

## CALIBRATING CELLULAR AUTOMATA OF LAND USE/COVER CHANGE MODELS USING A GENETIC ALGORITHM

J.F. Mas<sup>a\*</sup>, B. Soares-Filho<sup>b</sup>, H. Rodrigues<sup>b</sup>

<sup>a</sup> Centro de Investigaciones en Geografía Ambiental, Universidad Nacional Autónoma de México, 58190 Morelia, Mexico - jfmas@ciga.unam.mx

<sup>b</sup> Centro de Sensoriamento Remoto, Universidade Federal de Minas Gerais, Belo Horizonte 31270-900, MG, Brasil - britaldo@csr.ufmg.br

### Commission II, WG II/4

**KEY WORDS:** Stochastic spatial simulation, Genetic algorithm, Amazon deforestation, landscape pattern, fragmentation, connectivity

### ABSTRACT:

Spatially explicit land use / land cover (LUCC) models aim at simulating the patterns of change on the landscape. In order to simulate landscape structure, the simulation procedures of most computational LUCC models use a cellular automata to replicate the land use / cover patches. Generally, model evaluation is based on assessing the location of the simulated changes in comparison to the true locations but landscapes metrics can also be used to assess landscape structure. As model complexity increases, the need to improve calibration and assessment techniques also increases. In this study, we applied a genetic algorithm tool to optimize cellular automata's parameters to simulate deforestation in a region of the Brazilian Amazon. We found that the genetic algorithm was able to calibrate the model to simulate more realistic landscape in term of connectivity. Results show also that more realistic simulated landscapes are often obtained at the expense of the location coincidence. However, when considering processes such as the fragmentation impacts on biodiversity, the simulation of more realistic landscape structure should be preferred to spatial coincidence performance.

### 1. INTRODUCTION

Spatially explicit land use / land cover change (LUCC) models aim at simulating the patterns of change on the landscape (Paegelow et al., 2013). Many of the models are based on a inductive pattern-based approach: In this approach, LUCC is modelled empirically using past LUCC spatial distribution and rate to develop a mathematical model that estimates the change potential as a function of a set of explanatory spatial variables and the expected amount of change (Paegelow and Olmedo, 2005; Mas et al., 2014). In prospective modelling, allocation procedures are used to simulate the projected amount of change in the most likely locations. In order to simulate landscape structure and fragmentation pattern, the simulation procedures of most computational LUCC models use a cellular automata (CA) that intends to replicate the land use/cover (LUC) patches.

Generally, the assessment of model performance is based on the spatial coincidence between a simulated map and an observed LUC map for the same date and does not evaluate the model ability to simulate the landscape pattern as, for instance, the size, the shape and the distribution of patches. Landscapes metrics can be used to assess simulated landscape structure (Mas et al., 2012). As model complexity increases, the need to improve calibration techniques also increases. This study aims at applying a genetic algorithm to optimize cellular automata's parameters to simulate deforestation in a region of the Brazilian Amazon.

### 2. MATERIALS

Dinamica EGO freeware (hereafter DINAMICA) is an environmental modeling platform for the design of space-time models. It has been applied to a variety of studies, such as modeling tropical deforestation (Soares-Filho et al., 2001, 2002, 2006; Cuevas and Mas, 2008), urban growth (Almeida et al., 2005), fire regimes

(Silvestrini et al., 2011) and, landscape patterns (Pe'er et al., 2013; Soares Filho et al., 2003), among others. We chose it due to its flexibility and computing efficiency (Mas et al., 2014).

We used a portion of one of the 12 case-study areas from Soares-Filho et al. (2013), comprising a TM-Landsat scene map from the PRODES project (INPE, 2011). The study area is located in the State of Pará along the road between Santarém and Cuiabá. Deforestation maps encompassing the years 1997 and 2001 were rasterized into a 250-m raster. As spatial drivers of deforestation, we selected only three variables from the dataset: distance to previously deforested lands, proximity to roads and, elevation in order to have a simple model easier to interpret.

### 3. METHODS

DINAMICA uses transition probability maps that are based on a Bayesian method of conditional probability known the weight of evidence method. These maps of probability are used to simulate landscape dynamics using both Markov matrices to project the quantity of change and a cellular automata (CA) approach to reproduce spatial patterns (Soares-Filho et al., 2002, 2010). Two complementary CA are available: the Expander, that simulates the expansion of previously formed patches and, the Patcher that generates new patches through a seeding mechanism. The behaviour of the CA is controlled by four main parameters: the mean patch size, the patch size variance, the isometry and the prune factor. Increasing patch size value leads to simulated maps with a less fragmented landscape; increasing the patch size variance leads to a more diverse landscape in term of size of the patches. Setting the isometry value greater than one leads to create more isometric patches. Increasing the prune factor allows simulated changes to occur in less likely areas. With a prune factor of one, the model becomes almost deterministic, that is changes are restricted to the areas with higher change probability (Soares-Filho et al., 2002; Mas et al., 2014).

DINAMICA has also a genetic algorithm tool which has been

\*Corresponding author

used to calibrate the weights of evidence (Soares-Filho et al., 2013). Genetic algorithms are based on the Darwinian mechanisms of evolution and attempt to mimic the natural evolution of a population by selection, combination, and mutations (random changes in genes). Each chromosome is composed of genes that define its characteristics. In computational models, a chromosome is a string of numbers encoding the parameters that the genetic algorithm attempts to optimize. First, a population of chromosomes is created randomly. Subsequently, the genetic algorithm generates new individuals from the existing ones by means of processes of selection, crossover and mutation. The parents are selected from the best chromosomes, according to a fitness criterion. To increase genetic heterogeneity, mutation and crossover operators randomly changes or interchange some genes on a chromosome sequence. A new generation is created by copying the most successful individuals, by crossing-over them and by mutating some chromosome sequences. These processes generate a new population of chromosomes with a greater chance of including a near-optimum solution for the problem. This evolution process iterates until the fitness-stopping criterion is satisfied. Genetic algorithms have been shown to be able to optimize multi-parameter function Wang (1997). In the present study, the genetic algorithm optimizes the four main parameters of the Parcher CA: the mean patch size, the patch size variance, the isometry and, the prune factor.

Forest Maps of 1997 and 2001 were overlaid in order to map deforested and conserved forest area. This deforestation map allowed to compute a matrix of Markov, the weights of evidence and a map of probability using the three explanatory variables (see appendix). Then deforestation was simulated from 1997 to 2001 using the 1997 as initial LUC map, the matrix of Markov to calculate expected annual deforested area and the set of weights of evidence to compute the probability of change. The variable distance from previously deforested area is a dynamic variable, that is computed at each annual iteration of the simulation. It is worth noting that, in the present case, the training (or calibration) period is the same than the simulation period because the objective of the study is fitting the model and not testing its prospective ability.

The CA calibration aimed at fitting the parameters in order to simulate a landscape similar to true landscape with regards to the size of the patches of deforestation and its general spatial distribution (e.g. avoiding that simulated changes concentrate in most likely areas, near previous deforested area when in true landscape there are also a little quantity of changes in remote areas). We calculated three indices which depict the landscape characteristics: The mean area of the deforestation patches (DPMA), the standard deviation of the deforestation patches area (DPASD) and the mean distance to deforested patches in remote areas ( $MDFP_r$ ) which is the mean distance of forest cells to deforested area taking into account only cells located at a larger distance than the average distance. Fitness was assessed through the difference of the indices for the true change and the simulated change maps. The fitness criterion was computed by the weighted sum of three fitness components (equation 1).

$$F = \frac{w_1}{\delta DPMA} + \frac{w_2}{\delta DPASD} + \frac{w_3}{\delta MDFFP_r} \quad (1)$$

where  $w_1, w_2, w_3$  = pondering weights,  
 $\delta DPMA$  = Absolute value of the difference of index DPMA between simulated and true map,  
 $\delta DPASD$  = Absolute value of the difference of index DPASD between simulated and true map,  
 $\delta MDFFP_r$  = Absolute value of the difference of index  $MDFP_r$  between simulated and true map.

Final simulated map was assessed by visual inspection and through the computing of the spatial coincidence between true change and simulated changes during 1997-2001.

#### 4. RESULTS

Figures 1, 2 and 3 present the forest in 1997, 2001 and the deforested areas during the period 1997-2001 respectively. Figure 4 shows the map of probability of deforestation according to the weights of evidence based on the three explanatory variables (maps of explanatory variables are in appendix). It can be observed that change presents a clear spatial pattern (patches) and occurred mainly on high probability areas but also in less likely areas (e.g. remote areas).

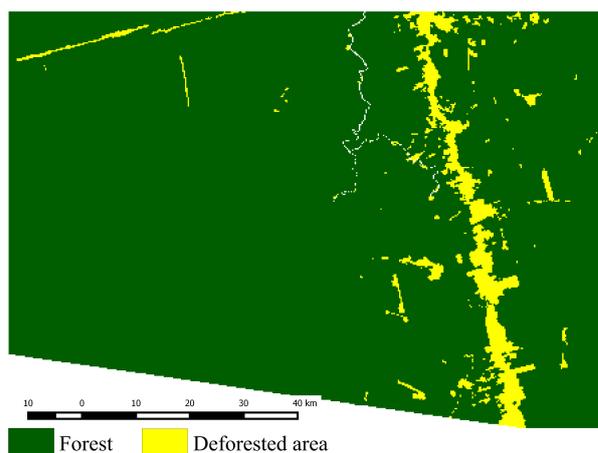


Figure 1: Land use / cover map (1997)

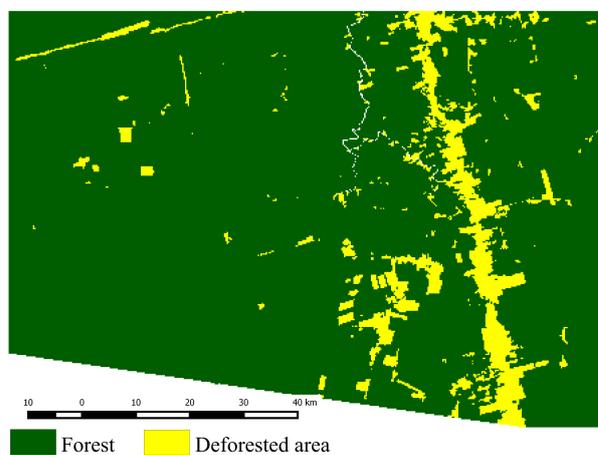


Figure 2: Land use / cover map (2001)

Figure 5 represents the simulated LUCC map from the modelled which CA was calibrated by the genetic algorithm. As comparison, figure 6 shows the simulated LUCC map thresholding the cells with the highest probability to simulate the change. It can be observed that the map obtained using CA is more realistic in term of landscape structure because it presents patches of deforestation of broadly the same size and distribution than the true map of change (Figure 3). However, the unrealistic map obtained without CA has a higher coincidence with true map (29%) than the more realistic map obtained with the CA calibrated by the genetic algorithm (22%).

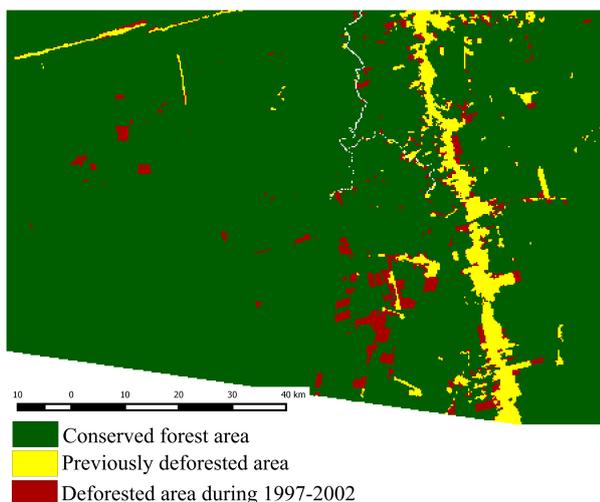


Figure 3: LUC map (1997-2002). Previously deforested area refers to forest area cleared before 1997. Some deforestation patches are located in remote areas

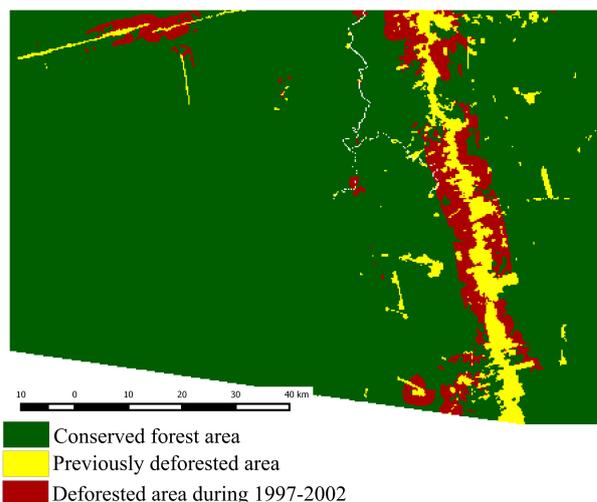


Figure 6: Simulated LUC map (1997-2002) applying a threshold to the probabilities. Previously deforested area refers to forest area cleared before 1997

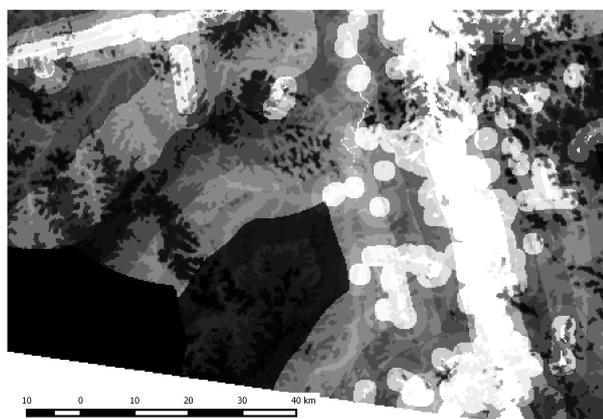


Figure 4: Map of probability of deforestation (1997)

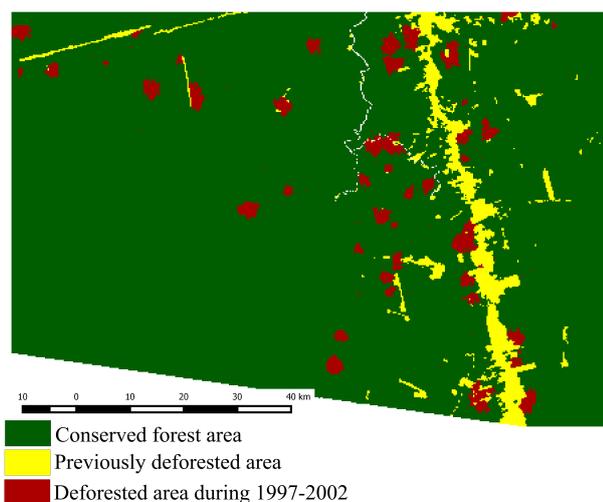


Figure 5: Simulated LUC map (1997-2002) using the CA calibrated by the genetic algorithm. Previously deforested area refers to forest area cleared before 1997

## 5. CONCLUSION

We found that the genetic algorithm was able to calibrate the model's CA to simulate more realistic landscape in term of connectivity. Different spatial patterns can be observed depending on human activities contributing to deforestation as cattle ranching, shifting cultivation, commercial agriculture or logging (Anwar and Stein, 2015; Lorena and Lambin, 2007). This approach can be used to calibrate LUC models and other types of models aiming at simulating landscape patterns. Results show also that more realistic simulated landscapes are often obtained at the expense of the location coincidence. However, when LUC modelling is used to assess processes such as the fragmentation impacts on biodiversity, the simulation of more realistic landscape structure and change dynamics should be preferred to spatial coincidence performance (Malanson et al., 2007; Mas et al., 2012).

## ACKNOWLEDGEMENTS

This study was supported by the *Consejo Nacional de Ciencia y Tecnología* (CONACYT) and the *Secretaría de Educación Pública* through the project *¿Puede la modelacin espacial ayudarnos a entender los procesos de cambio de cobertura/uso del suelo y de degradacin ambiental?*-Fondos SEP-CONACyT 178816.

## References

- Almeida, C. M. D., Monteiro, A. M. V., Câmara, G., Soares-Filho, B. S., Cerqueira, G. C., Pennachin, C. L. and Batty, M., 2005. GIS and remote sensing as tools for the simulation of urban landuse change. *International Journal of Remote Sensing* 26(4), pp. 759–774.
- Anwar, S. and Stein, A., 2015. Spatial pattern development of selective logging over several years. *Spatial Statistics* 13(0), pp. 90–105.
- Cuevas, G. and Mas, J., 2008. Land use scenarios : a communication tool. In: M. Paegelow and M. T. Camacho-Olmedo (eds), *Modelling Environmental Dynamics*, Springer Berlin Heidelberg, Berlin, chapter Land use s, pp. 223–246.
- INPE, 2011. PRODES: Assessment of Deforestation in Brazilian Amazonia.

Lorena, R. and Lambin, E., 2007. Linking spatial patterns of deforestation to land use using satellite and field data. In: Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International, pp. 3357–3361.

Malanson, G. P., Wang, Q. and Kupfer, J. A., 2007. Ecological processes and spatial patterns before, during and after simulated deforestation. *Ecological Modelling* 202(34), pp. 397–409.

Mas, J.-F., Kolb, M., Paegelow, M., Camacho Olmedo, M. T. and Houet, T., 2014. Inductive pattern-based land use/cover change models: A comparison of four software packages. *Environmental Modelling & Software* 51, pp. 94–111.

Mas, J.-F., Pérez-Vega, A. and Clarke, K. C., 2012. Assessing simulated land use/cover maps using similarity and fragmentation indices. *Ecological Complexity* 11, pp. 38–45.

Paegelow, M. and Olmedo, M. T. C., 2005. Possibilities and limits of prospective GIS land cover modelling compared case study: Garrotxes (France) and Alta Alpujarra Granadina (Spain). *International Journal of Geographical Information Science* 19(6), pp. 697–722.

Paegelow, M., Camacho Olmedo, M. T., Mas, J.-F., Houet, T. and Pontius Jr., R. G., 2013. Land change modelling: moving beyond projections. *International Journal of Geographical Information Science* 27(October), pp. 1691–1695.

Pe'er, G., Zurita, G. A., Schober, L., Bellocq, M. I., Strer, M., Müller, M. and Pütz, S., 2013. Simple Process-Based Simulators for Generating Spatial Patterns of Habitat Loss and Fragmentation: A Review and Introduction to the G-RaFFe Model. *PLoS ONE* 8(5), pp. e64968.

Silvestrini, R. A., Soares-Filho, B. S., Nepstad, D., Coe, M., Rodrigues, H. and Assunção, R., 2011. Simulating fire regimes in the Amazon in response to climate change and deforestation. *Ecological Applications* 21(5), pp. 1573–1590.

Soares-Filho, B., Moutinho, P., Nepstad, D., Anderson, A., Rodrigues, H., Garcia, R., Dietzsch, L., Merry, F., Bowman, M., Hissa, L., Silvestrini, R. and Maretti, C., 2010. Role of Brazilian Amazon protected areas in climate change mitigation. *Proceedings of the National Academy of Sciences of the United States of America* 107, pp. 10821–10826.

Soares-Filho, B., Rodrigues, H. and Follador, M., 2013. A hybrid analytical-heuristic method for calibrating land-use change models. *Environmental Modelling and Software* 43, pp. 80–87.

Soares-Filho, B. S., Assunção, R. M. and Pantuzzo, A. E., 2001. Modeling the Spatial Transition Probabilities of Landscape Dynamics in an Amazonian Colonization Frontier. *BioScience* 51(12), pp. 1059.

Soares Filho, B. S., Corradi Filho, L., Coutinho Cerqueira, G. and Leite Araújo, W., 2003. Simulating the spatial patterns of change through the use of the DINAMICA model. In: INPE (ed.), *Anais do XI Simpósio Brasileiro de Sensoriamento Remoto - SBSR*, Belo Horizonte, MG, Brasil, pp. 721–728.

Soares-Filho, B. S., Coutinho Cerqueira, G. and Lopes Penachin, C., 2002. DINAMICA - A stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological Modelling* 154, pp. 217–235.

Soares-Filho, B. S., Nepstad, D. C., Curran, L. M., Cerqueira, G. C., Garcia, R. A., Ramos, C. A., Voll, E., McDonald, A., Lefebvre, P. and Schlesinger, P., 2006. Modelling conservation in the Amazon basin. *Nature* 440, pp. 520–523.

Wang, Q., 1997. Using genetic algorithms to optimise model parameters. *Environmental Modelling & Software* 12, pp. 27–24.

## APPENDIX

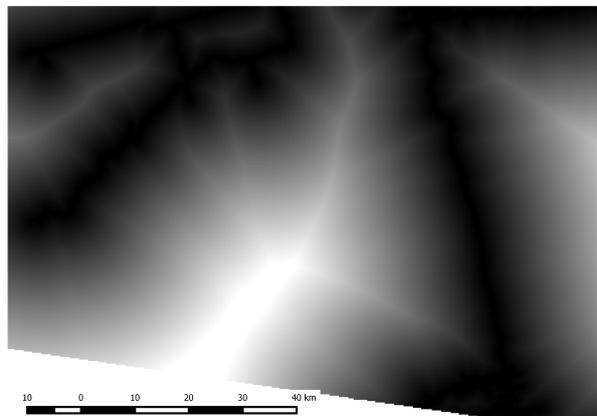


Figure 7: Map of distance from roads used as explanatory variable

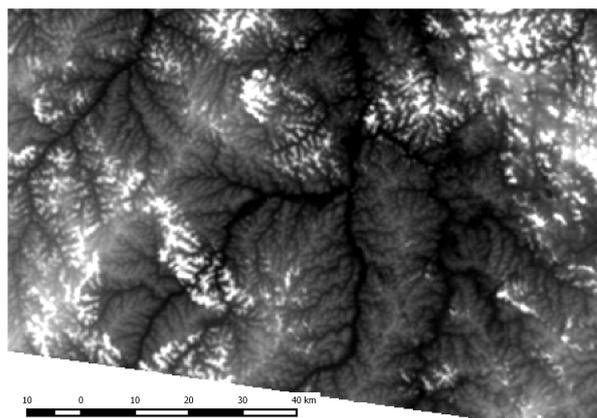


Figure 8: Digital model elevation used as explanatory variable

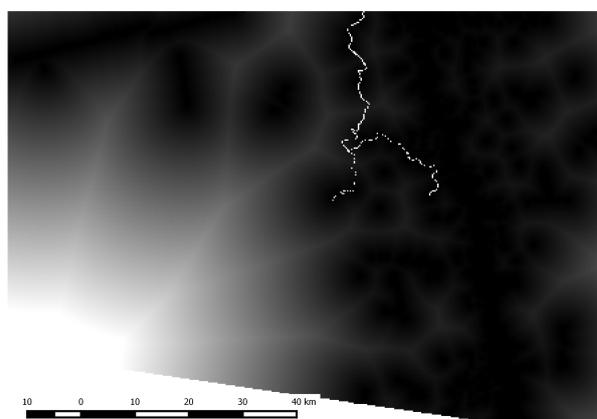


Figure 9: Map of distance from previously deforested area (1997). This map is updated at each modeling step