the role of the PNSC as an indispensable tool for preserving LUCC stability and to guarantee its functions.

KEYWORDS: Markov chains, LUCC, Remote sensing, Environmental monitoring.

RESUMEN: En este artículo los procesos de alteración de la utilización y ocupación del suelo (LUCC) son investigados recorriendo-se a técnicas de teledetección y a cadenas de Markov en las municipalidades de Sintra y Cascais (Portugal) entre los anos de 1989 y 2000. El papel del Parque Natural de Sintra-Cascais (PNSC) es evaluado. Los resultados demuestran que, dentro del PNSC, el LUCC presente depende del pasado inmediato del uso y ocupación del suelo siguiendo un comportamiento Markoviano. Fuera del PNSC, LUCC es aleatorio y no sigue un proceso Markoviano. Estimativas del LUCC para el ano de 2006 son presentadas para el área dentro del PNSC. Estos resultados refuerzan el papel del PNSC como una herramienta indispensable para preservar la estabilidad del LUCC y garantizar sus funciones.

PALABRAS CLAVE: Cadenas de Markov, LUCC, Teledetección, Monitorización ambiental.

1. INTRODUCTION

Remote sensing can be used to acquire spatio-temporal series of geographical data and to perform land use and land cover change (LUCC) analysis [1-8]. Obtained data can be processed using geographical information system (GIS) techniques and varied modelling approaches thus providing useful information for environmental monitoring and analysis [9-14]. In this study, stochastic modeling with Markov chains is the approach selected for studying LUCC in the municipalities of Sintra and Cascais, Portugal. Other studies have investigated this phenomenon with Markov chain models. [15]

used a first order Markov chain to make quantitative comparisons of the land use changes in the Niagara region, Canada, between 1935 and 1981. [16] presented a Markov-based model to study and predict forest cover in the Upper Midwest, USA. This approach could include important bias [17]. More recently, [1] used a first order Markov chain model to study land use and change analysis in the Zhujiang Delta, China.

In this article, we investigate if the LUCC, inside and outside PNSC are Markov Chains, i.e., if future land use and land cover is dependent of the present land use and land cover. The objective is to evaluate if the PNSC is an important factor in LUCC of Sintra and Cascais municipalities.

2. STUDY AREA

The study area analyzed in this research comprises the Sintra and Cascais municipalities in Portugal with an area of approximately 416 Km^2. These municipalities are integrated in the Lisbon Metropolitan Area (Figure 1).

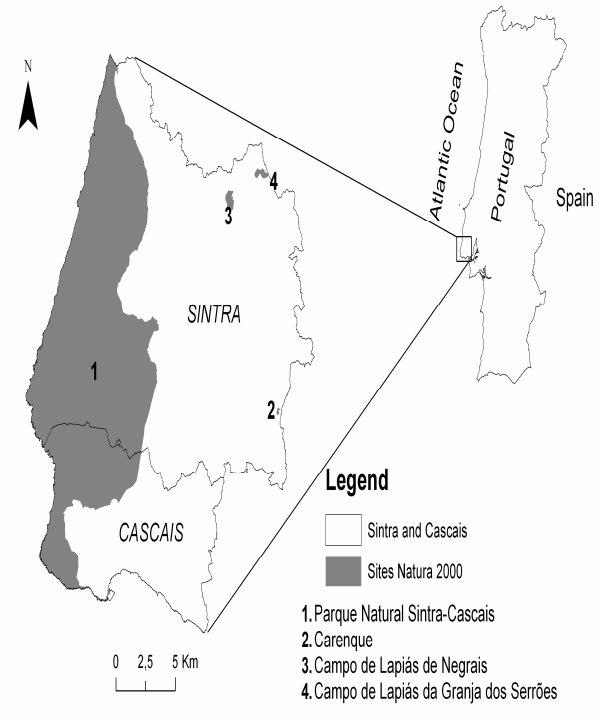

Figure 1. Sintra and Cascais Municipalities and Natura 2000 protected areas

There are four Natura 2000 protected areas in the study area. The largest one, the PNSC, has an area of 145 Km^2 and represents, approximately, 35% of total area. The vegetation of the PNSC is composed by Mediterranean and Western-Mediterranean species from which about 10% are endemic. In this last group, are included species like the *Armeria pseudarmeria*, *Dianthus cintranus cintyranus* and the *Omphalodes kusyn-skianae* which are considered threatened species at Community level. Endangered phauna species include *Rhinolophus hipposideros*, *Rhinolophus euryale*, *Putorius putorius*, *Bubo bubo*, *Hieraaetus fasciatus* and *Falco peregrinus* among many others [18].

It was in 1981 that the government created the Protected Landscape Area of Sintra-Cascais (*Área de Paisagem Protegida de Sintra-Cascais* by the *Decreto-lei* 292/81). The creation of this area had as objective to "preserve the natural, cultural and esthetical values inside its areas". All actions taken inside its areas were subject to strict authorizations from legal entities and included, among others, the introduction or change of new economic activities, urbanization, construction of roads and railroads, change the morphology of the terrain, destruction of natural vegetation, and introduction of new animal and plant species. In 1994, this area changed its name to PNSC (*Decreto-lei* 19/93) and the protected areas national network was created (*Decreto regulamentar* nº 8, 9/94).

According to the census [19], Sintra and Cascais municipalities have faced significant demographic growth between 1991 and 2001 (29%). The strong construction pressure in recent years may threaten PNSC functions.

3. DATA AND METHODS

3.1 Data

Two Landsat Thematic Mapper 5 (TM) and one Landsat 7 Enhanced Thematic Mapper

Plus (ETM+) images were used in this research (Table 1). The 1989 and 2000 images were downloaded from Global Land Cover Facility of the University of Maryland (USA). The 1994 TM image was specifically acquired for the purpose of this research.

Table 1. Satellite images used in this study

Characteristics / Sensor	Landsat TM	Landsat TM	Landsat ETM +
Date	14-03-1989	8-02-1994	24-06-2000
Path – Row	204-033	204-033	204-033
Spatial resolution	28.5 m	30 m	28.5 m

3.2 Methods

3.2.1 Image preprocessing

A 1:25,000 scale vectorial layer with administrative boundaries was used to create a subset of the images corresponding to the extent of Sintra and Cascais municipalities. The 1989 and 2000 images were previously geometrically and radiometrically corrected by USGS Earth Resource Observation Systems Data Center (EROS) to a quality level of 1G. The same quality level was available for the 1994 image by the European Space Agency. Both 1989 and 2000 images were already orthorectified to a UTM (Universal Transverse Mercator) projection using WGS (World Geodetic System) 84 datum. The 1994 image was co-registered to the 2000 image with a root mean square less than half a pixel (0.49). Both 1989 and 2000 images had a 28.5m pixel resolution. The image of 1994 was resampled to match this resolution using nearest-neighbor algorithm. This research is based on the detection of changes on surface reflectances of objects. This reason justifies the use of a relative radiometric correction with image regression [20] over 1989 and 1994 images. Brightness values of pixels of all the bands of 1989 and 1994 images were calibrated with image of year 2000 to create a linear regression equation. This procedure minimized effects caused by using time-series of satellite data collected in different dates and with different sun angles [20].

3.2.2 Classification écheme

This study aims to analyze the global trend of LUCC for the Sintra and Cascais municipalities. For this reason, the adopted land use/cover classification scheme included three generalized classes (Table 2).

Table 2. Classification scheme used in this study

Class	Description
Woodland	Coniferous, deciduous, and mixed forests.
Grassland	Grasses, scrubland, pastures, cropland, agriculture land, golf courses and other herbaceous vegetation.
Impervious	Areas with absence of vegetation cover. Houses, roads, dispersed warehouses, commercial and industrial buildings, airports, beaches and bare soil.

About 90 training samples were selected for each image. These training samples were as pure as possible and their location was maintained over the three images. Images were classified using the maximum-likelihood algorithm implemented in Clark Labs - Idrisi Kilimanjaro software [21]. Image classification was made over bands 2, 3, 4 and 5 because they were found to be the ones that best discriminated considered classes. Accuracy of classified maps was evaluated using 150 sample points systematically distributed. These points were converted into cells with the same resolution of the satellite images (28.5m) and classified as woodland, grassland and impervious. Selected pixels had to be pure instead of mixed pixels to ensure that the correct class was identified for each pixel [22]. These pixels were chosen using large-scale aerial photos and 1:25,000 scale land use/cover maps. Whenever it was not a pure pixel, the closest pure pixel was selected. Confusion matrices were used to compare ground information determined by the inspection of large-scale images and 1:25,000 scale land use/cover maps with the classification results.

3.2.3 Markov chains and LUCC

Stochastic processes generate sequences of random variables $\{X_n, n \in T\}$ by probabilistic laws. In this article, index n represents time. The process is considered discrete in time and T = {0, 5, 10 …} years approximately, which is a reasonable time unit for studying land use/cover change phenomenon. If the stochastic process is a Markov process then the sequence of random variables will be generated by the Markov property (1), formally:

$$P [X_{n+1} = a_{in+1} \mid X_0 = a_{i0}, ..., X_{in} = a_{in}] = P[X_{in+1} = a_{in+1} \mid X_{in} = a_{in}] \quad (1)$$

Where the double index means, in our study, for $n \in T$ and T = {0, 5, 10,…} and i the range of possible values that ai can assume, in this case the 3 classes defined previously. When the range of possible values for a_i is either finite or infinite denumerable, as in this study, the Markov process may be referred as a Markov chain. To demonstrate that land use/cover change in Sintra-Cascais area is a Markovian process, one must prove that: there is a statistical dependence between X_{n+1} and X_n (2); and that statistical dependence is a first-order Markov process (3).

$$P(X_n = a_n \mid X_{n-1} = a_{n-1}) \neq P (X_n = a_n) * P(X_{n-1} = a_{n-1}) \quad (2)$$

$$P [X_n = a_n \mid X_{n-1} = a_{n-1}] = P[X_n = a_n, X_{n-1} = a_{n-1}] / P [X_{n-1} = a_{n-1}] \quad (3)$$

A first-order Markov process is a Markov process where the transition from a class to any other does not require intermediate transitions to other states. The statistical dependence can be tested as in any contingency table [23] displaying the land use/cover change between X_n and X_{n-1}. In our study, this test is performed for the land use/cover change between 1994 and 2000. To infer from the association or independence between the land use/cover classes in different years from the contingency table, the random variable, with the chi-square distribution will be defined by (4):

$$\chi^2 = \sum_i \sum_j ((N_{ij} - M_{ij})^2 / M_{ij}) \quad (4)$$

Where N will be the contingency matrix displaying the land use/cover change between 1994 and 2000, and M the contingency matrix with the expected values of change assuming the independence hypotheses (Murteira, 1990). χ^2 measures the distance between the observed values of land use/cover change and the expected ones assuming independence and must be high enough to prove (2), for 4 degrees of freedom. The same non-parametric test will be used to test the Markov property. In this case, the values to be compared with the observed ones will be calculated from the Chapman-Kolmogorov equation (5) [24], assuming that these variables are generated by a first-order Markov process:

$$P(X_n = a_n \mid X_m = a_m) = P(X_1 = a_1 \mid X_m = a_m).P(X_n = a_n \mid X_1 = a_1), \quad m \leq l \leq n \quad (5)$$

As far as concerned in this study, the Chapman-Kolmogorov equation states that transition probabilities from years 1989 to 2000 can be calculated by multiplying the transition probabilities matrix from years 1989 to 1994 by the transition probabilities matrix from years 1994 to 2000 (6).

$$\chi^2 = \sum_i \sum_j ((N_{ij} - O_{ij})^2 / O_{ij}) \quad (6)$$

As the name itself indicates the transition probabilities matrix will be estimated by the contingency matrix displaying the relative frequencies of LUCC in a certain period of time.

4. RESULTS

4.1 Image classification and accuracy assessment

Three land use/cover maps were produced, respectively, for years 1989, 1994 and 2000 using the maximum-likelihood algorithm (Figures 2-4).

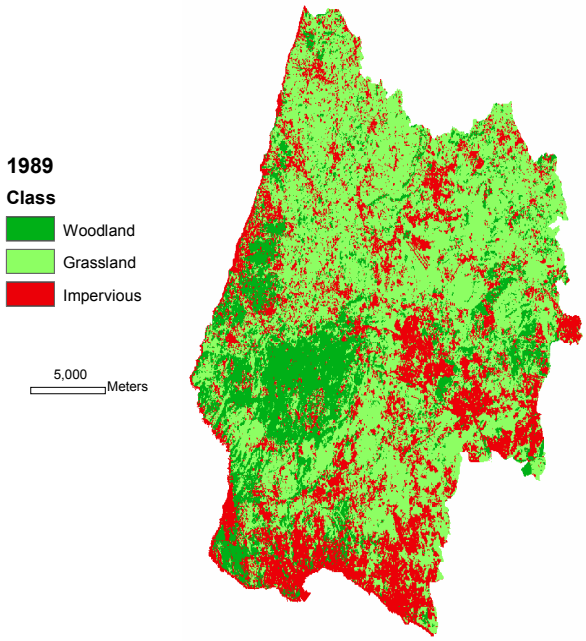

Figure 2. Land use/cover map for year 1989

The 2000 ETM+ image accuracy was assessed using large-scale orthophotos and land cover maps. There was no ground truth data available for this study to assess 1994 image accuracy. However, the classification methodology used for image of year 1989 was replicated for classifying 1994 image. Both images were collected in the rainy season at the same time of the day using the same satellite sensor. For these reasons, we assumed that the overall accuracy of 1994 classification should be identical to the 1989 classification. Overall accuracies obtained for 1989 and 2000 images were, respectively, 88.8% and 90.7%. The Kappa indices for years 1989 and 2000 were, respectively, 85.3% and 87.1%. These values are considered above the minimum value (85%) stipulated for interpretation accuracy in the identification of land use and land cover categories from remotely sensed data [25].

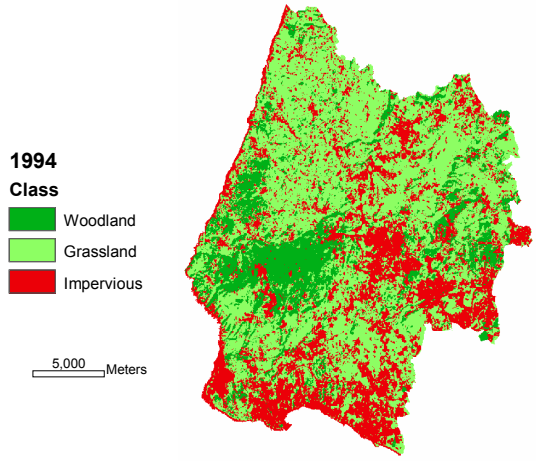

Figure 3. Land use/cover map for year 1994

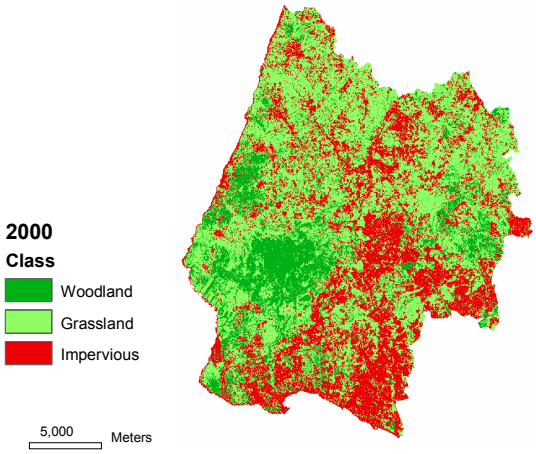

Figure 4. Land use/cover map for year 2000

4.2 Hypothesis testing

As described in the methodology, the main hypothesis to be tested in this study is that LUCC in the study area is generated by a first order Markov process. This will be our H_0. To prove H_0 two subsidiary hypotheses must be verified: H_1 - land use/cover in different time periods is not statistically independent and H_2- LUCC in the study area is a Markov process. For the purpose of the analysis inside and outside PNSC, six contingency tables were used to quantify land cover changes between years 1989 and 1994, 1989 and 2000 and 1994 and 2000 for each area (Tables 3-8).

Tables 3-5. Contingency tables inside PNSC (W: Woodland; G: Grassland; I: Impervious; T: Total)

		W	G	I	T	Δ Km²	Δ %
1994	W	33.8	7.0	2.9	43.7	-7.0	-13.9
	G	10.4	49.0	11.7	71.1	7.5	11.7
	I	6.6	7.6	14.9	29.1	-0.4	-1.5
	T	50.7	63.7	29.5	143.9	46.3	32.2

1994

		W	G	I	T	Δ Km²	Δ %
2000	W	29.4	9.6	4.1	43.1	-0.6	-1.4
	G	13.1	48.6	12.6	74.3	3.2	4.5
	I	1.2	12.9	12.4	26.5	-2.6	-8.9
	T	43.7	71.1	29.1	143.9	53.5	37.2

1989

		W	G	I	T	Δ Km²	Δ %
2000	W	33.8	6.7	2.6	43.1	-7.7	-15.1
	G	15.0	46.4	12.9	74.3	10.7	16.8
	I	1.9	10.5	14.1	26.5	-3.0	-10.3
	T	50.7	63.7	29.5	143.9	49.6	34.5

Tables 6-8. Contingency tables outside PNSC (W: Woodland; G: Grassland; I: Impervious; T: Total)
Tablas 6-8. Tablas de contingencia fuera del PNSC

1989

		W	G	I	T	Δ Km²	Δ %
1994	W	13.9	10.4	2.5	26.8	-1.8	-6.2
	G	8.4	118.8	21.4	148.6	-11.7	-7.3
	I	6.2	31.2	59.5	96.9	13.5	16.2
	T	28.5	160.4	83.4	272.2	80.0	29.4

1994

		W	G	I	T	Δ Km²	Δ %
2000	W	10.4	6.3	3.7	20.4	-6.4	-23.7
	G	13.7	98.5	27.6	139.7	-8.9	-6.0
	I	2.7	43.8	65.6	112.1	15.2	15.7
	T	26.8	148.6	96.9	272.2	97.8	35.9

1989

		W	G	I	T	Δ Km²	Δ %
2000	W	12.0	6.4	2.0	20.4	-8.1	-28.5
	G	12.4	106.7	20.7	139.7	-20.6	-12.9
	I	4.1	47.3	60.7	112.1	28.7	34.5
	T	28.5	160.4	83.4	272.2	92.8	34.1

The χ^2 value obtained to measure the association between the contingency table 1989-2000 inside PNSC (Table 5) and the Chapman-Kolmogrov equation is 0.605039. This value is clearly below the critical value of the distribution for a significance level of 0.950 which is 0.710721. This result allows the assumption that LUCC is a Markovian process inside the natural park. For the remaining area, where the LUCC has been more dynamic, the chi-square calculated to measure the association between the Chapman-Kolmogrov matrix and the contingency table 1989-2000 is now clearly above the critical value of the distribution for 0.950 confidence level with a value of 1.286519.

4.3 LUCC estimates for year 2006

To use Markov chains to predict future land use/cover inside de PNSC, one must prove that the Markov chain is stationary. However, this property can only be defined for recurrent Markov chains [24]. A Markov chain is said to be recurrent if it is certain that the chain will return to the same state, but uncertain when that will happen (aperiodic). We have no reason to assume that LUCC is a recurrent aperiodic chain; therefore we must ignore the stationary equation. On the other hand, it is reasonable to assume that LUCC inside PNSC will continue to be this Markov chain in the near future because all the factors affecting this process will continue to be regulated by the park administration. Considering all these, future land use/cover quantities for year 2006 were estimated for the area inside PNSC (Table 9).

This prediction is not spatial because Markov chains assume spatial independence of the area units. LUCC inside the PNSC are not predicted to be significant between 2000 and 2006. This conclusion reinforces the importance of the PNSC in maintaining LUCC dynamics stable inside its area.

Table 9. Contingency table 2000-2006 inside PNSC

		2000					
		W	G	I	T	Δ Km^2	Δ %
2006	W	29.0	12.9	1.2	43.1	0.4	0.8
	G	10.0	50.8	13.5	74.3	-0.9	-1.2
	I	3.7	11.5	11.3	26.5	0.5	2.1
	T	42.7	75.2	26.0	143.9	52.8	36.7

Estimates for what happens outside the PNSC are not presented because there was no significant statistical evidence to state that it was a Markov chain. This fact also means that LUCC outside the PNSC is not dependent of current LUCC. It may follow any other probabilistic law but not a Markovian one.

5. CONCLUSIONS

This paper describes an integrated approach of remote sensing and stochastic modeling techniques in explaining LUCC in Sintra-Cascais area. It was found that the behavior of LUCC inside the PNSC was a Markov process between years 1989 and 2000. The land use/cover dynamics of the area outside the park did not follow a Markovian behavior. The transition mechanism of LUCC outside the park is very unstable for the defined land use/cover scheme. This means that it does not depend on the previous land use/cover. These findings reinforce the existence of the PNSC as an important factor in the stability of this highly dynamic area.

Although Markov chains constitute a good tool for describing and projecting LUCC quantities, they are insufficient for spatial explicit LUCC predictions, because they assume statistical independence of spatial units. However, LUCC modelers can use Markov transitions coupled with spatially explicit models like cellular automata and/or linear extrapolation models. The methodology here presented can be employed to investigate if it is correct or not to use Markov transition probabilities in their modeling processes. Future research includes the experimentation of spatially explicit models to better understand the LUCC dynamics of this area.

REFERENCES

[1] WENG, Q. Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modelling. Journal of Environmental Management. Vol. 64. 2002. pp. 273-284.

[2] GLUCH, R, Urban growth detection using texture analysis on merged Landsat TM and SPOT-P data. Photogrammetric Engineering & Remote Sensing. Vol. 68. 2002. pp. 1283-1288.

[3] KWARTENG, A.; P. CHAVEZ. Change detection study of Kuwait city and environs using multi-temporal Landsat Thematic Mapper data. International Journal of Remote Sensing, Vol. 19. 1998. pp. 1651-1662.

[4] MASEK, J.; F. LINDSAY; S. GOWARD. Dynamics of urban growth in the Washington DC metropolitan area, 1973-1996, from Landsat observations. International Journal of Remote Sensing. Vol. 21. 2000. pp. 3473-3486.

[5] ZHANG Q.; WANG J.; PENG X.; GONG P.; SHI P. Urban built-up land change detection with road density and spectral information from multi-temporal Landsat TM data. International Journal of Remote Sensing. Vol. 23. 2002. pp. 3057-3078.

[6] STEFANOV, W.; RAMSEY M.; CHRISTENSEN P. Monitoring urban land cover change: An expert system approach to land cover classification of semiarid to arid urban centers. Remote Sensing of Environment. Vol. 77. 2001. pp. 173-185.

[7] MUCHER, C.; STEINNOCHER K.; KRESSLER F. Land cover characterization and change detection for environmental planning of Pan-Europe. International Journal of Remote Sensing, 2000. Vol. 21. pp. 1159-1181.

[8] CABRAL, P.; GILG J.-P.; PAINHO M. Monitoring urban growth using remote sensing, GIS and spatial metrics. Proceedings of SPIE Optics & Photonics: Remote sensing and modeling of ecosystems for sustainability. 2005. San Diego, USA. pp. 1-9.

[9] BAKER, W. A review of models of landscape change. Landscape Ecology. Vol. 2. 1989. pp. 111-133.

[10] CHENG, J. Modeling Spatial & Temporal Urban Growth. 2003. PhD thesis Utrecht University. Faculty of Geographical Sciences.

[11] BENENSON, I.; TORRENS P.,Geosimulation: Automata-based modeling of urban phenomena. Chichester: John Wiley & Sons, Ltd. 2004.

[12] WHITE, R.; ENGELEN G. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. Computers, environment and urban systems, Vol. 24. 2000. pp. 383-400.

[13] BATTY, M. Cities and complexity: Understanding cities with cellular automata, agent-based models and fractals. Cambridge, London: MIT Press. 2005.

[14] CABRAL, P.; ZAMYATIN A. Three land change models for urban dynamics analysis in the Sintra-Cascais area. Proceedings of First Workshop of the EARSEL SIG on Urban Remote Sensing: Challenges and solutions. 2006. Berlin.

[15] MULLER, M.; MIDDLETON J. A. Markov model of land-use change dynamics in the Niagara Region, Ontario, Canada. Landscape Ecology. Vol. 9. 1994. pp. 151-157.

[16] BROWN, D.; PIJANOWSKI B.; DUH J. Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. Journal of Environmental Management. Vol. 59. 2000. pp. 247-263.

[17] HUBER, W. Estimating Markov transitions. Journal of Environmental Management. Vol. 61. 2001. pp. 381-385.

[18] ICN. Instituto da Conservação da Natureza e Biodiversidade. [cited 2004 November 2004]. Available from: URL: http://www.icnb.pt. 2004.

[19] INE, Instituto Nacional de Estatística. Lisboa. 2003.

[20] JENSEN, J.R. Introductory Digital Image Processing: A remote sensing perspective. Second edition, edited by K.C. Clarke. New Jersey: Prentice Hall. 1996.

[21] IDRISI, Idrisi Kilimanjaro. Clark Labs: Worcester, MA. 2004.

[22] GONG, P.; HOWARTH J. The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. Photogrammetric Engineering & Remote Sensing. Vol 56. 1990. pp. 67-73.

[23] MURTEIRA, B. Probabilidades e estatística. Vol. II, 2ª Edição, McGraw-Hill de Portugal. 1990.

[24] KIJIMA, M. Markov processes for stochastic modeling. Stochastic Modeling Series, London: Chapman & Hall. 1997.

[25] ANDERSON, J.R.; HARDY E.; ROACH J.; WITMER R. A Land Use And Land Cover Classification System For Use With Remote Sensor Data. U.S. Geological Survey: Washington, DC. 1976.