



HAL
open science

Benchmarking of LUCC modelling tools by various validation techniques and error analysis

Martin Paegelow, Maria Teresa Camacho Olmedo, Jean-François Mas,
Thomas Houet

► **To cite this version:**

Martin Paegelow, Maria Teresa Camacho Olmedo, Jean-François Mas, Thomas Houet. Benchmarking of LUCC modelling tools by various validation techniques and error analysis . *Cybergeo: Revue européenne de géographie / European journal of geography*, 2014. hal-01447925

HAL Id: hal-01447925

<https://univ-tlse2.hal.science/hal-01447925>

Submitted on 27 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Martin Paegelow, María Teresa Camacho Olmedo, Jean-François Mas et Thomas Houet

Benchmarking of LUCC modelling tools by various validation techniques and error analysis

Introduction

- 1 Modelling land use / land cover changes (LUCC) can provide useful information for decision making or for understanding complex social and ecological interactions (Paegelow et al., 2013). The usefulness of LUCC models may be measured by the accuracy of the model output. In this context authors define accuracy as the degree of correctness of simulated land use / land cover (LUC) compared to observed LUC. There are many tools designed for LUCC model validation or for error analysis.
- 2 The intent of this paper is to provide insights that will help users choose appropriate LUCC models. To accomplish this, various map comparison techniques for validation and error analysis are described and applied to simple LUCCs simulated by the three following common modelling tools: CA_Markov, LCM (Eastman 2009) and Dinamica Ego (Soares-Filho *et al.* 2002).
- 3 First authors will implement mentioned models to analyze the accuracy of model outputs focusing on error in quantity, allocation, error in the modeled LUC state and its importance measured as categorical distance, error in LUCC components such as persistence and change and accuracy in spatial pattern and congruence of model outputs. Independently of this study, authors will determine the only impact of training dates on the amount of predicted LUCC.

Overview of map comparison based on validation techniques and error analyses

- 4 Following Torrens (2011), validation involves assessing the success of a model while Croks and Heppenstall (2012) provide the following explanation: “Verification is the process of making sure that an implemented model matches its design, validation is the process of making sure that an implemented model matches the real world”. According to Coquillard and Hill (1997), model validation includes the following progressive steps: verification (also called internal validation, Does the model work accurately?), calibration (Does the model correctly simulate a known situation?) and validation (Does the model correctly predict an unknown situation?). To improve the robustness and the acceptance of a model, the data at the validation date must be model unknown, in other words not used in model building and calibration (Paegelow and Camacho 2008). Otherwise, the simulation becomes a calibration step.
- 5 Rykiel (1996) distinguishes between “conceptual” and “operational” validation. Conceptual validations ensure that the assumptions underlying the conceptual model are correct or justifiable. Operational validations measure the model output accuracy. When modelling the future, only a partial validation may be conducted by comparing the results to an expert’s knowledge, by judging the model robustness by measuring the output stability during iterative model runs or by determining the degree of output congruence between different models that use the same data set and parameters. In addition, data weighting and transformations provide helpful information regarding model performance. Gómez Delgado and Tarantola (2006) propose a sensitivity analysis to test model stability. These authors use several indices to measure the variability of model results based on changing input parameters. Gomez Delgado and Barredo (2005) describe a method to assess risk when using model outputs and Jokar Arsanjani (2012) focus on model data and drivers of uncertainty.
- 6 A large panel of statistical tools exists to measure the accuracy for hard and soft predictions. A validation of a hard prediction results from a comparison made between simulated and

observed LUCs. However, a soft prediction is evaluated to balance potential changes or LUC suitability with observed LUC or LUCC. In this case, the area under the ROC (Relative Operating Characteristics) curve (Pontius and Schneider 2001) is often measured. Eastman *et al.* (2005) and Pérez-Vega *et al.* (2012) evaluate the potential for change in dynamic areas relative to persistent ones (DiP, Difference in Change Potential). Hard validation is more common and uses more developed statistical tools. These statistical tools focus on the following aspects: accuracy of quantity, LUCC components, landscape pattern, model congruence in the sense of similar results between different models, and error analysis.

7 Regarding quantitative agreement, model users distinguish between matching the sum of the LUC area and the pixel by pixel comparison, which includes allocation (Torrens 2011). Overall agreement may also be obtained by using statistical indices, such as Chi-square or Kappa (Pontius 2002). However, Pontius and Millones (2011) indicate that the Kappa index is not convenient for LUCC model validation because this index assumes randomness and needs to convert the sample matrix to an estimated population matrix. Authors recommend for map comparison easier indices like quantity and allocation disagreement. Various validation techniques that consider changes were developed. For example, Pontius (2000) and Pontius *et al.* (2004a, 2004b, 2008) propose a technique that splits the LUCC-budget into gain, loss, net change and swap. Pontius *et al.* (2008) also developed several statistical LUCC indices to determine accuracy, including a figure of merit, a ratio between correct predicted changes and the sum of observed and predicted changes.

8 Other validation approaches focus on allocation agreement. Fuzzy logic-based indices (Hagen 2003, Hagen-Zanker *et al.* 2005, Rodrigues *et al.* 2007) measure the allocation agreement and overcome the restrictions induced by hard pixel limits, pattern quantifications and the exclusive cell state. Procrustes analysis (Jackson 1995) compares the fit between different matrices by linearly transforming one grid as rotation, translation or scaling to achieve the best fit with the reference grid. Kuhnert *et al.* (2005) propose raster map comparison algorithms and varying window size and weighting techniques.

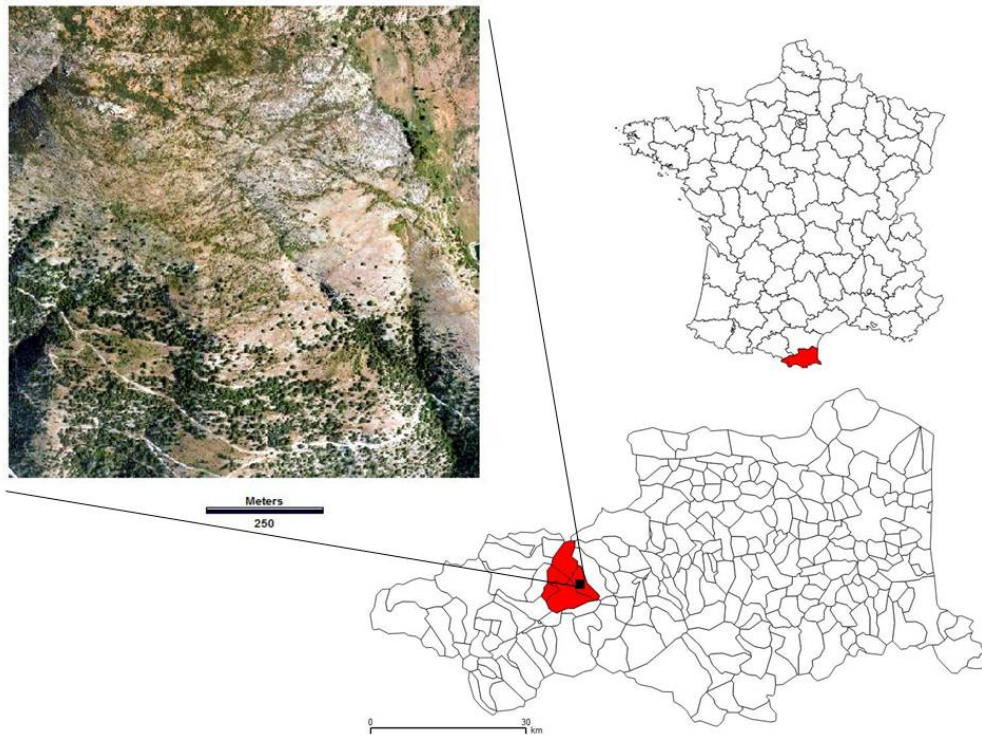
9 Spatial analysis measurements account for land patterns and their distribution and shapes (White *et al.* 1997) at multiple scales (Gaucherel 2007, Gaucherel *et al.* 2008) and are mainly inspired by landscape ecology metrics (Forman 1995, McGarigal and Marks 1995, Botequilha *et al.* 2006). Error analysis emphasizes conceptual and model parameter inaccuracy by measuring errors in allocation and predicted state of LUC (Pontius 2000, Pontius and Petrova 2010). Several papers provide a comprehensive review of validation techniques that are designed for spatial models (Turner *et al.* 1989, Pontius *et al.* 2004a, Paegelow and Camacho 2008, Shirley and Battaglia 2008).

Data set and study area

Study area

10 Garrotxes is an 8 750 ha catchment located in the western part of the Eastern Pyrenees (France). The lowest area in this region is located in the SE and varies from 650 m at the confluence of the Têt river to 1 000 m. A Mediterranean climate dominates this low area. In contrast, the upper region reaching 2400 m is influenced by the mountain climate. The western area of this region is characterized by a ponderous geomorphologic relief on granite. This area is composed of early terrace cultivation and coniferous forests (*Pinus uncinata* and *P. sylvestris*). The east bank forms a large, steep and south facing area that overlies schist and is used as a pasture. The demographic maximum, which occurred during the decade of 1820, corresponded with intensive use of all natural resources in this area. According to the Napoleon cadastre, a quarter of the Garrotxes catchment was terraced for crops in 1826. Today, the crops have entirely disappeared. The population fell from 1832 inhabitants in 1826 to 94 inhabitants in 2008. Crop terraces were transformed into pastures prior to becoming shrub or forest areas. Currently, the crops grown in this area are marginal. In addition, the near future likely depends on the intensity of pastoral activity and management, which will determine how far the forest spreads (Paegelow and Camacho 2005).

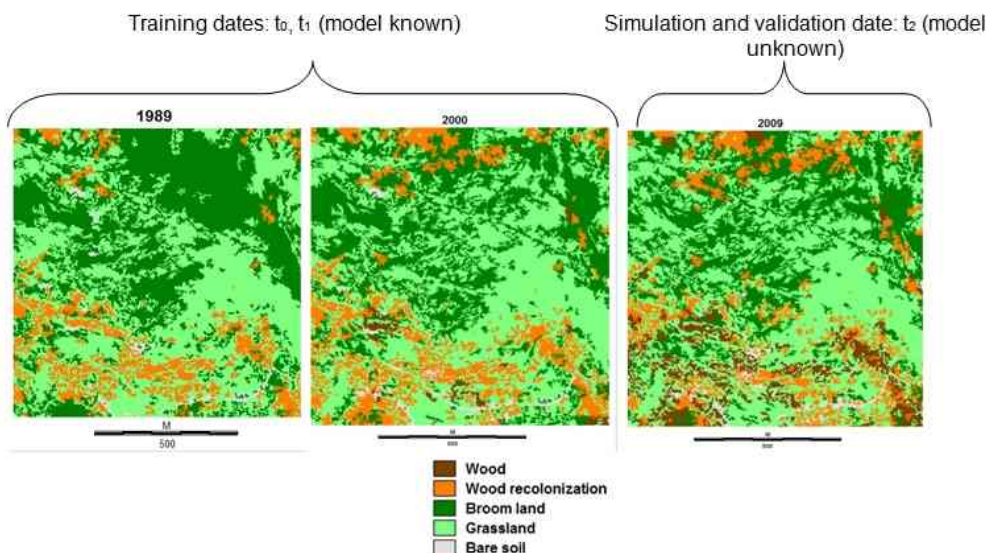
Figure 1: Location of the study area



Top right: France and department of Eastern Pyrenees. Down right: Garrotxes basin formed by 5 municipalities and position of test area (black square). Top left: 1 km² sample area. IGN 2004 orthophotograph

Data set

Figure 2: Land cover maps: training dates (1989, 2000) and validation date (2009) for the simulations



Data source: aerial photographs, IGN (Institut Géographique National), France

11 The sample area that was considered in this paper forms a 1x1 km square located on the eastern pastoral slope of the Garrotxes. The land cover is described by the following categories: wooded area (*Quercus ilex* in the southern area and *Pinus uncinata* in the higher northern areas), wood recolonization, meaning area with scattered trees, broom land (mainly *Genista purgans*), grassland and bare soil. LUC maps are available for 1980, 1989, 1995, 2000, 2004 and 2009. Coarser resolution maps with more detailed categories exist for the entire Garrotxes catchment area. The only remaining land use in this area is pastureland.

Model data set

- 12 The data sources that are used to produce the LUC maps include aerial photographs, DEM (digital elevation model, derived from topographic maps at 1:25,000 by interpolation) and field surveys after 1994. LUC maps for 1989 (t_0) and 2000 (t_1) are used to train the models while data for 2009 (t_2) are employed for model validation. The LUC maps are derived from segmentation and maximum likelihood classification by using panchromatic aerial photographs (2000 and earlier) or color orthophotographs (2004, 2009) (Paegelow and Camacho Olmedo, 2010). Classification errors are corrected by hand and the results are validated with an *in situ* survey. Several potential drivers explain the LUCCs, including the type of land cover, elevation, slope, aspect, distance from thalwegs and pasture fire management (maps with the location and approximate date of the fire). All of these data are in raster format with a pixel resolution of 0.5 m. These drivers are available for all of the used models. In addition, they are used by all of the models but may have different weights.

Methods

Modelling approaches and tools

- 13 Three model tools incorporating different modelling approaches were run using the dataset. All of these approaches use Markov chains to predict the quantity of LUCC but present differing techniques how to get knowledge about spatial distribution and how to allocate land categories or land changes in space. These expected quantities are spatially allocated using suitability or change potential maps. The three following modelling approaches are commonly used and debated by authors:
- A supervised method to assign expected conditional probabilities performed by Markov chains in space using a multi-criteria evaluation. CA_Markov implemented this method with the Idrisi software (Houet and Hubert-Moy 2006, Eastman, 2009).
 - A dynamic model that spatially assigns expected conditional probabilities performed by Markov chains and uses a neural network which is a self-learning algorithm, such as the Land Change Modeler (LCM). This model was implemented in the Idrisi software (Eastman 2009, Villa *et al.* 2006).
 - A dynamic cellular automaton that uses weights of evidence and is able to account for landscape pattern and sojourn time in the Dinamica Ego freeware (Soares-Filho *et al.* 2002, Mas *et al.* 2011).

Model implementation

- 14 The three modelling tools use Markov chains to compute the expected conditional probabilities for land cover categories. Nevertheless, the software algorithms are slightly different, as shown by Mas *et al.* (2011, 2014), and the default options are different. The LCM model computes a Markov chain based on a 100 % confidence level and the Markov module, which is used in CA_Markov, allows the user to specify the probability that LUC is incorrect. Called proportional error, the documentation of Markov module suggests a value of 0.15 for remote sensing data. To assure comparability, we computed a 0.15 proportional error Markov chain and introduced this matrix into all models. In our case study land cover maps for 1989 and 2000 are used as inputs for the three models. LUCC simulations were done on a yearly basis and 2009 was extracted for comparison with the 2009 validation map.

The main differences between the modelling approaches are listed below:

- CA_Markov is a static model and the LCM and Dinamica models are dynamic because they can update drivers and land cover factors, such as distance, at each simulation step
- CA_Markov is an expert driven process that spatially allocates expected categorical LUC by using categorical suitability maps. In the LCM model, the user introduces relevant drivers and the type and strength of the LUCCs are determined with a multilayer perceptron, which is an automatic machine learning algorithm. The Dinamica Ego model establishes relationships between the LUCCs and the drivers by using conditional

probabilities that are called weights of evidence. Otherwise, expert editing is possible in the Dinamica Ego model. The user is able to make manual corrections through a graphic tool and can select / blend out transitions to be modeled. The LCM model allows the user to choose transitions and to merge some of them. However, the CA_Markov model can implicitly manage a selection by using land cover constraints.

- As previously mentioned, the Dinamica Ego model simulates landscape patterns by controlling the mean patch size, its variance and isometry of simulated patches by splitting the LUCC into two processes. These two processes use the following cellular automata: i) the spreading or regression of the existing patch borders or ii) the creation of new patches. LCM does not provide any tool to manage this aspect and the CA_Markov model proposes a contiguous filter to down-weight suitabilities and to reduce the salt and pepper effect. The net effect is a lower number of larger landscape patches. Form and isometry of the patches cannot be directly modeled by the CA_Markov model unless they are expressly implemented in suitability images.
- The LCM and Dinamica models offer soft prediction maps for each considered LUC transition.
- Another main conceptual difference between these three modelling approaches is the modeled object. While the CA_Markov model models the LUC categories, the LCM and Dinamica Ego models explicitly model transitions between categories (Camacho Olmedo *et al.*, 2013).

Map comparison techniques performed for validation and error analysis

ROC validation considering suitability / transition potential maps

15 Hard prediction must make a unique choice that affects one LUC for each pixel. Therefore, this simulation is binary and may or may not be accurate when prediction errors cannot be evaluated by simply asking ‘how inaccurate is the simulation?’ Some modelling tools provide soft prediction maps that indicate the vulnerability of the land to change. The LCM and Dinamica models offer change potential maps for each considered LUC transition, which are expressed as rankings. The CA_Markov model uses suitability maps for each LUC category, which are computed by multi criteria evaluation (MCE) (Eastman *et al.* 1995).

16 Relative operating characteristic (ROC) (Pontius and Schneider 2001) is, in this context, a measure of the spatial likelihood between a binary presence/absence reference map for a considered LUC category or LUCC transition and a suitability map for this category or a map expressing the transitions propensity to change, respectively. The ROC ranks these suitability or vulnerability scores into n classes and calculates the quantity of true (presence in reference map) and false (absence) positives. The hypothesis is that the high scores in the comparison map are more likely to be truly positive. ROC results vary between 1 (perfect fit) and 0 (perfect negative association) with 0.5 indicating a random relationship. This number is referred to as the AUC (area under curve) proportion. Pontius and Schneider (2001) provide a graphic illustration for this technique. In this context, 10 % thresholds are applied and the resulting ROC values are expressed as a percentage from 0 to 100.

17 Specifically, the spatial likelihood, which is the “agreement in terms of location of cells in a category” (Eastman 2009), was measured for MCE suitability maps based on the training data set and then compared to the given LUC category at the validation date. This method measures the spatial likelihood for both: persistence and occurred changes. In contrast, the soft prediction potential is tested on LCM and Dinamica Ego modeled LUCC transitions and is compared to the corresponding observed transitions between 2000 and 2009. In addition, we also tested the prediction potential of the MCE suitability maps for changes. For this test, the authors masked in the suitability maps all areas outside the considered and observed LUC transition between 2000 and 2009. This test makes the ROC results comparable and provides information regarding the spatial likelihood for the following CA_Markov soft prediction potentials: overall, confusing persistence and changes, and only changes.

Validation by quantity

18 The classic method for estimating model accuracy is to fit the quantity. Authors first compared the observed categorical areas from 2009 to the simulated areas. A synthetic indicator of disagreement is half the sum of the absolute differences between the observed and model specific simulated LUC (Pontius *et al.* 2004b), called here the overall absolute difference. In parallel, the accuracy is measured by a pixel by pixel comparison between the observed LUC matrix (lines) and the simulated one (columns). The overall accuracy is expressed by the sum of the diagonal pixels converted into a percentage of area. The null hypothesis is used as a reference for the LUCC model accuracy. The null hypothesis corresponds with only persistence. In this case, the null hypothesis showed no change between 2000 and 2009. In addition, a synoptic accuracy index, the Cramer's V, is computed.

Validation considering land cover persistence and changes

19 Pontius (2000) and Pontius *et al.* (2004a, 2004b) established a comprehensive way to analyze LUCC and to measure the accuracy of the model outputs based on the LUC persistence and changes. In this context, we compared the LUCC budget from between 2000 and 2009 to the LUC in 2000 and the simulated LUC in 2009, respectively, with the CA_Markov, LCM and Dinamica models. First, we compared the overall LUCC before considering the individual categorical LUCCs. Both of these values are expressed by the three following indicators:

- Total change - expresses the overall change (gains and losses) between two LUC maps (dates).
- Absolute net change - the absolute balance of the sum of gains and losses for each LUC category (e.g., 2% gain and 4 % loss results in an absolute net change of 2%).
- Swap - the difference between total change and absolute net change that expresses a change of allocation without a change of quantity.

20 A more thorough LUCC analysis by Klug *et al.* (1992), Perica and Fofoula-Georgiou (1996) and Pontius *et al.* (2008) splits the map comparison between the observed and simulated LUCs into percent correct and percent error as follows:

A = Error due to observed change predicted as persistence

B = Correct due to observed change predicted as change with correct gaining category

C = Error due to observed change predicted as change but as wrong gaining category

D = Error due to observed persistence predicted as change

21 These components allow for the calculation of the three following derived measurements:

- Figure of merit is the ratio of $B / (A+B+C+D)$ and expresses the overlap between the observed and predicted change. This value ranges from 0 (no overlap) to 100 % (perfect overlap).
- Producer's accuracy is the ratio of $B / (A+B+C)$ and expresses "the proportion of pixels that the model predicts accurately as change, given that the reference maps indicate observed change" (Pontius *et al.* 2008).
- User's accuracy is the ratio of $B / (B+C+D)$ and measures the proportion of the pixels that the model predicts accurately as change when all model predicted changes are given.

Validation by spatial pattern analysis

22 In addition to quantitative accuracy measurements, the landscape pattern agreement is a supplementary validation approach. In ecology, landscape pattern indices, such as shape, compactness, diversity and fragmentation, have been commonly measured since the 1960s. In this context, we simply calculate the two following basic pattern indicators: the number of patches for each observed and simulated LUC and its average size.

Validation by congruence of models

23 To measure the congruence of model accuracy and the individual model contributions, the three simulation maps of correct predicted land are intersected by the logical operator AND. The intersection score measures the congruence of the LUCC models. In addition,

all supplementary contributions of the two model combinations (CA_Markov and LCM, CA_Markov and DINAMICA, LCM and Dinamica) are calculated with the remaining individual contributions. Likewise, the same analysis is performed to distinguish the correct predicted persistence and change.

Error analysis

24 Error analysis provides information regarding model specific logic and the underlying conceptual approaches. In other words, error analysis allows the modeler to better understand the model. In addition, validation techniques are completed by analyzing the possible origins of error. Following this distinction, one may consider LUCC analysis and the figure of merit as validation techniques and error analysis. The question of 'how wrong is the prediction?' may be split into the two following components: error in the modeled LUC state and error in allocation.

Error in the modeled LUC state

25 Various techniques are used to measure disagreement. While quantitative data (e.g., percent of tree cover) can measure the magnitude of inaccuracy, categorical data generally needs to be transformed, (for example, based on evidence likelihood) to compute quantitative data. Ahlqvist (2008) provides fuzzy change estimation about the closeness of LUC categories. Here, the authors propose to evaluate the magnitude of error between the predicted and observed LUC. The LUC is expressed by qualitative data (categories). However, the used LUC legend is a ranking that reflects spontaneous vegetation succession from bare soil to wood. There is no physical limit for spreading vegetation, such as altitude or soil, that block this evolution. In addition, authors are unable to indicate a specific distance between categories. Therefore, they rank the categories from 5 (bare soil) to 1 (wood). Inaccurate prediction is measured by the absolute categorical distance between the observed and simulated LUC. This error distance scale reaches from 1, e. g. observed grassland and predicted broom land or bare soil, to 4, e.g., observed wood and predicted bare soil.

Error in allocation

26 For each LUC category observed in 2009, a distance map is calculated. This distance map is crossed with simulation errors (error categories A, C and D in figure 5). For each wrong predicted patch of a given LUC category, we retained the minimal distance to the nearest correct location and then computed, for each LUC category, the average distance in meters for all errors.

Impact of training dates on Markovian transition matrices

27 The main difference between the three LUCC models lies in the spatial allocation of the predicted Markov chain quantities. Thus, the two following questions emerge: i) Does the Markov chain constitute the appropriate time prediction model? ii) Is the Markov chain able to correctly predict the past and present LUCC? To answer these questions, the authors compare the observed LUC rates from the six available dates to the various Markov transition matrices that were performed by using only two LUC maps. This analysis attempts to determine the impact of the chosen training dates on the predicted LUCC. Otherwise, this comparative analysis may emphasize the limitations of the Markov chains, depending on the degree of temporal instability in the system.

28 To do so, authors apply Markov chain on all available dates: two previous training dates are used to predict the amount of LUC at the next date. The disagreement between the observed and the Markov chain predicted LUC quantities are measured as half of the sum of the categorical absolute differences. This test is i) independent from LUCC modelling performed by CA_Markov, LCM and Dinamica and ii) driven afterwards. When performing LUCC modelling, 2009 validation data still was not available.

Results

ROC validation considering suitability / transition potential maps

29 Table 1 expresses ROC results for MCE suitability maps which include persistence and change (on the left) and ROC results restricted to changing areas (on the right) . The ROC results for

the MCE suitability maps, which are used in the CA_Markov model for spatial allocation of the predicted Markov quantities include both persistence and change (cf. 2nd column “Overall CA_Markov”). Although persistence is dominant, it is not surprising that the categorical LUC ROC scores are higher than those performed for specific changes (cf. columns 2, 3 and 4 entitled “Changes”). All LUC categorical suitability maps have high spatial likelihood scores. For example, the un-weighted simple average is approximately 94 %.

30 The right side of table 1 is entitled “Changes” and contains the ROC values for the LCM and Dinamica EGO that were performed with the specific LUC transition potential maps and were compared with the corresponding changes observed between 2000 and 2009. The spatial prediction potential of the two LUCC models is always lower than the soft prediction potential of the CA_Markov for both persistence and changes. When comparing the LCM and Dinamica Ego performances, the Dinamica Ego generally obtains higher ROC scores, which may be caused by the Dinamica Ego specific tools that are used to model the spatial pattern. This software distinguishes between linear growth on existing pattern limits and disconnected appearance of new pattern. The two modules controlling this process, called “Exapnder” and “Patcher” provide also the ability to set the mean patch size and form. However, the LCM obtains some transition specific ROC values that are higher than those in the Dinamica Ego. Considering the LUC changes to wood, the LCM’s soft prediction potential to model changes from broom land to wood fit the observed changes. However, the soft potential for transitioning from wood recolonization to wood is lower in the LCM than in the Dinamica Ego.

31 The CA_Markov suitability maps are compared to the LCM and Dinamica Ego transition potential maps. The third column of table 1 contains ROC scores for the CA_Markov suitability maps that had been compared to the specific changes mentioned in the table. We changed the reference map in the ROC test from the LUC category observed in 2009, including persistence as change, to the specific observed transition that occurred between 2000 and 2009. The suitability map results regarding the spatial likelihood that were restricted to changes were even less than those applied to the entire LUC. However, the CA_Markov soft prediction potential is, on average, comparable to the applied LUCC models, which approach modelling transitions rather than LUC. On average, the CA_Markov suitability maps obtained intermediate results relative to the LCM and Dinamica Ego maps. There is no evident relationship between categorical net change and swap proportions (figure 4).

Table 1

LUC 2009	Overall CA_Markov	CHANGES			
		CA_Markov	LCM	Dinamica	2000 to 2009 LUC transition
1 – wood	95.8	94.0	30.7	69.0	2 to 1
		82.4	92.1	69.6	3 to 1
2 – wood recolonization	87.8	63.8	78.0	63.6	3 to 2
3 – broom land	92.9	45.1	63.8	73.8	4 to 3
4 – grassland	96.3	34.2	56.9	73.4	3 to 4
		38.0	22.8	79.0	5 to 4
5 – bare soil	95.6	95.2	40.5	84.0	4 to 5
average	93.7	64.7	55.0	73.2	

ROC likelihood for the CA-Markov categorical suitability compared with the observed LUC in 2009 (2nd column OVERALL). In addition, the CA_Markov categorical suitability and the LCM and Dinamica predicted transition vulnerability compared with the observed 2000 to 2009 LUC transition values (columns entitled CHANGES).

Validation by quantity

32 All models underestimate wood and overestimate wood recolonization (Table 2). The Dinamica model was the closest to reality and the LCM model produces a quantity of error that is three times larger.

Table 2

Observed	Simulation 2009
----------	-----------------

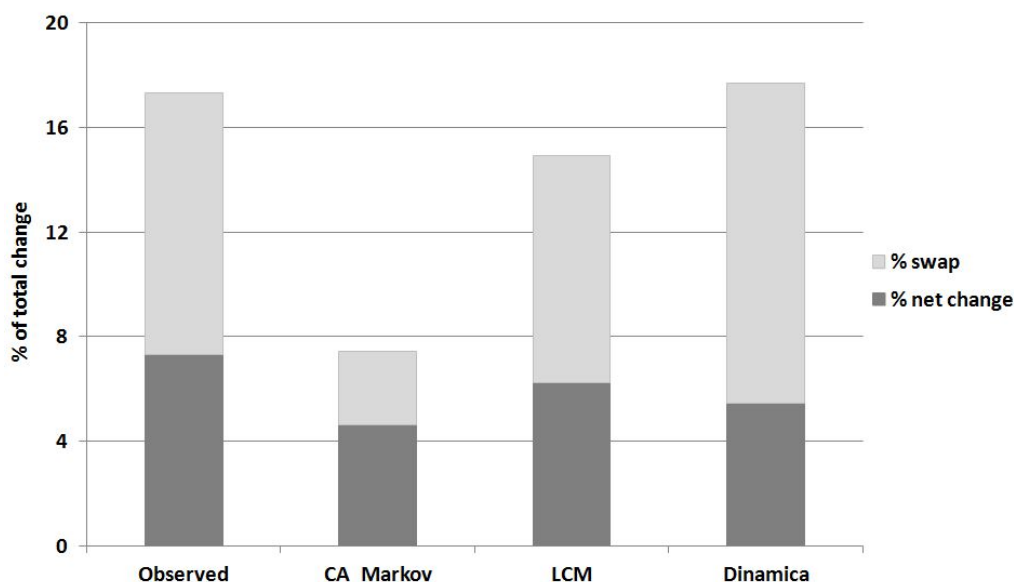
			Quantity in %			Pixel by pixel fit in %		
	2000	2009	CA-Markov	LCM	Dinamica	CA-Markov	LCM	Dinamica
Wood	1.50	8.81	4.75	5.17	6.83	3.48	3.58	3.76
Wood recolonization	15.95	13.64	16.60	17.54	14.28	8.29	8.30	6.48
Broom land	34.45	30.25	30.54	31.96	31.75	24.19	24.02	21.51
Grassland	46.76	45.96	46.35	43.02	45.68	43.60	39.32	39.09
Bare soil	1.34	1.34	1.75	2.31	1.46	1.34	0.63	0.91
Overall absolute difference			8.12	13.15	4.51			
Accuracy of null (no change)		82.66						
Accuracy of LUCC model						80.90	75.85	71.75

Accuracy in quantity and pixel by pixel (%) between the observed 2009 (validation map) and simulation maps computed by CA_Markov, LCM and Dinamica Ego.

33 The pixel by pixel comparison between the observed and predicted LUC values for 2009 integrates the quantity and allocation accuracy. The right side of table 2 contains the categorical and overall accuracies. Categorical scores depend on the land cover and model. For example, all models show a better fit for grassland. In contrast, wood recolonization, which is the most dynamic category, and wood and broom land have lower pixel by pixel scores than the observed LUC. The overall accuracy, which is always lower than in the null hypothesis (82.66 %), results in a different model ranking than that observed for only quantity. For example, the CA_Markov model has the best pixel by pixel prediction score and the Dinamica model has the worst. Statistical measurements between the observed and simulated LUC parameters, such as the Cramers V (CA_Markov 0.72; LCM 0.59; Dinamica 0.60), show results that are close to the overall prediction accuracy that is obtained from the pixel to pixel comparison.

Validation considering land cover persistence and changes

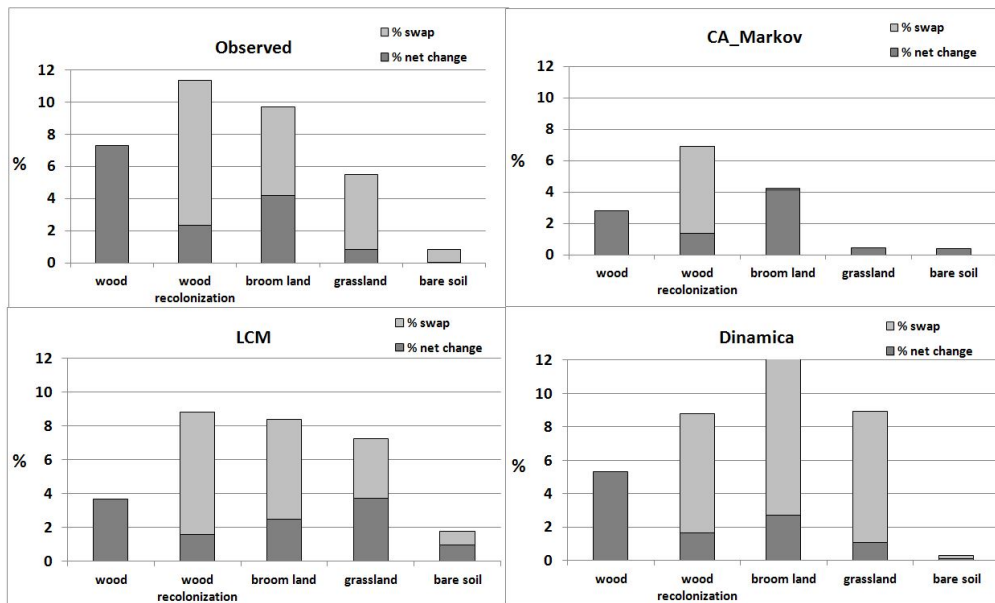
Figure 3



Observed LUCC budget from 2000 to 2009 compared with the LUCC budgets from 2000 (observed) and 2009 as simulated by CA_Markov, LCM and Dinamica. The bar length expresses the % of total change, the dark grey portion indicates net change and the grey part indicates the swap proportion.

34 The observed LUCC budget resulted in a total change of 17.34 % (figure 3). Therefore, the null hypothesis which means no change between 2000 and 2009 is 82.66 % correct. The Dinamica model is close to reality and the CA_Markov model underestimates the LUCCs. Thus, the best accuracy for the CA_Markov model resulted from persistence. The split of total change into net change and swap shows that net change is the minor process (7 %) and swap the major phenomenon (10 %). The LCM model ratio is the closest to reality and the CA_Markov model is the only model that gives a larger net change than swap value.

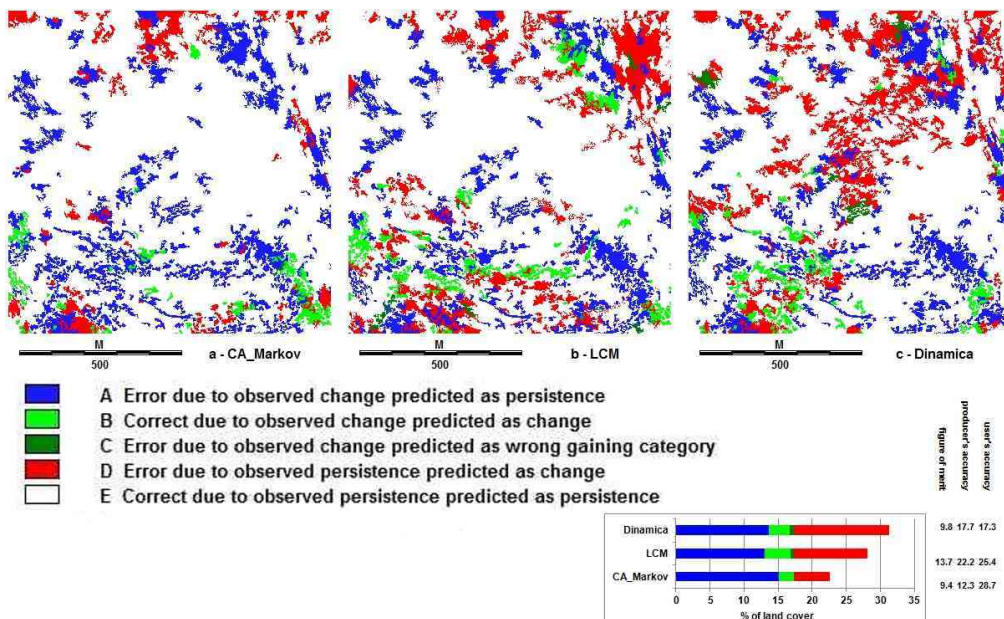
Figure 4



Total change, net change and swap proportion for each LUC category. Top left: observed, top right: CA-Markov, bottom left: LCM and bottom right: Dinamica. The bar length expresses the % of total change, the dark grey portion indicates net change and the grey part indicates the swap proportion.

35 Figure 4 provides more detailed information regarding LUCCs based on categories. The following considerations can be made. All models accurately modeled change in wood as 100 % growth (net change). Although these proportions are lower than reality and vary depending on the used model. All models underestimate the dynamic character of wood recolonization. This category represents a transition state between shrubby broom land, and woodland. In reality, the gains by wood recolonization from broom land are nearly balanced by losses in other areas, especially to wood. The swap component is underestimated in the CA_Markov model except for wood recolonization.

Figure 5



Validation maps for the CA_Markov, LCM and Dinamica models and the sources for the percent correct and percent error during validation.

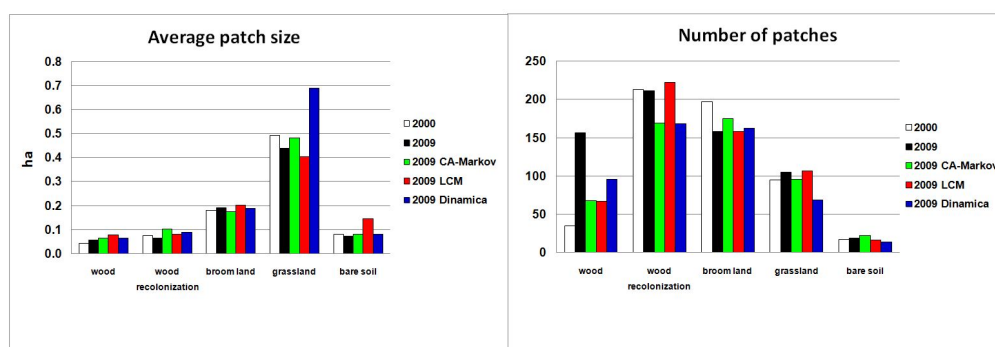
36 Figure 5 is inspired by Pontius *et al.* (2008). The bar diagram values correspond with the maps. The fifth category, which is correct due to observed persistence predicted as persistence, is not included in the bar diagrams and corresponds to the percent missing from 100. Figure 5 illustrates that land cover persistence varies from 69 % in the Dinamica model to 77 %

in the CA_Markov model. As previously mentioned, persistence is dominant in Garrotxes. Considering that the main component of change is swap (cf. figure 4), the LUCCs in Garrotxes are not easy to model. In addition, figure 5 contains the main prediction error components (error due to observed change predicted as persistence and *vice versa* and represented in blue and red), which are nearly equivalent with the Dinamica and LCM models. However, in the CA_Markov model, the error that is due to observed persistence predicted as change is three times smaller as the opposite. Another interesting fact is that the error due to observed change predicted as wrong gaining category is very small for all models. For example, in the CA_Markov model, this error is essentially nonexistent (0.04%).

37 All performed LUCC simulations obtain low figure of merit scores. However, the LCM model produces the highest relative scores. Regarding the figure of merit, the LCM producer's accuracy, which is the correct predicted proportion of changes for all observed changes, is the less low. In contrast with the producer's accuracy, the user's accuracy, which expresses part of the correct predicted change given all model predicted changes, is the highest in the CA_Markov model followed by the LCM model.

Validation by spatial pattern analysis

Figure 6



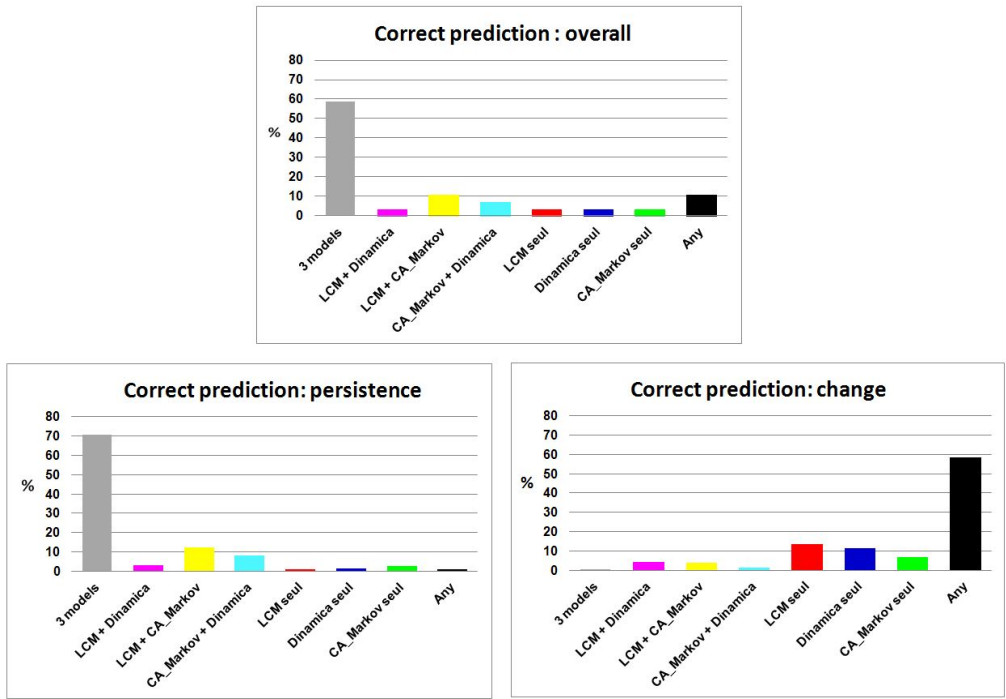
The number of patches and the average patch size (ha) for observed land cover in 2000 and 2009 and the CA_Markov, LCM and Dinamica simulated land cover in 2009.

38 Figure 6 shows landscape pattern agreement between the observed and simulated landscapes for 2009. In addition, it represents patch number and size for the observed landscape in 2000, which allows for comparison between the observed and simulated changes. While the number of patches is partly dependent on the categorical LUC area, the average patch size is not. The most visible observed change is the increase of wood patches and the decrease of broom land patches. As previously mentioned, all models underestimate the wood gain. The number of simulated wood patches is approximately half of the observed wood patches in all models other than the Dinamica model, which is the only model in which patch criteria are specified. Regarding the other LUC categories, the simulated number of patches is close to reality and the individual LUCC model performances are heterogeneous. Figure 6 shows that all models generate landscape size patches nearby observed values except for the Dinamica model for grassland and the LCM model for bare soil. These two LUC categories have the highest proportion of swap compared to net change (Figure 4). In addition, the standard deviations remain close to the average area scores (data not shown).

Validation by congruence of different models

39 Figure 7 contains the overall, the persistence and the change correct predicted scores. First, the overall correct prediction congruence, measured as the intersection of the three used models, is high (58.6 %) according to the individual model accuracy (the lowest LCM is approximately 71.8 %, Table 2) and the described methodological differences. Supplementary contributions to this prediction base are produced by the two model pairs. Overall, the prediction is highest for pairs involving the CA_Markov model. The CA_Markov model contribution has the highest remaining individual model contribution.

Figure 7

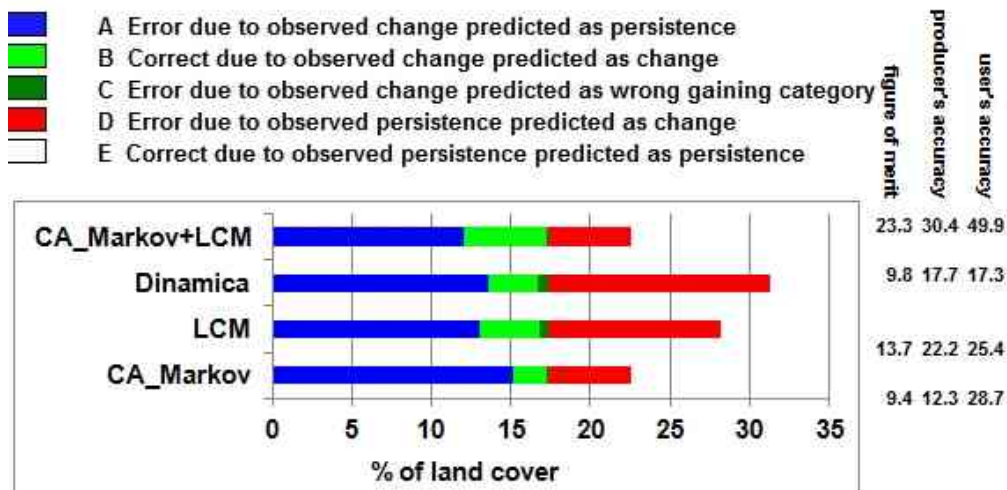


Overall, persistence and change correctly predicted by all models (intersection), the supplementary contributions by pairs of models and the remaining individual model scores. Scores are given in per cent of observed.

40 Focusing on the persistent land between 2000 and 2009, similar results are observed with a higher level (70.4 % of the correct prediction based on the intersection of all models). Overall, less than 1 % of the land is incorrectly predicted by any of the models. The supplementary CA_Markov model contribution is substantial. Regarding only change, any model is able to correctly simulate 58.3 % of the observed land change. In addition, the amount of correct predicted change by all models is close to zero (0.5 %). The LCM and Dinamica model contributions are higher than the CA_Markov model contribution.

41 The different abilities of the models for predicting persistence and change suggest that the models can be combined. Figure 8 shows a very simple combination. Essentially, the LCM (best figure of merit) model component with correct pixels due to observed change predicted as change (B) is inserted into the CA_Markov (best land cover persistence score) prediction map. The resulting overall pixel by pixel accuracy of 82.68 % is roughly equal to that of the null hypothesis although the figure of merit is approximately doubled.

Figure 8



Sources of percent correct and percent error in validation including the combination of LCM correct predicted change (B) inserted into the CA_Markov model simulation map.

Error analysis

Error in the modeled LUC state

42 Table 3 shows the amount of categorical error distance (left) which sum is the complement to 100 percent in table 2. For the CA_Markov model, table 3 indicates that 18.48 % of the land suffers from a prediction error in 1 category. This means that the wrong LUC category is adjacent to the correct category into the caption. 0.62 % is affected by prediction error 2 categories wide. The right side of the table expresses these categorical distance prediction errors as proportions and the sum of proportions is 100. Any model predicts errors that are characterized by a categorical distance greater than 2. More than 93 % of the prediction errors for each tested model have a categorical distance of only one category to observed LUC. Although persistence is the major phenomenon, the CA_Markov obtains the highest scores.

Table 3

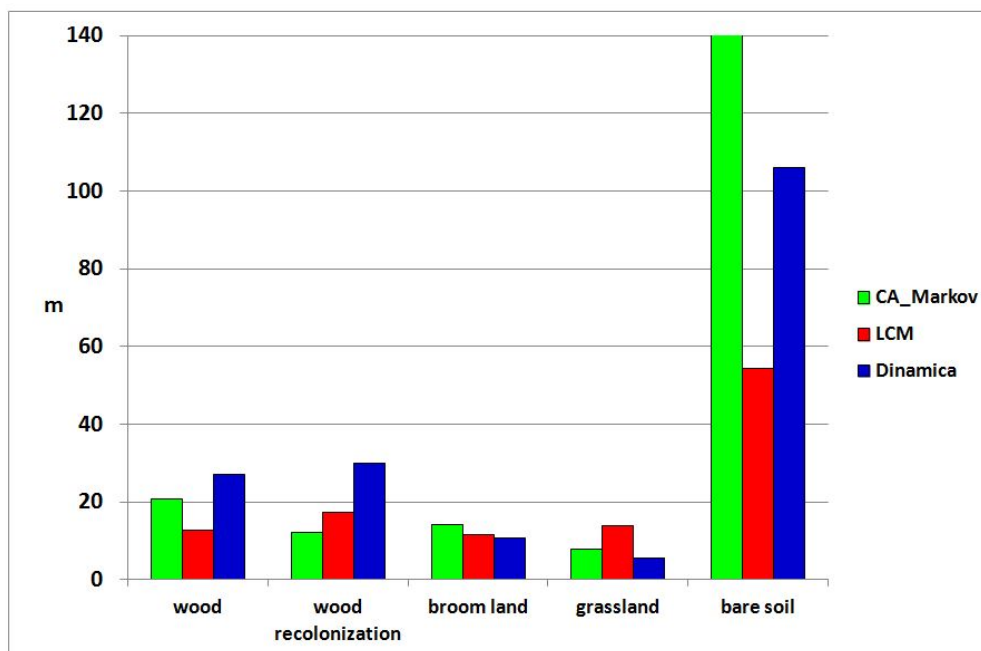
Error distance	% inaccurate predicted			% proportional error		
	CA_Markov	LCM	Dinamica	CA_Markov	LCM	Dinamica
1	18.48	21.43	26.90	96.75	93.68	95.24
2	0.62	1.45	1.35	3.25	6.32	4.76
3	0	0	0	0	0	0
4	0	0	0	0	0	0
Sum	19.10	22.88	28.25	100	100	100

Categorical distance for prediction errors (% of LUC) and the percent of proportional error (sum of errors = 100) by categorical distance for the CA_Markov, LCM and Dinamica models. The prediction errors correspond to categories A, C, and D in figure 5.

Error in allocation

43 Figure 9 shows the average distance (m) of prediction errors to the next observed LUC category in 2009 for the different LUCC models. Two general considerations are made. First, the model specific differences are less than the categorical divergence. Secondly, an important error distance affects bare soil, which is the category with the lowest extent and error (cf. table 2). Down weighting the average error distance by the error importance (percent of area), the CA_Markov (13.9 m) and LCM (14.0 m) models obtain nearly the same average allocation error score. In contrast, the Dinamica model induces an average down weighted distance error of 17.5 m.

Figure 9



The average distance of prediction errors (m) to the next corresponding observed LUC (2009) in the CA_Markov, LCM and Dinamica models. The prediction errors correspond to categories A, C and D in figure 5.

Impact of training dates on Markov transition matrices

44 Table 4 shows the observed and Markov chain predicted LUC. The two previous dates were used to predict the next one (italic lines: 1995, 2000, 2004 and 2009). The chosen Markov chain prediction for 2009, used in the implementation of the three LUCC models, is based on 1989 and 2000 (last line, bold and italic). As shown in the right column of table 4, the quantity error varies depending on the dates used and the resulting LUCC. In our example, the closeness of the Markov chain predicted quantity with the observed LUC extent does not depend on the duration of the considered period. Predicting land cover for 2009 by Markov chain using the two earlier training dates is more accurate than for 2004. Table 4 also shows that LUC prediction for 2009 based on 1989 and 2000 is more correct than using 2000 and 2004 as training dates however 2000 and 2004 represent more recent data.

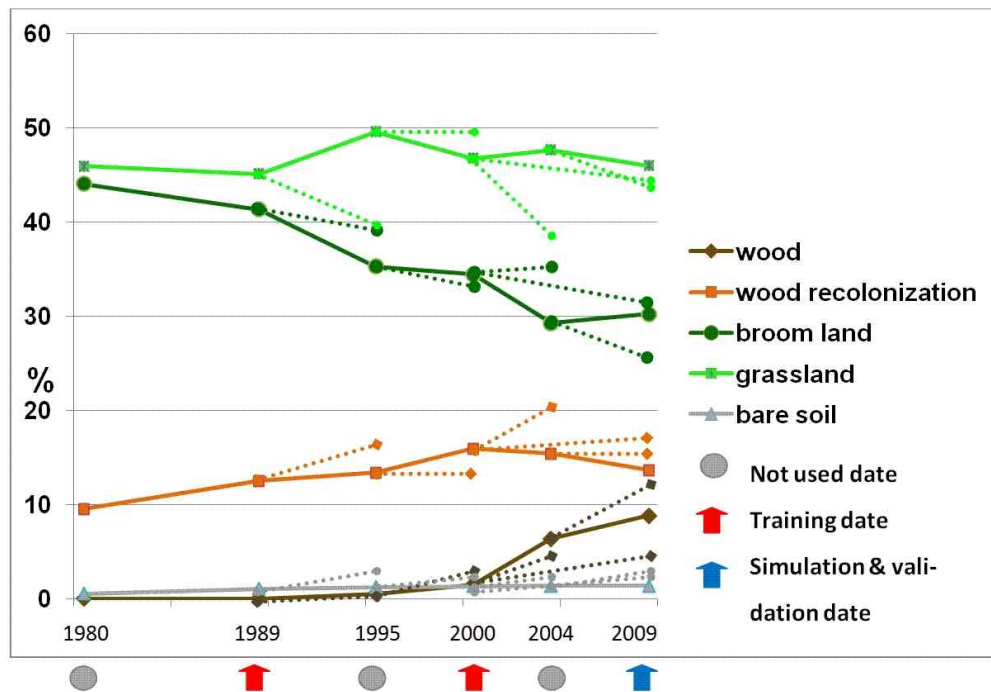
Table 4

	Wood	Wood re-colonization	Broom land	Grassland	Bare soil	Markov chain dates	Half of sum of abs. differences
1980	0.00	9.53	44.03	45.92	0.52		
1989	0.00	12.49	41.34	45.09	1.08		
1995	0.52	13.40	35.26	49.58	1.23		
<i>1995</i>	<i>0.51</i>	<i>16.90</i>	<i>38.94</i>	<i>39.91</i>	<i>3.73</i>	<i>80+89->95</i>	<i>9.68</i>
2000	1.50	15.95	34.45	46.76	1.34		
<i>2000</i>	<i>2.85</i>	<i>12.86</i>	<i>32.85</i>	<i>49.13</i>	<i>2.31</i>	<i>89+95->00</i>	<i>4.69</i>
2004	6.36	15.41	29.25	47.63	1.34		
<i>2004</i>	<i>4.48</i>	<i>20.29</i>	<i>35.18</i>	<i>38.21</i>	<i>1.84</i>	<i>95+00->04</i>	<i>11.31</i>
2009	8.81	13.64	30.25	45.96	1.34		
<i>2009</i>	<i>12.70</i>	<i>15.24</i>	<i>25.91</i>	<i>42.89</i>	<i>3.26</i>	<i>00+04->09</i>	<i>7.40</i>
<i>2009</i>	<i>4.97</i>	<i>17.73</i>	<i>31.77</i>	<i>43.26</i>	<i>2.27</i>	<i>89+00->09</i>	<i>6.54</i>

Observed and Markov chain simulated (italic) LUC (%). The last column indicates the absolute differences between the observed and Markov chain predicted land cover quantities.

45 Figure 10 expresses the differences between the observed (continuous lines) and Markov predicted LUCs (dotted lines). This graphic representation for each category shows that Markov chain is continuing the past trend. Otherwise, the figure illustrates that choosing a larger training period (e.g., 1989 – 2000 in comparison to 2000 – 2004) allows the modeler to focus more on general trends and ignore certain short-term variations. Therefore, the curves for broom land and grassland are a demonstrative example. The Markov matrix from 2000 and 2004 produces results that are far from reality. However, the Markov matrix from 1989 and 2000 resulted in a predicted area that is close to the observed one in 2009, which fit the more general trend.

Figure 10



Observed land cover (in % of total area) from 1980 to 2009. Dotted lines show land cover proportions that were estimated by the Markov chain (cf. table 4).

Discussion

46 For this dataset, applying various validation and error analysis techniques to map comparisons produces contrasting results. Although all models use the same Markov transition matrix, only quantitative validation (cf. table 2, left) shows differences in the fit with the observed LUC in 2009. As previously mentioned, the Markov transition matrix use differs in each modelling tool. The CA_Markov model uses predicted areas directly and only precedes the plausible allocation of predicted quantities. In contrast, LCM and Dinamica Ego model transitions and allow the inhibition or selection of specific changes. The authors avoided improbable changes. Otherwise, Mas *et al.* (2011) showed that the Markov algorithms in the considered models slightly differ. The most common map comparison technique, a pixel by pixel comparison between the simulated and observed LUC, provides a higher prediction score in the CA_Markov model than in the LCM and Dinamica models. This result occurs because persistence is the dominant process. The CA_Markov models LUC. The two other tools are especially designed to handle changes. Furthermore, these results correspond with the predicted amount of change and its swap component, in which both are substantially lower in the CA_Markov model (figure 3). Applied to this dataset, the CA_Markov model is a more conservative modelling tool in that the model is closer to the null hypothesis, which is also noted in other studies (Paegelow and Camacho, 2005, Mas *et al.* 2011). Figure 5 confirms this fact and shows that the CA_Markov model performs a larger fit with reality for persistence (category E) than the LCM and Dinamica models and performs the lowest fit for the correct predicted change (category B). Regarding the realism of the predicted spatial pattern (figure 6), we cannot note any evident advantage for any model. However, the Dinamica model expressly manages this aspect. Regarding the magnitude of error, all considered model tools perform LUC maps which categorical errors that are more than 90 percent minimal that means a class distance of 1. The allocation errors, as measured by distance between the false predicted and closest observed LUCs are more dependent on the LUC categories than on the chosen model.

47 Combining the model outputs (figure 7) provides complementary information. First, the large score of correct predicted LUC by intersecting the three predicted maps with Boolean and operator illustrates a base similarity of the applied models. This model output congruence may be useful for evaluating the predicted LUCCs for later and unknown dates. Secondly,

figure 7 illustrates the larger proximity of the LCM model and the Dinamica model relative to the CA_Markov model regarding the additional prediction gain from combining individual predicted LUCs by pairs. Furthermore, this figure underlines the previous observation that the CA_Markov model obtains a larger prediction score than the LCM and Dinamica models for persistence whereas LCM and Dinamica are modelling changes closer to reality as CA_Markov. Considering this result, a combination of the highest obtained LUCC component scores (i.e., persistence and change) may improve the model accuracy. The insertion of correct predicted changes by LCM into the CA_Markov prediction map which is most accurate overall and also for persistence doubles the figure of merit. However the resulting overall accuracy score is just above the baseline hypothesis which means no change. The comparison of the observed and modeled LUCCs and the analysis of their components, such as persistence, change and figure of merit, permits the characterization of the model behavior and allows the model outputs to be combined to improve the overall prediction score.

48 The relative importance of the presented validation techniques also depends on the objective of the model. If the model is used for prediction, the accuracy of the estimated amount of change is just as important as its allocation. In contrast, if the simulation aims to make prospective LUCC models by designing various scenarios, the modeler generally implements quantitative objectives regarding the expected LUC area or changes. Therefore, LUCC analysis, spatial pattern validation techniques and error analysis are more appropriate. Furthermore, the presented map comparison techniques, especially the figure of merit when computed for outputs of various models, provide useful information regarding the model performances concerning persistence and change, change components, such as the net change and swap and the realism of landscape, and allows the modeler to choose the most appropriate model based on the project objectives.

49 The results and interpretations presented here are specific to this case study and large series of tests with varying scales, types and speeds of LUCCs are necessary to reach more general considerations.

50 The LUCC model independent tests with various training dates used for Markov chains show that quantity agreement depends on the choice of these dates. This finding indicates the importance of disposing key dates because the Markov chain is strongly dependant on previous trends. Disposing of few LUC dates increases random chance because Markov chain determines the overall model accuracy. If the number or the quality of LUC dates does not allow tracing past trends or if past trends are not significant for future evolution, it is advisable to support the trend based simulation, called also baseline scenario, by various scenarios deliberately breaking with Markovian conditional transitions calculated on a basis too incomplete or becoming obsolete. By varying quantitative assumptions, this therefore geopropective model (Houet and Gourmelin, 2014) implements the allocation of these hypotheses and designs plausible futures.

Conclusions and outlook

51 The specificity of any LUCC model tool does not allow for strict comparisons. However, the three tested tools represent commonly used LUCC modelling approaches that use past dynamics and Markov matrices to simulate LUCC and show a high congruence of correct predicted LUC. The only choice of LUC dates for calculating conditional transition probabilities may change the simulation map. Thus, we showed that the most recent data are not always the most appropriate data. In our case, the LUCC depends on the unknown fire management schedule that is designed to stop spontaneous recolonization and maintain pastureland value. While most popular GIS embedded modelling tools only use two LUC maps for their simulation, a longer time series analysis appears necessary to estimate the degree of significance of the chosen training dates.

52 In addition, this paper shows that prediction maps vary even when the same data set is used in various modelling tools. *Per se* no model provides more accurate results than another. The set of validation and error analysis techniques emphasizes the diverse specific prediction rankings. Consequently, the model choice depends on the modelling objectives. For example, do we

want to predict LUC or are we interested by only changes? Do we focus on error analysis to better understand complex environmental dynamics? Once these questions are answered, do we consider the more important quantity, allocation accuracy or shapes of the landscape close to reality? Answering these questions is even more important than prediction errors are high. In this study, model accuracy is always lower than in the null hypothesis.

53 A perspective that should be thoroughly studied concerns the statistical level of the modeled LUC. In this study LUC is expressed by qualitative categories and each of them includes heterogeneity. As previously mentioned, different LUC categories do not have the same weight for future changes and land potential in this case study. Although bare soil, grassland and broom land allow pasture use, once the scattered trees exceed a critical threshold density that prohibits fire management, the land will be lost to pasture and become wooded. Consequently, focusing on wooded land, handled as quantitative variable (as a percent of covered land), may produce different results. This means not only paying attention to a particular LUC but also eliminates model unknown intra categorical LUC variation and helps to minimize data uncertainty.

54 “Despite its apparently scientific nature, modelling is a matter of judgement” (Abdou *et al.* 2012). Even if highlighted results are specific and need to be confirmed by applying a large range of datasets that represent various spatial and time scales, data origins and dynamics, the use of complementary map comparison techniques may help inform users regarding model choice so that they can select the most accurate software bricks to combine them.

Bibliographie

Abdou, M., Hamill, L., Gilbert, N., 2012, “Designing and Building an Agent-Based Model”. In: Heppenstall et al. (eds) *Agent-Based Models of Geographical Systems*. Springer Verlag, pp 141-166.

Ahlqvist, O., 200, “Extending post-classification change detection using semantic similarity metrics to overcome class heterogeneity: A study of 1992 and 2001 US National Land Cover Database changes”. *Remote Sensing of Environment* 112(3): 1226-1241

Botequilha, A., Miller, J., Ahern, J., McGarigal, K., 2006, *Measuring Landscapes. A planner’s handbook*. Washington: Island Press.

Camacho Olmedo, M.T., Paegelow, M. and Mas, J.F., 2013. “Interest in intermediate soft-classified maps in land change model validation: suitability versus transition potential”. *International Journal of Geographical Information Science*, 27 (12), 2343-2361.

Coquillard, P. and Hill, D.R.C., 1997, *Modélisation et simulation d’écosystèmes. Des modèles déterministes aux simulations à événements discrets*. Paris: Masson.

Crooks, A.T. and Heppenstall, A.J., 2012, “Introduction to Agent-Based Modelling”. In: Heppenstall et al. (eds). *Agent-Based Models of Geographical Systems*. Springer Verlag, 85-108.

Eastman, J.R., Jin, W., Kyem, P.A.K., Toledano, J., 1995, “Raster procedures for multi-criteria/multi-objective decisions”. *Photogrammetric Engineering and Remote Sensing*, 61(5), 539–547.

Eastman, J.R., 2009, Idrisi Taiga, Guide to GIS and Image Processing, manual version 16.02, Clark University, 342 p

Forman, R.T.T., 1995, *Land Mosaics: The Ecology of Landscapes and Regions*. Cambridge: EEUU.

Gaucherel, C., 2007, “Multiscale heterogeneity map and associated scaling profile for landscape analysis”. *Landscape and Urban Planning*, 82 (3): 95-102 (2007).

Gaucherel, C., Alleaume, S. and Hély, C., 2008, “The Comparison Map Profile method: a strategy for multiscale comparison of quantitative and qualitative images”. *IEEE Transactions on Geoscience and Remote Sensing*, 46 (9): 2708-2719, 2008

Gómez Delgado, M. and Barredo, J.I., 2005, *Sistemas de Información Geográfica y evaluación multicriterio en la ordenación del territorio (GIS and multicriteria evaluation for urban and regional planning)*. Madrid: Ra-Ma.

Gómez Delgado, M. and Tarantola, S., 2006, “Global sensitivity analysis, GIS and multi-criteria evaluation for a sustainable planning of hazardous waste disposal site in Spain”. *International Journal of Geographical Information Science*, 20, 449-466.

- Hagen, A., 2003, "Fuzzy set approach to assessing similarity of categorical maps". *International Journal of Geographical Information Science*, 17(3), 235–249.
- Hagen-Zanker, A., Straatman, B., Uljee, I., 2005. "Further developments of a fuzzy set map comparison approach". *International Journal of Geographical Information Science*, 19, 769-785.
- Houet, T. and Hubert-Moy, L., 2006. "Modelling and projecting land-use and land-cover changes with a cellular automaton considering landscape trajectories : an improvement for simulation of plausible future states". *EARSeL eProceedings*, n°5, pp. 63-76
- Houet, T. and Gourmelon, F., (2014), "La géoprospective – Apport de la dimension spatiale aux approches prospectives". *Cybergeo*, 8 p., <http://cybergeo.revues.org/26194>
- Jackson, D.A., 1995, "PROTEST: a procrustean randomization test of community environment concordance". *Ecoscience*, 2, 297-303.
- Jokar Arsanjani, J., 2012, *Dynamic Land-Use/Cover Change Simulation: Geosimulation and Multi Agent-Based Modelling*, Springer Theses, Springer Verlag.
- McGarigal, K. and Marks, B.J., 1995. *FRAGSTATS: Spatial pattern analysis program for Quantifying Landscape Structure*. USDA For. Serv. Gen. Tech. Rep.
- Kuhnert, M., Voinov, A., Seppelt, R., 2005. "Comparing raster map comparison algorithms for spatial modelling and analysis". *Photogrammetric Engineering & Remote Sensing*, vol. 71 (8), 975-984.
- Klug, W., Graziani, G., Grippa, G., Pierce, D., Tassone, C. (eds), 1992, *Evaluation of long range atmospheric transport models using environmental radioactivity data from the Chernobyl accident: the ATMES Report*. London: Elsevier.
- Mas, J.F., Kolb, M., Houet, T., Paegelow, M., Camacho Olmedo, M.T., 2011, "Eclairer le choix de modèles de simulation des changements des modes d'occupation et d'usages des sols" . *Revue Internationale de Géomatique*, v21, n° 3, pp. 405-430.
- Mas J.F., Kolb M., Paegelow M., Camacho Olmedo M.T., Houet T., 2014, "Inductive pattern-based land use/cover change models: A comparison of four software packages". *Environmental Modelling & Software*, v 51 January 2014 P.94-111
- Paegelow, M. and Camacho Olmedo, M.T., 2005, "Possibilities and limits of prospective GIS land cover modelling - a compared case study: Garrotxes (France) and Alta Alpujarra Granadina (Spain)". *International Journal of Geographical Information Science*, 19(6), 697-722.
- Paegelow, M. and Camacho Olmedo, M.T. (eds), 2008, *Modelling Environmental Dynamics. Advances in geomatic solutions*. Springer, series: Environmental Science and Engineering.
- Paegelow M., Camacho Olmedo M.T., 2010, "Modelos de simulación espacio-temporal y Teledetección: el método de la segmentación para la cartografía cronológica de usos del suelo". *Serie Geográfica – Universidad de Alcalá* n° 16; p.19-34
- Paegelow M., Camacho, Olmedo M.T., Mas J.F., Houet T., Pontius Jr. R.G., 2013, "Land change modelling: moving beyond projections", *International Journal of Geographical Information Science*, 27:9, 1691-1695
- Perica, S. and Foufoula-Georgiou, E., 1996, "Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions". *Journal of Geophysical Research*, 101 (D21), 26347-26361.
- Pérez-Vega, A., Mas, J.F., Ligmann-Zielinska, A., 2012, "Comparing two approaches to land use/cover change modelling and their implications for the assessment of biodiversity loss in a deciduous tropical forest", *Environmental Modelling and Software*, 29(1), 11-23
- Pontius Jr., R.G., 2000, "Quantification error versus location error in comparison of categorical maps". *Photogrammetric Engineering & Remote Sensing*, 66 (8), 1011-1016.
- Pontius Jr., R.G., 2002, "Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions". *Photogrammetric Engineering & Remote Sensing*, 68,1041–1049.
- Pontius Jr., R.G. and Millones, M., 2011, "Death to kappa and to some of my previous work: a better alternative". *International Journal of Remote Sensing*, 32 (15), 4407-4429.
- Pontius Jr., R.G. and Schneider, L.C., 2001, "Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA", *Agriculture, Ecosystems & Environment*, 85, 239–248.
- Pontius Jr., R.G., Huffaker, D., Denman, K., 2004a, „Useful techniques of validation for spatially explicit land-change models". *Ecological Modelling*, 179 (4), 445-461.

- Pontius Jr., R.G., Shusas, E. and McEachern, M., 2004b. "Detecting important categorical land changes while accounting for persistence". *Agriculture, Ecosystems & Environment*, 101, 251–268.
- Pontius Jr., R.G. and Petrova, S.H., 2010, "Assessing a predictive model on land change using uncertain data". *Environmental Modelling & Software*, 25, 299-309
- Pontius Jr, R.G., Boersma, W., Castella, J.C., Clarke, K., de Nijs, T., Dietzel, C., Duan, Z., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C.D., McConnell, W., Sood, A.M., Pijankowski, B., Pidhadia, S., Sweeney, S., Trung, T.N., Veldkamp, A.T., Verburg, P.H., 2008, "Comparing the input, output, and validation maps for several models of land change". *Annals of Regional Science*, 42 (1), 11-27.
- Rodrigues, H.O., Soares Filho, B.S., de Souza Costa, W.L., 2007, "Dinamica EGO, uma plataforma para modelagem de sistemas ambientais". *Anais XIII Simposio Brasileiro de Sensoriamento Remoto*, INPE, 3089-3096.
- Rykiel, E.J.J., 1996, "Testing ecological models: the meaning of validation". *Ecological Modelling*, 90, 229-244.
- Shirley, L.J. and Battaglia, L.L., 2008, "Projecting fine resolution land cover dynamics for a rapidly changing terrestrial-aquatic transition in Terrebone basin, Louisiana, U.S.A." *Journal of Coastal Research*, 246, 1545-1554.
- Soares-Filho B.S.; Pennachin, C. L.; Cerqueira, G., 2002, "DINAMICA – a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier". *Ecological Modelling*, 154(3): 217 – 235
- Torrens, P.M., 2011, "Calibrating and validating cellular automata models of urbanization". In: Yang, X., ed. *Urban Remote Sensing. Monitoring, Synthesis and Modelling in the Urban Environment*. John Wiley & Sons, Ltd, 335-345.
- Turner, M.G., Constanza, R., Sklar, F.H., 1989, "Methods to evaluate the performance of spatial simulation models". *Ecological Modelling*, 48, 1-18.
- Villa, N., Paegelow, M., Camacho Olmedo, M.T., Cornez, L., Ferraty, F., Ferré, L., Sarda, P., 2006, "Various approaches for predicting land cover in Mediterranean mountains". *Communication in Statistics*, v36 (1) 73-86.
- White, R., Engelen, G., Uljee, I., 1997, "The use of constrained cellular automata for high resolution modelling of urban land use dynamics". *Environment and Planning, B: Planning and Design*, 24, 323-343.

Pour citer cet article

Référence électronique

Martin Paegelow, María Teresa Camacho Olmedo, Jean-François Mas et Thomas Houet, « Benchmarking of LUCC modelling tools by various validation techniques and error analysis », *Cybergeo : European Journal of Geography* [En ligne], Systèmes, Modélisation, Géostatistiques, document 701, mis en ligne le 22 décembre 2014, consulté le 06 janvier 2015. URL : <http://cybergeo.revues.org/26610> ; DOI : 10.4000/cybergeo.26610

À propos des auteurs

Martin Paegelow

GEODE UMR 5602 CNRS Université de Toulouse
Jean Jaurès, 5 allées Antonio Machado 31058 Toulouse Cedex
paegelow@univ-tlse2.fr

María Teresa Camacho Olmedo

Universidad de Granada,
Campus de la Cartuja, 18077 Granada, Spain
camacho@ugr.es

Jean-François Mas

Centro de Investigaciones en Geografía Ambiental,
Universidad Nacional Autónoma de México (UNAM),
Antigua Carretera a Pátzcuaro No. 8701, Col. Ex-Hacienda de San José de La Huerta, C.P. 58190
Morelia Michoacán MÉXICO
jfmasc@ciga.unam.mx

Thomas Houet

GEODE UMR 5602 CNRS Université de Toulouse
Jean Jaurès, 5 allées Antonio Machado 31058 Toulouse Cedex
thomas.houet@univ-tlse2.fr

Droits d'auteur

© CNRS-UMR Géographie-cités 8504

Résumés

This study focuses on various validation and error analysis techniques that are based on map comparisons. After a literature review, authors apply these techniques to analyze the accuracy of LUCC models in terms of quantity, pixel by pixel correctness and LUCC components such as persistence and change. In addition, the fidelity of the spatial patterns and the congruency of the simulation maps from different modelling tools are tested. Finally, an error analysis is conducted that focuses on the magnitude of allocation errors and the magnitude of errors in predicted land use / cover classes. In addition, the impact of training dates on Markov chain predicted LUCC is analyzed.

Mentioned techniques of validation and error analysis are illustrated by modelling LUCC of a small study area in the Eastern Pyrenees (France), where current LUCC are driven by spontaneous reforestation, decreasing pastureland and minimal anthropogenic disturbance. This very simple data set is used with three different tools (CA-Markov, LCM and Dinamica Ego) that represent commonly used modelling approaches and their methodological characteristics are highlighted. Applied to this specific dataset, contrasting results occur for different software programs that can help users choose an appropriate modelling approach according to specific model objectives.

Cette étude montre plusieurs techniques de validation et d'analyse d'erreurs basées sur la comparaison de cartes. Après une revue de la littérature, les auteurs appliquent ces techniques pour analyser l'exactitude de simulations LUCC en termes de quantité, de comparaison pixel par pixel et de composantes LUCC telles que la persistance et le changement. En outre la fidélité des patrons spatiaux et le degré de congruence des cartes simulées par différents logiciels sont testés. Une analyse des erreurs focalise sur la magnitude des erreurs en termes de localisation et d'écart aux catégories d'occupation et d'usage du sol. Enfin, l'impact des dates d'entraînement sur la quantité des changements prédits par chaînes de Markov est analysé.

Ces techniques de validation et d'analyse d'erreur sont illustrées par la modélisation des changements d'occupation et d'usage du sol d'un site d'études de petite taille dans les Pyrénées Orientales (France) où les changements actuels sont régis par la reforestation spontanée, la diminution du pastoralisme et un impact anthropique minimal. Ce jeu de données alimente trois outils de modélisation (CA-Markov, LCM et Dinamica Ego) qui représentent des approches de modélisation couramment utilisées et dont les caractéristiques méthodologiques sont signalées. Pour ce jeu de données spécifique, des résultats contrastants sont observés pour les différents outils. Ceux-ci peuvent aider le modélisateur dans le choix d'un outil de modélisation appropriée en fonction des objectifs de modélisation.

Entrées d'index

Mots-clés : analyse d'erreur, LUCC, modélisation, simulation, validation

Keywords : error analysis, LUCC, modelling, simulation assessment, validation