Assessing yearly transition probability matrix for land use / land cover dynamics

Jean-François Mas [1], Ernesto Vega [2]

[1] Centro de Investigaciones en Geografía Ambiental
Universidad Nacional Autónoma de México (UNAM)
jfmas@ciga.unam.mx

[2] Centro de Investigaciones en Ecosistemas
Universidad Nacional Autónoma de México (UNAM)
evega@oikos.unam.mx

Abstract

In order to generate land cover projections and model land use and cover changes (LUCC), probability-based transition matrices are obtained through the overlaying of two maps of two different dates. However, the observation interval may differ because maps are not available every year or at a constant time interval and a matricial algorithm is commonly used to adjust the matrix. However, although the obtained yearly matrix is mathematically correct, it does not necessary represent adequately the yearly transitions due to the spatial coincidence of various transitions over the entire period of time and because some transitions that are observed over a large period are not possible over a shorter (e.g. yearly) period. In this paper, a novel approach, based on a genetic algorithm (GA), is applied to adjust yearly transition matrices taking into account criteria to produce more realistic transitions. A LUCC model was used to produce land cover maps at different dates using transition rules defined by the user. Yearly matrices were then obtained from these maps by both matricial and genetic algorithms and were compared. For certain periods of time, the matricial algorithm was unable to find a yearly matrix or found a matrix with impossible transitions. The GA approach was able to find realistic matrices from all the periods.

Keywords: Markov matrix, modelling, land cover change.

1. Introduction

Land use and cover changes (LUCC) become a critical issue due to its relation with climate change and biodiversity loss. The development and application of models that improve our capability to predict trends in LUCC is nowadays an essential requirement of planning process (Lambin et al., 2001; Pontius Jr. and Schneider, 2001). LUCC monitoring is often carried out through the comparison over time of classified remotely-sensed images. Probability-based transition tables, such as Markovian models, are obtained from area-based transition through the overlaying of two land use/cover (LUC) maps or classified images of two different dates in a GIS framework. These matrices are used to generate land cover projections under the assumption that the transition rule is invariant over time. The area

vector of LUC categories after c years can be obtained by the product of the area vector in the current year and the transition matrix expressing the transition rule. However, a common problem is that observation interval may differ among several observation periods because remotely-sensed images or LUC maps are not available every year or at a constant time interval. If the observation intervals differ, the transition probabilities cannot be compared directly. Therefore, they should be adjusted on the basis of the same observation time to allow the comparison. Moreover, in LUCC modeling, it is often desirable to use a yearly transition matrix in order to set the modeling step in one year.

Equation (1) indicates a formula (hereafter the matricial algorithm) commonly used to calculate the yearly matrix (Flamenco et al., 2007; Mas et al., 2004, Soares-Filho et al., 2002).

$$A^{1/c} = V.H \, V^{-1}. \tag{1}$$

Where $A^{1/c}$ is the yearly matrix corresponding to A (Note that $A^{1/c}$ multiplied by itself c times is equal to the original matrix A), V is the eigenvector of matrix A (n-by-one column vector), V^{-1} the inverse of V and, H is a n-by-n matrix in which the diagonal elements are filled by the values of the eigenvalues of the matrix A and out-diagonal elements are set to 0.

Takada et al. (2010) developed a computer program to solve the equation (1) and reported that there are c^n solutions, including matrices that have imaginary numbers, negative values or values above one, which is not acceptable for a probability matrix. These authors proposed to use the obtained matrices with negative elements close to zero setting these negative values to zero. In some cases, there are no matrices with small negative values and no yearly matrix can be provided.

However, although these yearly matrices are mathematically correct, they do not necessary represent adequately the yearly transitions because: i) The changes observed by comparison of the beginning and the end of the period are only a part of the changes which occurred. For instance, a change from category 1 to 2 followed by a change from 2 to 1 is not detected at all and a change from category 1 to 2 followed by a change from 2 to 3 seems to be a single direct change from 1 to 3, ii) the change patterns may vary along the period of observation: the matrix A is not the result of the product of only one yearly matrix but rather of different yearly matrices and, iii) some transitions that are possible over a large period are not possible over a shorter (yearly) period (for example the transition from pasture to mature forest must transitate by secondary forest).

In this paper a novel approach, based upon a genetic algorithm (GA), is applied to adjust yearly transition matrices taking into account criteria to produce more realistic transition probabilities.

2. Material

DINAMICA EGO, an environmental modeling platform for dynamic spatial models, was used to develop a LUCC model (Soares-Filho et al., 2002, Rodrigues et al., 2007). We used a R based GA called genalg developed by Willighagen (2005) to optimize, a set of floating point numbers (chromosomes) representing transition probabilities. Matricial computing was also carried out with R (R Development Core Team, 2011) and a DOS program developed by Takada et al. (2010).

3. Methods

3.1. Genetic algorithms (GAs)

A GA is a search heuristic that mimics the biological evolution process for solving problems in a wide domain. A GA operates through a simple cycle of stages: i) Creation of a "population" of strings called chromosomes, which encode candidate solutions, ii) evaluation of each string, iii) election of best strings and, iv) genetic manipulation to create new population of strings.

Each cycle in GAs produces a new generation of possible solutions for a given problem. In the first phase, an initial population, describing representatives of the potential solution, is created to initiate the search process. The elements of the population are encoded into bit-strings, called chromosomes. The performance of the strings, often called fitness, is then evaluated with the help of an evaluation function, representing the constraints of the problem. Depending on the fitness of the chromosomes, they are selected for a subsequent genetic manipulation process. It should be noted that the selection process is mainly responsible for assuring survival of the best-fit individuals. After selection of the population strings, the genetic manipulation process consisting of two steps is carried out. In the first step, the crossover operation that recombines the bits (genes) of each two selected strings (chromosomes) is executed. The crossover points of any two chromosomes are selected randomly. The second step in the genetic manipulation process is called mutation, where the bits at one or more randomly selected positions of the chromosomes are altered. The mutation process helps to overcome trapping at local maxima. The offsprings produced by the genetic manipulation process are the next population to be evaluated.

In this study, a chromosome represents the values of the probabilities of a candidate yearly matrix. The evaluation function was defined to select the "best" chromosome computing the power of the candidate matrix (chromosome) and comparing its values with the corresponding values in the entire period matrix computing the sum of the absolutes differences between the elements of the two matrices. An additional criterion penalized matrix which columns sum more than one.

3.2. Case study

The GA was applied to a simple case of LUCC to illustrate and compare the outputs with the results of the matricial algorithm. Suppose a study area with three categories: 1) mature forest, 2) secondary forest and, 3) pasture. The annual transition rules expressed by this matrix (Table 1) indicate that no direct transition can be done from mature forest to secondary forest and from pasture to mature forest (null transition probability for these two transitions).

Table 1: Annual probability transition matrix.

t \ t$_{+1}$	Mature forest	Secondary forest	Pasture land
Mature forest	0.8	**0**	0.2
Secondary forest	0.2	0.7	0.1
Pasture land	**0**	0.1	0.9

These annual LUCC were simulated over a period of 10 years. Using a set of change potential maps, modeled changes were more likely located in a certain region of the study area. Consequently, as in real situations, different transitions occured at the same place during the entire period of simulation. Simulated LUC maps for time t_0, t_1, t_7 and t_{10} were compared in order to produce transition matrices for different periods of time. These matrices were then transformed into yearly matrices using both matricial and genetic algorithms.

4. Results

Tables 2, 3 and 4 show the matrices of transition probability for increasing periods of time. In these matrices, transitions from mature forest to secondary forest and from pasture to secondary forest exhibit an increasing probability because these indirect transitions are more likely to occur over a larger period of time.

Table 2: Transition probability matrix for period t_0-t_5 (5 years).

t_0 \ t_5	Mature forest	Secondary forest	Pasture land
Mature forest	0.35	0.10	0.55
Secondary forest	0.37	0.25	0.38
Pasture land	0.00	0.28	0.72

Table 3: Transition probability matrix for period t_0-t_7 (7 years).

t_0 \ t_7	Mature forest	Secondary forest	Pasture land
Mature forest	0.28	0.15	0.57
Secondary forest	0.26	0.17	0.57
Pasture land	0.00	0.14	0.86

Table 4: Transition probability matrix for period t_0-t_{10} (10 years).

t_0 \ t_{10}	Mature forest	Secondary forest	Pasture land
Mature forest	0.22	0.19	0.59
Secondary forest	0.17	0.15	0.67
Pasture land	0.25	0.2	0.56

In order to obtain the yearly matrix, we first applied the matricial method. Slights differences in probabilities were found depending the program used to carry out the matricial method (DINAMICA, R and Takada's program), we present here the matrices obtained by the program developed by Takada et al. (2010). In the case of the matrix over 5 years, a solution with small negative elements was found (Table 5) and is very close to the original matrix (see table 1). In the case of the 7 years matrix, a yearly matrix was also found (Table 6) but it presents probability above zero for an impossible direct transition (Mature to secondary forest). Finally, no corresponding yearly matrix with positive elements nor matrix with a few small negative elements was found for the t_0-t_{10} period.

Table 5: Annual transition probability matrix obtained from matrix of period t_0-t_5.

t_0 \ t_{10}	Mature forest	Secondary forest	Pasture land
Mature forest	0.82	0	0.18
Secondary forest	0.22	0.72	0.07
Pasture land	0	0.11	0.89

Table 6: Annual transition probability matrix obtained from matrix of period t_0-t_7.

t \ t_{+1}	Mature forest	Secondary forest	Pasture land
Mature forest	0.79	**0.08**	0.13
Secondary forest	0.24	0.62	0.13
Pasture land	0.00	0.06	0.94

For the 5 and 7 years periods, the GA method gave matrices with very similar values of probability (Table 7). For the period t_0-t_{10}, the GA found a matrix presenting unrealistic values (higher probabilities for change transition than for permanence, table 8). The fitness function was modified to favor matrices with permanence probability higher than 0.7 and a more realistic matrix was found (Table 9).

Table 7: Annual probability matrix obtained by the GA from periods t_0-t_5 and t_0-t_7.

t \ t_{+1}	Mature forest	Secondary forest	Pasture land
Mature forest	0.89	0	0.11
Secondary forest	0.12	0.85	0.03
Pasture land	0	0.05	0.95

Table 8: Annual transition probability matrix obtained by the GA from period t_0-t_{10}.

t \ t_{+1}	Mature forest	Secondary forest	Pasture land
Mature forest	0.49	0	0.51
Secondary forest	0.61	0.11	0.28
Pasture land	0	0.28	0.72

Table 9: Annual transition probability matrix obtained from matrix of period t_0-t_{10} by the GA with an additional constraint (permanence probabilities preferentially above 0.7).

t \ t_{+1}	Mature forest	Secondary forest	Pasture land
Mature forest	0.73	0	0.27
Secondary forest	0.33	0.66	0.00
Pasture land	0	0.11	0.89

4. Discussion and conclusions

A method based on GA was proposed to determine yearly matrix from a multi-years matrix. It allows overcoming the problem of not finding solution encountered with the matricial algorithm. Another advantage is that the user can assign defined

values or ranges of allowed values to certain transitions to obtain a more realistic matrix. This can be essential if the model aims at simulating the landscape dynamics precisely (for instance a model of shifting agriculture and successional vegetation dynamics). Models which objective is only the elaboration of a final map of landscape can likely be based upon matrices with indirect transitions.

In the present study case, the LUC maps were elaborated allowing multiple transitions at the same place. However, the fitness value was based on matrix multiplication (that is, without taking into account the spatial patterns of change). To take into account the spatial coincidence of change, fitness could be calculated from the simulated map obtained by the model with a candidate matrix (chromosome). This approach is much more time consuming because the entire LUCC model has to be run for each candidate chromosome, but it is expected to give better results particularly in case where changes tend to occur at the same place.

Acknowledgments

References

Flamenco-Sandoval, A., Ramos, M.M., Masera, O.R. (2007), "Assessing implications of land-use and land-cover change dynamics for conservation of a highly diverse tropical rain forest". *Biology Conservation*, Vol. 138:131-145.

Lambin, E.F.; Turner, B.L.; Geist, H. J.; Agbola, S.B. et al. (2001), "The causes of land- use and land-cover change: moving beyond the myths". *Global Environmental Change*, Vol. 11: 261-269.

Mas, J.F., Velázquez, A. Díaz-Gallegos, J.R., Mayorga-Saucedo, R., Alcántara, C., Bocco, G., Castro, R., Fernández, T., Pérez-Vega, A. (2004), "Assessing land use/cover changes: a nationwide multidate spatial database for Mexico". *International Journal of Applied Earth Observation and Geoinformation*, Vol. 5(4):249-261.

Pontius Jr., R.G.; Schneider, L.C. (2001), "Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA". *Agriculture, Ecosystems and Environment*, Vol. 85:239-248.

R Development Core Team (2011), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rodrigues, H.O., Soares-Filho, B.S, Costa, W.L.S. (2007), "Dinamica EGO, uma plataforma para modelagem de sistemas ambientais". In: INPE, *Simpósio Brasileiro de Sensoriamento Remota*, Florianópolis, Brasil, pp. 3089-3096.

Soares-Filho, B.S., Pennachin, C.L., Cerqueira (2002), "G. DINAMICA – a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier". *Ecological Modelling*, Vol. 154(3): 217:235.

Takada, T., Miyamoto, A. Hasegawa, S.F. (2010), "Derivation of a yearly transition probability matrix for land-use dynamics and its applications". *Landscape Ecology*, Vol. 259(4): 561-572.

Willighagen, E.L., 2005, Genalg, http://cran.r-project.org/web/packages/genalg/