Chapter 4
Techniques for the Validation of LUCC Modeling Outputs

M. Paegelow, M.T. Camacho Olmedo and J.F. Mas

Abstract Validation is the third stage in the modeling process, after calibration and simulation, and also applies to scenarios. It is an essential part of the process in that the credibility of a model depends on the accuracy of its output. A large range of validation approaches and tools exist, many of which can also be used during the calibration stage. In this chapter we distinguish between purely quantitative validation techniques and those that also consider the spatial allocation of simulated land use/cover changes (LUCC). According to model outputs and objectives, simulation maps can be either hard-classified or soft-classified. While some validation techniques apply to both types of map (cross tabulation matrices and indices, congruence of model outputs), others are specific to one. Techniques such as LUCC indicators, feature and pattern recognition and error analysis are used to validate hard-classified simulation maps, while ROC is used to test soft-classified maps. We then look at a second validation approach based on LUCC dynamics such as LUCC components, intensity analysis, data uncertainty and the impact of spatial and temporal scales. Finally, we compare a group of the most common model software programs (those used by the contributors to parts II and III of this book), in order to list their validation capabilities.

Keywords Validation · Error analysis · Land use and cover changes · Modeling · Simulation assessment

---

M. Paegelow (✉)
Département de Géographie, Aménagement et Environnement,
Université de Toulouse Jean Jaurès, Toulouse, France
e-mail: paegelow@univ-tlse2.fr

M.T. Camacho Olmedo
Departamento de Análisis Geográfico Regional y Geografía Física,
Universidad de Granada, Granada, Spain
e-mail: camacho@ugr.es

J.F. Mas
Centro de Investigaciones en Geografía Ambiental, Universidad Nacional
Autónoma de México (UNAM), Morelia, Michoacán, Mexico
e-mail: jfmas@ciga.unam.mx

M.T. Camacho Olmedo et al. (eds.), *Geomatic Approaches for Modeling Land Change Scenarios*, Lecture Notes in Geoinformation and Cartography,
https://doi.org/10.1007/978-3-319-60801-3_4

1 Introduction

"Despite its apparently scientific nature, modeling is a matter of judgement" (Abdou et al. 2012). "However, the validity of a model should not be thought of as a binary event… model can have a certain degree of validity" (Croks and Heppenstall 2012). "Until more guidance is provided in the literature, calibration and validation will remain a key challenge" (Ngo and See 2012). Rykiel (1996) noted that there is no universal agreement as to how to evaluate the goodness-of-fit of validation. "Depending on their position on this spectrum, models may have different calibration and validation requirements… Models can be calibrated with vast quantities of detailed data, and using sophisticated procedures. They can be validated for historical time periods with high degrees of success. However, a model is only as good as the rules that drive its behavior. Good rules require good theory" (Torrens 2011). Spatial models cannot be validated in a rigorous way (Oreskes et al. 1994).

These quotations from the literature give us some idea both of the difficulty of designing a model that closely reflects future reality and the ambiguity or debate as to what validation actually means. Model validation becomes crucial in a world that produces an ever-increasing number of simulations and scenarios over a large thematic range. In order to give credit to the output of a model, we need information about its robustness and accuracy.

1.1 What Is Validated in Land Change Models?

In this chapter we begin by outlining that the validation techniques discussed here focus on path-dependent models, although there are others that are not path-dependent. Also known as the SAS (story and simulation) approach (Alcamo 2008), these models try to render contrasted, spatially explicit scenarios defined by experts or in a participatory manner: narratives which are then translated into quantitative scenarios (Houet et al. 2016). For their part, the path-dependent models produce scenarios known as trend scenarios or BAU (business as usual) scenarios.

Over the last decade, there has been an important and increasing interest in the validation of simulation models that predict changes over time, particularly from an initial time in the past to a subsequent time in the future (Pontius and Petrova 2010), with a focus on land use and cover change (LUCC), often simplified as land change (Jansen and Veldkamp 2011).

A model's credibility depends on its validation, and this general concept includes three stages, which have been widely endorsed: Verification, Calibration and Validation (Coquillard and Hill 1997; Torrens 2011; Croks and Heppenstall 2012; Ngo and See 2012). Verification refers to the entire process of certifying the correct internal operation of a model (including Face Validation and Sensitivity Analysis); during calibration (see Chap. 2 about calibration), the model is tested using several specific parameters and context-like training periods or dates; while validation involves evaluating the accuracy of the results produced by the model during the simulation stage (see Chap. 3 about simulation), including scenarios (see Chap. 5 about scenarios). North and Macal (2007) also state that "Verification is the process of making sure that an implemented model matches its design, validation is the process of making sure than an implemented model matches the real-world" (cited by Croks and Heppenstall 2012).

Calibration and validation are individually and separately defined, and the period used for calibration purposes may be different from or unknown in the validation period. While the first step refers to a date (t1) and/or a period prior to it (t0−t1), the second step is focused on simulations after t1, which is the point in time for which the predictive extrapolation with the horizon T (t1–T) begins. Pontius and Malanson (2005) highlight this difference, referring specifically to the confusion detected in several studies regarding the goodness-of-fit of the calibration stage for quantifying the predictive power of a model rather than using the goodness-of-fit of the validation stage. In fact, a good fit for calibration does not necessarily imply a good fit for validation or that the latter is an appropriate indicator of a model's predictive power (Pontius and Pacheco 2004). Following White et al. (2012), the time periods for calibration and validation must be sufficiently long to minimize the impact of unrepresentative details during the training period. Calibration and validation over short time periods are notoriously unreliable. Even an empirically excellent calibration may be fundamentally in error either because over-calibration tunes the model to idiosyncratic details of the particular data set or more fundamentally because the data set may be unrepresentative of the range of possibilities present in the system being modeled (Brown et al. 2005; Engelen and White 2007).

1.2 How to Validate Land Change Models?

Modeling land use/land cover changes (LUCC) can help us understand complex social and ecological interactions and provides useful information for decision-makers such as planners (Paegelow et al. 2013). The usefulness of LUCC models can be measured by the accuracy of their output.

According to Torrens (2011), validation evaluates the correctness of a model while Croks and Heppenstall (2012) described it this way: "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". Coquillard and Hill (1997) proposed that model validation should consist of three progressive steps: verification, (Does the model run correctly?), calibration (Does the model correctly simulate a known state?) and validation (Does the model correctly predict an unknown state?). "To improve the robustness and the acceptance of a model, the data at the validation date must be model unknown, in other words data that has not been used in the building and calibration of the model"

(Paegelow and Camacho Olmedo 2008). If not, simulation must be considered as a step in the calibration process.

Rykiel (1996) distinguishes between “conceptual” and “operational” validation. Conceptual validation warrants that the assumptions underlying the conceptual model are correct or plausible. Operational validations measure the accuracy of model output. When modeling the future, the model can be partially validated by comparing the results with expert knowledge, by assessing its robustness by measuring the constancy of model outputs during iterative model runs. A complementary technique is gauging the degree of congruence between the outputs of different software programs that use the same data set and parameters. Gómez Delgado and Tarantola (2006) tested model stability using sensitivity analysis. To this end they developed several indices to measure the variability of model outputs when input parameters are changed. In this context, Gomez Delgado and Barredo (2005) describe techniques to assess risk when using model outputs and Jokar Arsanjani (2012) focuses on model data and drivers of uncertainty.

There is a large range of statistical tools for measuring the accuracy of hard and soft predictions. Hard predictions can be validated by comparing between simulated and observed LUCs. However, a soft prediction is evaluated by comparing potential changes or LUC suitability with observed LUC or LUCC. This is often done by measuring the area under the ROC (Relative Operating Characteristics) curve (Pontius and Schneider 2001). Eastman et al. (2005) and Pérez-Vega et al. (2012) focused on the potential for change. With this in mind, they compared dynamic areas relative to persistent ones and developed a measure called DiP (Difference in Change Potential). Of the two forms of model output—hard or soft prediction—the validation of hard maps is more common and there is a larger spectrum of statistical tools. These tools focus on different aspects: accuracy of quantity and allocation, correctness of LUCC components, similarity of the landscape pattern, model congruence and error analysis.

As regards quantitative agreement, modelers distinguish between matching the sum total of the LUC area and the pixel-by-pixel comparison, which also evaluates matching in allocation (Torrens 2011). As a first step, an overall agreement may be obtained by calculating statistical indices, such as Chi-square or Kappa (Pontius 2002). However, Pontius and Millones (2011) indicate that the KIA (Kappa Index of Agreement) is not suitable for LUCC model validation because it assumes randomness. The sample matrix must therefore be converted into an estimated population matrix. The Chi-square index has the same drawback, as pixels cannot be considered as independent observations. For map comparison we recommend easier indices such as quantity and allocation disagreement. Various validation techniques that consider changes have been developed. For example, Pontius (2000) and Pontius et al. (2004a, b, 2008) propose a technique that splits the LUCC-budget into gain, loss, net change and swap (see Technical Notes in Part IV of this book). Pontius et al. (2008) also developed several statistical LUCC indices for determining accuracy, including a figure of merit (see Technical Notes in Part IV of this book), a ratio between correct predicted changes and the sum total of observed and predicted changes.

Further validation techniques focus on fuzzy allocation agreement, with indices (Hagen 2003; Hagen-Zanker et al. 2005; Rodrigues et al. 2007) that measure the relative allocation agreement and overcome the limitations produced by exclusive cell state and exact allocation (see Technical Notes in Part IV of this book). In the same way Procrustes analysis (Jackson 1995) performs pixel-by-pixel comparison by linearly transforming one grid as rotation, translation or scaling to achieve the best fit with the reference grid. Furthermore, Kuhnert et al. (2005) describe algorithms that test the similarity of raster matrices by using different weights and by varying the window size.

Spatial analysis measurements consider the distribution and shapes of land patterns (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). In addition, error analysis highlights conceptual and model parameter inaccuracy by measuring errors in simulated LUC categories or transitions and their allocation (Pontius 2000; Pontius and Petrova 2010).

There are several studies that provide a comprehensive review of the validation techniques designed for spatial models (Turner et al. 1989; Pontius et al. 2004a; Paegelow and Camacho Olmedo 2008; Shirley and Battaglia 2008; Sargent 2009), while van Vliet et al. (2016) provide the results of a large study about calibration and validation techniques applied in recent land change modeling papers.

These few lines of introduction are intended to outline the importance of setting objectives for LUCC modeling. Do we care about the entire space or should we focus only on changing land? Do we want to achieve quantitative accuracy or a realistic landscape or urban pattern? Evaluating the accuracy of a model is clearly a matter of assessing its true purpose: do we want a model that makes predictions or one that presents a range of plausible futures?

In this chapter, we will be focusing on three aspects of validation. We will begin by presenting validation methods and tools according to model outputs and objectives (Fig. 1). Model outputs may be *hard* (maps with the same legend as training LUC maps), or *soft* (simulation maps expressing the potential of places to become a particular land cover or land use). Modeling objectives may be different: focusing on accuracy in terms of quantity, of allocation, of realistic landscape patterns. A second aspect is that validation depends on LUCC dynamics, as manifested in the intensity or rate of land change and also in the impact of the particular spatial and temporal scales used. Thirdly we describe validation according to LUCC models.[1] A presentation of selected software validation tools is completed with a table comparing them.

[1]See the short presentations in Part V of this book about (in alphabetical order) APoLUS, CA_MARKOV, CLUMondo, Dinamica EGO, Land Change Modeler (LCM), LucSim, Metronamica and SLEUTH. The authors are also grateful to all contributors who helped us understand the different software packages.

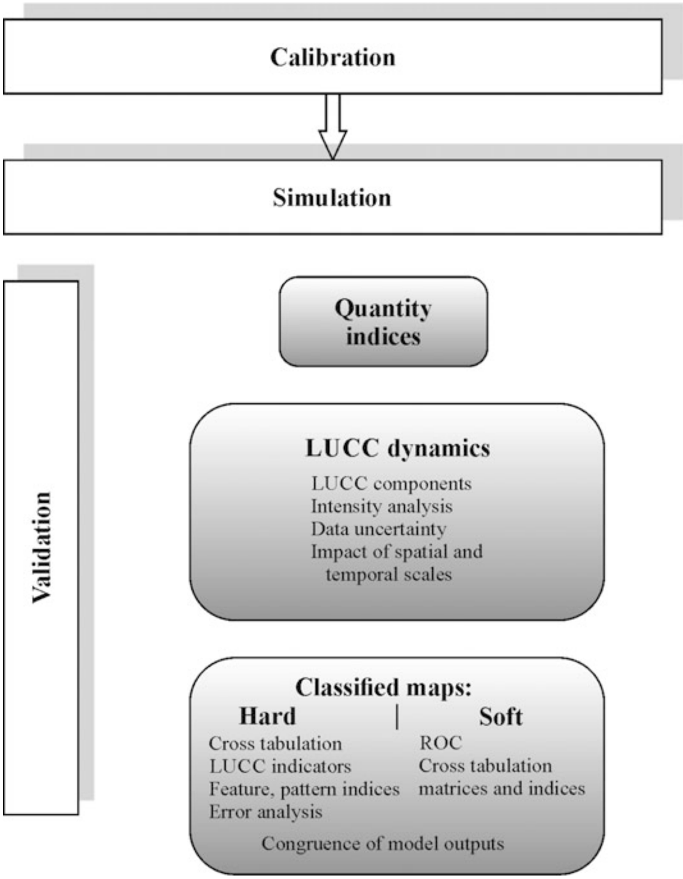

Fig. 1 General overview of validation techniques

2 Validation in Terms of Model Outputs and Objectives

Simulation Output Forms: Hard Versus Soft

As mentioned in the chapter in this book about simulation (see Chap. 3), model outputs can be split into two categories: hard outputs, in which each pixel in a raster map is assigned to exactly one category of land use or cover (LUC) (hard-classified map) and soft outputs, in which each pixel has a partial membership of several classes simultaneously (soft-classified maps). During the validation step, soft simulation results show the partial membership of a specific land use category or land transition and the level of membership indicates the degree of uncertainty. Most spatial land-change models focus on hard simulation results and their validation. In several cases, a quick reference to soft simulation is made, but only a few contributions focus exclusively on soft simulation results and their validation (Pérez-Vega et al. 2012; Wang and Mountrakis 2011; Conway and Wellen 2011).

2.1 Validation in Terms of Quantity Estimation

Modeling over time and space typically produces results about the quantity of land-use change (quantity) and where it takes place (allocation). Validation can focus on one or both of these output components (Fig. 2). Generally, both components are evaluated together. This is the domain of map comparison techniques using matrices to compute correct predictions as quantities correctly allocated. The spatial component does not only refer to prediction at the correct place. Validation focusing on allocation can also evaluate spatial shapes and patterns.

Evaluating only predicted quantities (cumulated area) without considering correct allocation is much easier than predicting the correct amount of land change at the correct place (Paegelow and Camacho Olmedo 2005; Paegelow et al. 2014).

The amount of expected land change may be predicted or given. The latter choice is made by "what happens if" scenarios that design a range of plausible futures. Quantitative prediction often uses a probabilistic approach such as Markov chains (see Technical Notes in part IV of this book). In this context, we will be specifically focusing on Markov chains and their implications on accurate

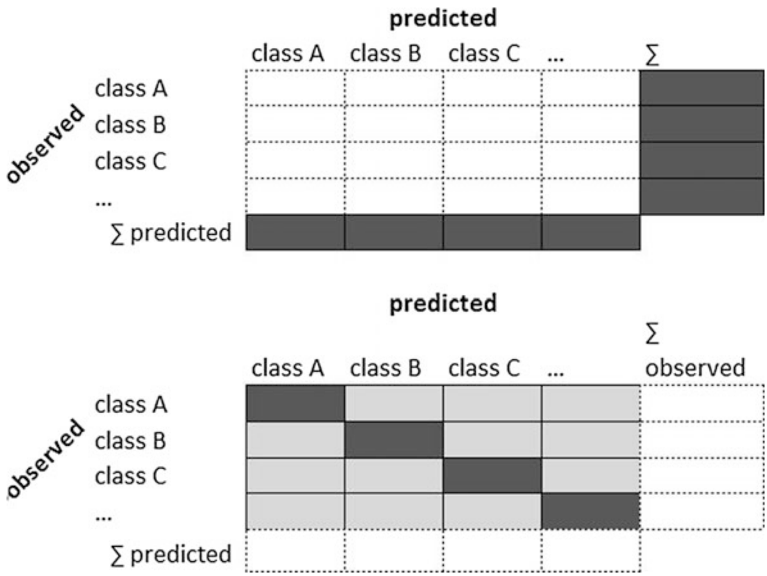

Fig. 2 Validation of cumulated surface (*above*) versus pixel-by-pixel matrix validation of quantity and allocation (*below*)

prediction. Two important aspects will therefore be analyzed: the impact of the software (Mas et al. 2011, 2014) and its algorithms and the assumed or specified level of confidence in training data.

When focusing exclusively on the quantitative aspect of model output, it is important to put the comparison between observed and simulated LUC at t_2 into perspective by also indicating former LUC quantities at t_1 (end of calibration—model known—period). This enables us to compare observed and modeled land change. We will come back to this point in more detail when discussing map comparison techniques in the next paragraph. As for integrating dynamics into quantitative validation, error analysis will be discussed further by taking into account the allocation aspect too.

2.2 *Hard Classified Maps*

The initial validation may be visual or qualitative (Torrens 2011), a more intuitive means of assessing the resemblance between model output and the validation data, e.g. simulated land use and observed land use. However, visual inspection only provides an initial impression and model accuracy has to be tested in other ways, generally statistically.

2.2.1 Pixel-by-Pixel Matrices and Comparison with the Null Model

For hard-classified maps, a full validation is the most common method, where comparisons between simulated and observed LUC referring to the same data are possible, i.e. both documents have the same nomenclature and temporal reference. The model's accuracy is evaluated by comparing simulated LUC with its reference image to a null, no change model (Pontius and Malanson 2005). In a relative minority of cases, researchers have compared different models or individual runs of the same model in different places and times (Pontius Jr. et al. 2008, cited by Torrens 2011). A large range of statistical tools may be used to assess the correctness of model output. The range of tools for comparing observed and simulated results or various different simulations, include the following pixel-matching techniques (performed on a pixel-by-pixel basis):

LUCC Indicators

Sohl et al. (2012) used this pixel-by-pixel technique (Fig. 3) to compare various LUCC scenarios by measuring the disagreement in quantity and allocation.

Prediction errors may be split into omission errors and commission errors for each class (Fig. 4). Omission refers to areas observed as change but not predicted as such. Commission error means the part of predicted change that, in fact, did not

Fig. 3 LUC matrix comparing observed and predicted LUC. Accurate prediction (*hits*) are located on the matrix diagonal (*dark cells*), errors in the rest of the matrix (*light cells*)

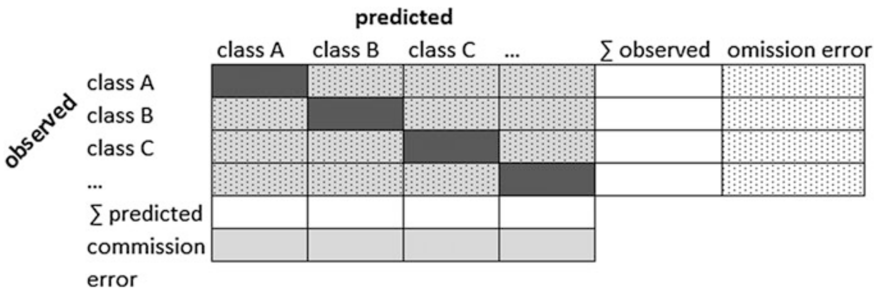

Fig. 4 Omission and commission errors

change. Commission is sometimes also referred to as consumer's accuracy and omission as producer's accuracy such as in cross tabulation techniques in remote sensing. In Fig. 4 omission is the total per line minus correct predicted (diagonal matrix cell), while commission is equal to the total per column minus correct predicted (diagonal matrix cell).

When introducing a third map into the comparison, e.g. observed LUC at the beginning of the simulation period (generally the last known date for the model is the end of the calibration period), it will be possible to compare observed and predicted change and to distinguish between hits (observed persistence or change predicted as such) and errors due to observed change predicted as persistence (omission), observed persistence predicted as change (commission) and observed change predicted as such, but with incorrect LUC categories.

Some software programs provide tools for cross validation between t_1 observed, t_2 observed and t_2 predicted by differentiating between 'Hits' (correctly predicted changes), 'misses' (omission errors) and 'false alarms' (commission errors).

These validation techniques rely on a technique of land change analysis. Pontius (2000) and Pontius et al. (2004a, b) established a comprehensive way of analyzing LUCC and measuring the accuracy of the model outputs based on LUC persistence and changes. They called this technique LUCC- budget (see the technical note about LUCC budget in Part IV of this book).

On the basis of previous research by Klug et al. (1992) and Perica and Foufoula-Georgiou (1996), Pontius et al. (2008) calculated various LUCC indices by splitting map comparison between the observed and predicted LUCs into percent correct and percent error distinguishing the following components:

A = Observed change predicted as persistence: error
B = Observed change predicted as such with correct LUC categories: correct
C = Observed change predicted as such but with incorrect LUC categories: error
D = Observed persistence predicted as change: error

These components allowed the following three derived measurements to be calculated:

- Figure of Merit—the ratio of B/(A + B + C + D) which expresses the overlap between observed and predicted change. This value ranges from 0 (no overlap) to 100% (perfect overlap).
- Producer's Accuracy—the ratio of B/(A + B + C) which 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—the ratio of B/(B + C + D) which expresses the part of the pixels accurately predicted as change compared to all model-predicted changes.

2.2.2 Disagreement Indices Based on Cross Tabulation

Krüger and Lakes (2015) present an innovative method for quantifying disagreement between different simulations using cross-tabulation techniques applied to binary maps (e.g. deforestation or not). Their disagreement index also includes quantity as allocation matching and may be used for hard classified maps as continuous probability simulations. They started with a well-known cross-tabulation matrix (Hagen-Zanker 2009; Mas et al. 2013) as shown in Fig. 5. "The diagonal from upper left to lower right represents agreement while the diagonal from lower left to upper right represents disagreement" (Krüger and Lakes op. cit.). By considering soft-classified maps as original simulation output and following Pontius and Milliones (2011), Krüger and Lakes considered the two disagreement cells of

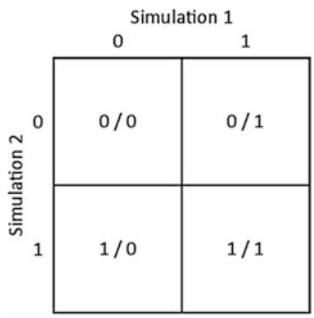

Fig. 5 Cross-tabulation between two binary simulation maps showing the four possible combinations

the matrix as a base to split the disagreement between different simulations into their quantity and allocation components. Their method allowed us to quantify the distance between two maps from the diagonal (perfect fit) in an orthogonal diagram whose two axes express quantity and allocation.

As the authors themselves make clear, their method is established for comparison between binary maps but can be extended to multi-categorical maps by splitting them into monothematic maps. However, we must bear in mind that when doing so, we lose the relations between LUC categories. This means for example that we cannot measure how wrong a simulation is by comparing simulated and observed LUC. Some errors could be considered more important than others, e.g. simulating woodland instead of shrubs could be a more important disagreement than simulating urban.

2.2.3 Fuzzy Logic Indices

There are various alternative techniques to hard pixel-by-pixel comparison. Indices based on fuzzy logic (Hagen 2003; Hagen-Zanker et al. 2005, 2009) (see Technical Notes in Part IV of this book) measure the agreement of location and overcome the limitations due to exclusive cell state and exact allocation. Some popular modeling software programs incorporate vicinity-based comparison tools measuring the fuzziness of location (Rodrigues et al. 2007), allowing a more gradual and flexible method than the classic cell-to-cell comparisons.

2.2.4 Procrustes Analysis

Jackson (1995) described the usefulness of Procrustes analysis. He compared the fit between different matrices by linear transformation (rotation, translation, scaling) of one grid to achieve the best fit with the reference grid. Pontius et al. (2004b) chose multiple resolutions to analyze the nature of allocation errors (cf. Sect. 2). More recently, Pontius et al. (2007) proposed a validation method that considered a nested stratification structure.

2.2.5 Feature and Pattern Recognition

Spatial analysis measurements take into account spatial pattern, its distribution and shapes (White el al. 1997). Many metrics were derived from landscape ecology such as shape, compactness, diversity and fragmentation (Forman 1995; McGarigal and Marks 1995; Botequilha et al. 2006). White et al. (2012) analyzed cluster size-frequency distributions. In addition to quantitative accuracy measurements, landscape pattern agreement offers a useful, supplementary validation approach. The simplest indicators are the size and shape of the patches. Dinamica EGO software allows us to model these parameters by average and standard deviation of patch size and the degree of compactness as a ratio between surface area and perimeter. Validation may be done by map comparison techniques that focus on the number, size and compactness of observed and simulated patches.

2.2.6 Error Analysis

Error analysis provides useful information about model logic and underlying conceptual approaches, so giving the modeler a better understanding of the model. In addition, the previously presented techniques can be completed by analyzing the possible origins of error. Seen from this point of view, LUCC analysis and Figure of Merit (see Technical Notes in part IV of this book) can be considered alongside validation techniques such as error analysis. Error analysis tries to answer the question 'how wrong is the prediction?' To do so, it generally focusses on two components: categorical or transitional errors and error in allocation.

LUC Category Errors

Various techniques measure disagreement between observed and simulated LUC. While quantitative data (e.g., percent of tree cover) enable us to measure the magnitude of inaccuracy, categorical data generally needs to be transformed into quantitative data or ordered on a scale before being analyzed. Ahlqvist (2008) offers a technique of fuzzy change estimation about the closeness between observed and simulated LUC categories. Paegelow et al. (2014) measured the magnitude of error between simulated and observed LUC expressed as categories. However, if LUC legends form a ranking order that reflects spontaneous vegetation succession from bare soil to woodland, land use intensity or other criteria that enable us to place LUC categories in an ordered scale, we can measure the parametric distance between observed and simulated LUC. Prediction error is measured by the absolute categorical distance between observed and simulated LUC. In many situations, modelers will probably have difficulties quantifying the exact distance between different LUC on an ordered scale. A possible coarse approach is to use equal distance between original categories. Paegelow et al. (2014) did so to rank LUC by the covering rate from bare soil to woodland.

Allocation Error

A large number of metrics can be calculated. Paegelow et al. (2014) created a distance map for each LUC category for which the considered LUC was the origin. The distance map was then crossed with simulation errors (omissions, commissions and prediction of false gaining categories). For each wrongly predicted patch of a given LUC category, these authors measured the minimum distance to the nearest correct location and then calculated the average for each LUC category.

2.2.7 Congruence of Model Outputs

Another form of validation consists of using the same data set to simulate LUCC with different models (Figs. 6 and 7). The closeness of the resulting simulation maps is measured and the degree of congruence is considered as an indicator of the stability of the model and the plausibility of the simulations (Paegelow et al. 2014). The same procedure also provides useful information about model robustness (Camacho Olmedo et al. 2015). Sohl et al. (2012) applied the same approach to multiple LUCC scenarios computed for the Great Plains in the United States, a procedure they described as "scenario diversity".

2.2.8 Other Approaches

Torrens (2011) proposed running models exhaustively (specifically in stochastic or probabilistic models). Several other authors use histograms (Conway and Wellen 2011) with several choices (equal weights, difference…) (Bone et al. 2011; Kamusoko et al. 2009), while Li et al. (2011) proposed a geographical simulation

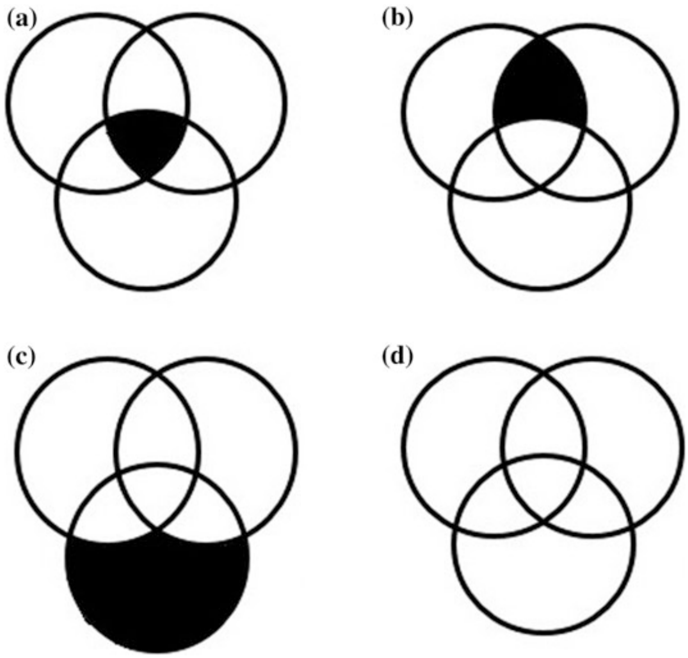

Fig. 6 Different congruence levels of simulation maps computed by three different models: **a** perfect intersection, which means total congruence of correctly predicted land use, **b** congruence of two models, **c** only one model gives correct prediction, **d** no model predicts correctly

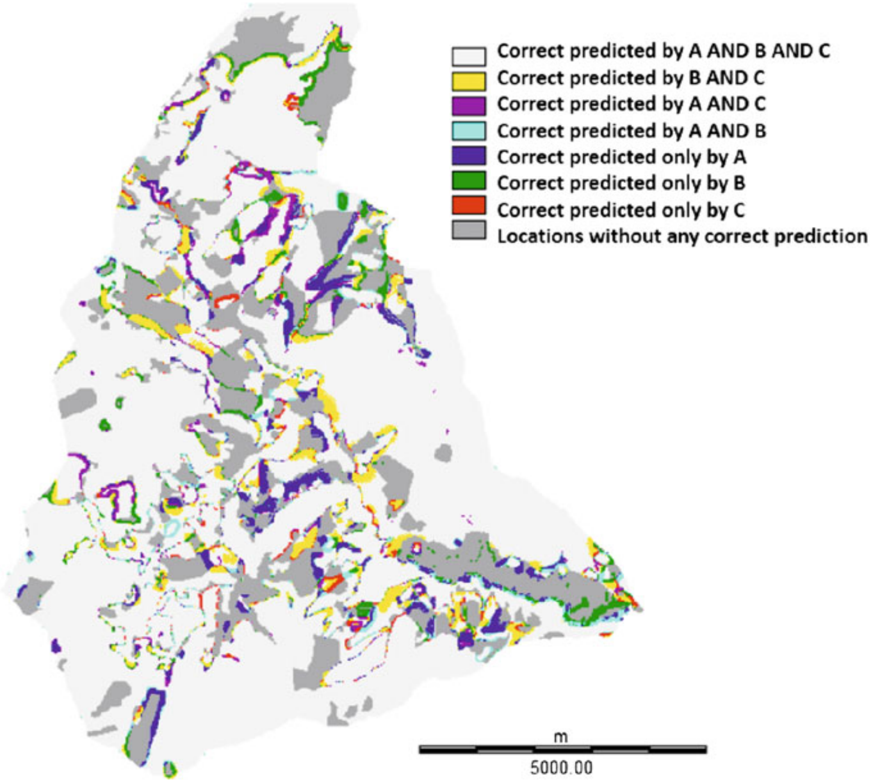

Fig. 7 Congruence of three simulations computed by *(A)* CA_MARKOV model, *(B)* Multi Layer Perceptron, *(C)* Statistical regression model, applied to Garrotxes catchment (Eastern Pyrenees, France)

and optimization system to model the reciprocal relationships between simulation and spatial optimization, including future simulations.

2.3 Soft-Classified Maps

2.3.1 Soft-Classified Maps

Of the various methods for assessing the accuracy of simulation maps, the first, most intuitive comparison method is usually visual or qualitative validation. This is also used in soft results (Torrens 2011) and in different types of superposition between soft-classified and real maps (observed and non-observed transition or land use) and in the analysis of frequency distributions (Yu et al. 2011; Alcamo et al. 2011; Camacho et al. 2013; Wang and Mountrakis 2011). Paegelow and Camacho

Olmedo (2005) compared the performance average and the standard deviation suitability scores for each candidate land cover with all of the other categories.

2.3.2 ROC

While hard prediction leads to cells being classified within one specific LUC category, some modeling programs provide soft prediction maps expressing the vulnerability of the land to change or suitability maps for each LUC category, which are computed by multi-criteria evaluation (MCE) (Eastman et al. 1995) (see Technical Notes in Part IV of this book).

In this context, Relative Operating Characteristic (ROC) (Hanley and McNeil 1982; Pontius and Schneider 2001) (see Technical Notes in Part IV of this book) is a measure of the spatial likelihood between a reference map and a suitability map. The reference is binary and shows the spatial distribution of a specific LUC category or transition, while the suitability map expresses the potential for this category or the propensity to change in the case of analyzing transitions. The procedure consists in ranking these suitability or vulnerability-to-change scores into *n* classes and computes the proportion of true (presence on reference map) and false (absence) positives. ROC assumes that the high scores in the comparison map are more likely to be truly positive. Pontius and Schneider (2001) provide a graphic illustration for this technique. Various other researchers have applied ROC in land change models (Wang and Mountrakis 2011; Alcamo et al. 2011; Lin et al. 2011; Jokar Arsanjani 2012; Ngo and See 2012), comparing different study areas (Paegelow and Camacho Olmedo 2005), calibration and validation periods (Conway and Wellen 2011) or different results after a number of drivers had been considered (Huang et al. 2012). Eastman et al. (2005) and Pérez-Vega et al. (2012) applied ROC and DiP to compare modeling approaches. Conway and Wellen (2011) compared ROC between the calibration and validation period. Pontius and Si (2013) introduced a variant of ROC: TOC—the Total Operating Characteristic, which enables the user to calculate the AUC, while also showing all the information in the contingency table for each threshold.

2.3.3 Cross Tabulation Matrices and Indices

This type of validation compares two or more types of soft-classified maps. All of the maps are likelihood maps. Nevertheless, overlay maps based on pixel matching (performed on a pixel-by-pixel basis) can be applied after reducing soft maps to several classes or binary maps, and this method can reach conclusions regarding the convergence of the results. This transformation makes it possible to use the most common validation techniques (Paegelow and Camacho Olmedo 2008). For example, Syphard et al. (2011) overlaid binary maps of urban predictions (only including land with a high-probability of development) for several future scenarios, in order to map and quantify where urban growth predictions converged over time.

They also carried out a data reduction by placing the probability images in classes. Another technique known as soft cross-tabulation involves a process of cross-tabulation on soft-classified maps, which preserves continuous values without reducing them into classes, performing a pixel-by-pixel comparison between two maps in which the pixel values have simultaneous memberships of more than one category (also called fuzzy classification). Pontius and Cheuk (2006) compared this method to existing techniques, and proposed that it should be applied to both hard-classified and soft-classified data at any scale. A cross-tabulation tool of this kind for soft-classified maps in which the spatial resolution can be varied is implemented in TerrSet software. The potential of ROC statistics within the framework of land change modeling is analyzed in detail in Mas et al. (2013).

Assessment methods developed for hard-classified maps that focus on the similarity or correspondence between them can also be used for soft-classified maps. The most commonly used tools are Spearman and Pearson correlation indices: similarity can be tested at ordinal and quantitative data level. Using the Spearman rank correlation, Conway and Wellen (2011) evaluated two suitability maps using histograms showing the degree of similarity between the two maps.

2.3.4 DIP—Difference in Change Potential

Difference in Change Potential (DiP) is an assessment technique measuring the difference between the mean potential in the areas of change and the mean potential in the areas of no change, as manifested in the form of hits (correct forecast of change) and false alarms (incorrect forecast of change) (Eastman et al. 2005; Pérez-Vega et al. 2012).

DiP is based on the Peirce Skill Score (PSS) defined as:

$$\mathrm{PSS} = \mathrm{H} - \mathrm{F}$$

where H is the mean potential in the areas of change and F is the mean potential in the areas of no change respectively, and PSS is the difference between them. A value of 1.0 indicates perfect agreement, while a value close to 0 shows random behavior (Pérez-Vega et al. 2012).

2.3.5 Other Validation Techniques/Crossing Techniques

A large number of studies combine various validation techniques. Wang and Mountrakis (2011) compared three models at both per-pixel and neighborhood levels. In the first, they included the confusion matrix, KIA, the receiver operating characteristic (ROC) curve, and multi-scale summary accuracy. The same authors recommended that the results obtained by binary comparison (accurate or not), the probability of change and the spatial accuracy of predicted change be compared.

Lin et al. (2011) use ROC, KIA, multiple resolution validation and landscape metrics to analyze the accuracy of model outputs.

3 Validation According to LUCC Dynamics

The relative importance of the validation techniques presented here also depends on the objective of the model. If the model aims to predict land change, the accuracy of the estimated amount of change is just as important as its allocation. By contrast, if the objective is to design plausible, contrasted, scenarios, the modeler implements quantitative targets with regard to the expected LUC area or changes. From this perspective, validation techniques focus more on spatial pattern and error analysis. Furthermore, the map comparison techniques presented above (particularly the Figure of Merit when computed for outputs of various models) provide useful information about the performance of the models in predicting persistence and change, change components such as net change and swap, and the realism of the landscape. They also allow the modeler to choose the most appropriate model according to the objectives.

Tests with various training dates used for Markov chains show that quantitative accuracy depends on the choice of these dates (see Chap. 7 in part II of this book). This finding shows why it is so important for the modeler to have the key dates at his/her disposal because the Markov chain is strongly dependent on previous trends. If relatively few LUC dates are available, this increases random chance because the Markov chain determines the overall accuracy of the model. If available LUC maps do not allow us to trace past trends or if these trends are not informative for future evolution, it is advisable to support trend-based simulation, also known as the baseline scenario, with various scenarios that deliberately break with Markovian conditional transitions calculated on a basis that is incomplete or becoming obsolete. By varying quantitative assumptions, this geoprospective model (Houet and Gourmelin 2014) implements the allocation of these hypotheses and designs plausible futures.

3.1 Intensity of Dynamics

3.1.1 Splitting Dynamics into Components of Interest: Persistence, Net Change and Swap

LUCC allows us to analyze observed and simulated land change at different levels. The first level is obtained by cross-tabulation of the whole area (Fig. 8). An example from a study of dynamics in a typical European mountain region first shows: persistence (sum of diagonal cells) which amounts to about 97.09%. This means that land use has changed in less than 3% of the study area. Having said this,

		from: 2000						
		Coniferous forest	Deciduous forest	ood recolonization	Broom land	Grassland	Crops	total 2009
to 2009	Coniferous forest	37.37	0.00	0.03	0.20	0.01	0.00	37.61
	Deciduous forest	0.00	8.04	0.00	0.01	0.05	0.00	8.10
	ood recolonization	0.19	0.00	18.91	1.56	0.83	0.00	21.49
	Broom land	0.01	0.00	0.00	25.68	0.01	0.00	25.71
	Grassland	0.01	0.00	0.00	0.00	7.09	0.00	7.10
	Crops	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	total 2000	37.58	8.05	18.94	27.45	7.98	0.00	100.00

Fig. 8 LUCC 2000–2009 in Garrotxes (French Pyrenees); data in percent of study area (8750 ha)

some LUC categories underwent important changes. Expressed as a percentage of the surface area in 2000 (start date), most LUC components remained stable. For example, all land changes from and to coniferous forest (gains and losses) totaled only 1.2% of its extent in 2000. At the other end of the scale, changes in the "wood recolonization" category amounted to 13.9% of the land it occupied in 2000. This means that global or dominant persistence can mask important individual transitions.

Validations may be performed at individual LUC category or transition level or at a global level by considering the overall change potential map (superposition of all the maps) (Pérez-Vega et al. 2012). If we set persistence aside to focus exclusively on change, the accuracy of predicted land change is considerably lower (Brown et al. 2005).

3.1.2 Intensity Analysis

Another means of analyzing model accuracy is to put it into perspective with the intensity of land change (Pontius et al. 2013). *Intensity* is the amount of land change per time unit (e.g. the annual rate). Land change intensity may be analyzed by comparing the amount of change over the study period in several LUC categories (Fig. 9) or by comparing their rate of change over different time periods (Fig. 10). Figure 9 shows that three LUC categories (coniferous and deciduous forest, crops) were more persistent, while broom land and wood recolonization underwent more significant changes over the study period.

For the extent 1942–2009, Fig. 10 shows two intervals in which there was a slow rate of change (1980–1989; 2000–2009), one interval that was close to the average (1942–1962) and two intervals characterized by fast dynamics (1962–1980; 1989–2000). This shows that model accuracy is highly dependent on the comparison interval selected.

Runfola and Pontius (2013) proposed a number of indices based on the difference between individual change rates and the average annual rate of change. For their part, Aldwaik and Pontius (2012) developed tools to measure the intensity of land change at three levels: interval, category and transition. They created indices

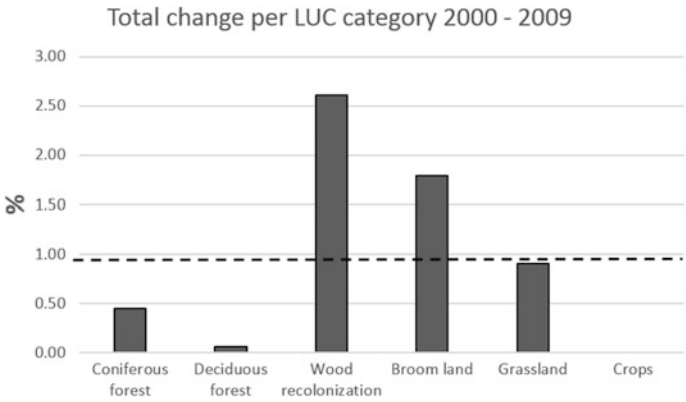

Fig. 9 Total change (expressed in % of the entire study area) per LUC category, Garrotxes 2000–2009. The *dotted line* shows the average LUCC rate

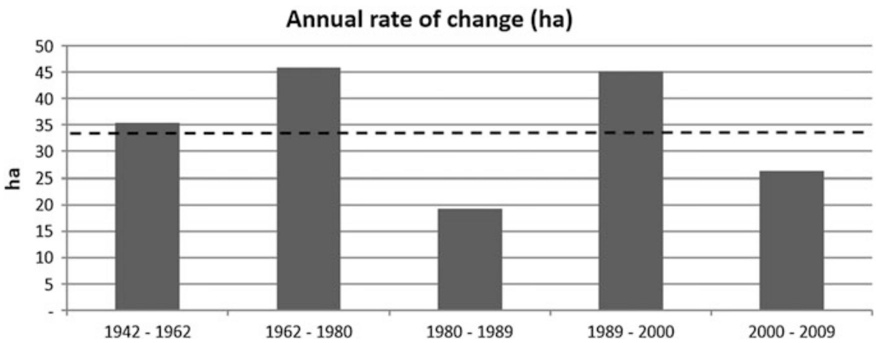

Fig. 10 Annual rate of change in ha (all categories) over the different time periods, Garrotxes. The *dotted line* is the average rate of change over the extent 1942–2009

based on cross-tabulation matrices and distinguished between slow and fast intensities of change with respect to average annual change over different intervals within the whole time extent. They also explored the relative importance of changes (Fig. 8) by unraveling the annual rate of change, expressed in area units or percent, as a proportion of the study area and the amount of annual change expressed in percent of the total area covered by each LUC category. This is important for measuring changes affecting small areas involving relatively less significant (in terms of area) LUC categories. Huang et al. (2012) applied these intensity measures to a coastal watershed in south-eastern China and qualified the categories in which total change was below or above the average as "dormant" versus "active" categories, which respectively "avoid" or "target" transitions.

3.1.3 Data Uncertainty

Pontius et al. (2006); Pontius and Lippitt (2006) proposed a way of using model accuracy measurements to extrapolate predictive uncertainty. Pontius and Petrova (2010) considered the question of whether map error can explain the differences between LUC maps from two points in time. This paper is unique in that it was the first in this series to consider how the level of accuracy in the reference maps influences the interpretation of model validation, and it examines the results for each entry in several cross-tabulation matrices, rather than just overall agreement (Pontius and Millones 2011). This alternative approach had a major impact because most LUCC simulations rely on category data to calibrate and validate the model, and these data often do not have a clear level of accuracy or error structure. The issue of data misclassification within LUCC models has only recently been explored, as have the procedures to follow when the available error information is incomplete. For example, Pontius and Petrova (2010) developed a method for evaluating predicted results when the level of accuracy of the reference data is unknown (Conway and Wellen 2011). Uncertainty in the data is often related to the statistical level of LUC data. This is because most studies are based on qualitative data, which means that LUC is described by categories. The coarser the legend, the higher the uncertainty of the data due to intra-category variance (Paegelow et al. 2014).

3.2 Impact of Spatial and Temporal Scales (Resolution)

Jansen (2006) distinguish three dimensions of scale: (1) space, (2) time and (3) the organizational hierarchy as constructed by the observer. This organizational hierarchy is synonymous with the variation in the semantic contents of data expressed as differences in categorization (Feng and Flewelling 2004). Of these three dimensions, scientists paid little attention to the latter. In fact, so little that this dimension was not even included in the definition of scale cited above (Jansen and Veldkamp 2011). The organization of the data which is finally expressed in the legends of LUC maps is also a critical point, as mentioned above, about which we feel we must insist.

3.2.1 Impact of Spatial Resolution

The concept of scale and resolution is closely linked to the level of detail available in geographic data. Scale refers to printed maps and the level of detail for a given scale is expressed by the minimum mapping unit. The notion of resolution is closely linked to numerical data, especially in raster format and is expressed by the pixel size.

Pontius et al. (2004b) showed that spatial resolution impacts on LUCC components as net change and swap. Using an example of LUC maps for several towns

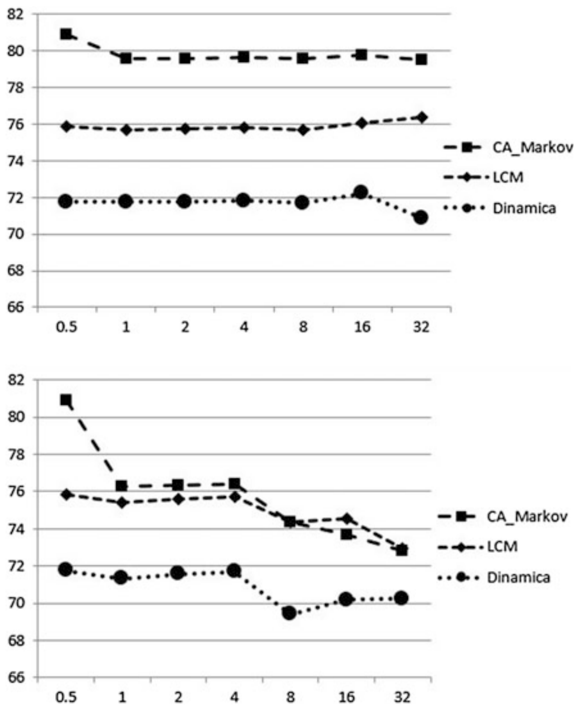

Fig. 11 Varying spatial resolution (geometric sequence) in cross-tabulation between observed LUC and LUC simulated by three land change model tools: CA_MARKOV, LCM and Dinamica Ego applied to pasture land in Eastern Pyrenees. The abscissa shows the spatial resolution in meters while the Y-axis is the percentage of correctly simulated LUC. The *top* figure shows the accuracy rate by pixel thinning and the *bottom* one shows the impact of applying the majority rule

in central Massachusetts, they discovered that the swap component in LUCC budgets is related to spatial scale. The coarser the spatial resolution, the lower the swap. Varying resolution may have different effects when it comes to validating hard-classified land change simulation. We performed pixel-by-pixel cross tabulation between LUC simulated by three models and observed (model unknown) LUC on pasture land in the Eastern Pyrenees by varying the spatial resolution (geometric sequence) and the method of calculating pixel values (pixel thinning and majority). As Fig. 11 shows, the prediction score remains almost stable with coarser resolution when the pixel-thinning technique is applied, while it falls with coarser resolution when the majority rule is applied.

3.2.2 Impact of Temporal Resolution

The influence of scale or resolution—in our case the duration of the time interval—is well known in various disciplines (Allen and Starr 1982; Kim 2013). Several recent studies have formalized the impact of time intervals on the amount of change (Burnicki et al. 2007; Lee et al. 2009; Liu and Deng 2010).

Using various data sources and resolutions, Colas (2016) observed that, as in the case of spatial resolution, short time intervals generate a high rate of change while change intensity decreases with longer intervals. Figure 12 underlines this finding

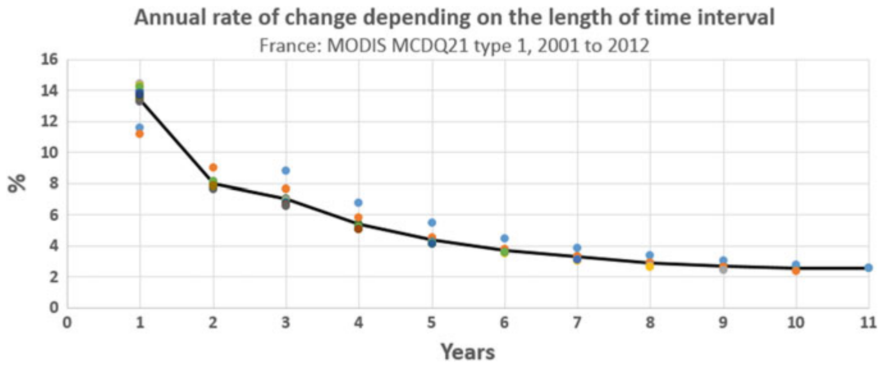

Fig. 12 Annual rate of change (%) depending on the length of the time interval (years). Applied to MODIS MCDQ21 type 1 data for France, 2001–2012

by using MODIS 250 m MCDQ21 data with a type-1 legend for France. The available data are for 2001–2012. The figure shows that the intensity of change decreases exponentially with increasing length of time intervals.

4 Validation According to LUCC Models

The modeling software packages discussed here use either internal validation tools implemented within the modeling program, or external techniques such as parent software, GIS or specific raster tools such as Map Comparison Kit (Visser and de Nijs 2006), especially recommended for CLUMondo (Table 1).

As regards those with built-in validation techniques, all the software packages we considered except for CLUMondo offer cross tabulation to compare hard predictions to observed data. The majority of programs also do this for soft prediction maps, while only TerrSet and Dinamica EGO allow a validation of this kind with multiple resolutions. For their part, Dinamica EGO and APoLUS allow a spatial validation by fuzzy allocation.

With the exception of LucSim and CLUMondo all programs offer various similarity indices for comparing maps. The situation varies more with regard to comparison tools, in that they do not all have tools that offer omissions and commissions and pattern analysis. All of the software packages we considered do however perform a quantitative validation and most of them use ROC statistics.

Table 1 Comparing LUCC models in the validation stage

	Pattern-based models (PBM)				Constraint CA-based models (CCAM)			
	CA_Markov TerrSet	LCM TerrSet	Dinamica EGO	CLU-Mondo	Metronamica	APoLUS	SLEUTH	LucSim
Cross-tabulation for hard classified maps	Yes	Yes	Yes	No, external	Yes	Implemented in parent software(R)	Model creates transition and contingency matrices	Yes
Cross tabulation for soft classified maps	Yes	Yes	Yes: DIP	No	Yes	Implemented in parent software(R)	No	No
Cross tabulation for multiple resolutions	Yes	Yes	YES	No, external	No	Implemented in parent software(R)	Multi-resolution can be used in calibration	No
Fuzzy coincidence	No	No	Yes	No, external	No although available in MCK1	Implemented in parent software(R)	No	No
Map comparison similarity indexes	Cramer's V, KIA, KIA multiple resolutions	Yes	Yes	No, external	Yes, mainly through accom-panying MCK	Ksim, KsimF (MCK, currently working on native R solution)	Model uses 13 statistics based on data matching. Post comparison must be performed independently	No
Map comparison showing correctly predicted changes, omissions, commissions	Yes	Yes	Yes	No, external	Yes, mainly through accom-panying MCK	No. Per category map or Ksim	Post comparison must be performed independently	Confusion matrix
Pattern analysis	Compactness ratio, landscape metrics	Yes	Yes, various	No, external	Yes, mainly through accom-panying MCK	Various pattern based (SDMTools, Fragstats, MCK)	Post comparison must be performed independently.	No
Quantity	Yes	Yes	Yes	Yes	Yes	Yes	YES	Yes
ROC statistics	Yes	Yes	Yes	Yes	Yes	No	No	No

5 Concluding Remarks

Everyone agrees on the importance of model validation. The credibility of the model depends on it. However, the specific nature of each land change model software program and its various options make detailed comparisons impossible. On the other hand, the efforts undertaken by the scientific community in recent years are beginning to bear fruit. Modelers—and critical model users—have never had as many tools at their disposal for assessing the credibility of a simulation or that enable them to focus on particular aspects such as quantitative accuracy, in particular the accuracy of LUCC components, or to pay more attention to landscape pattern similarity. Nevertheless, the impressive array of techniques for calculating validation indices should not make us forget certain limitations. Firstly, the fact that in this chapter we have focused on path-dependent modeling approaches (Houet et al. 2016), while the validation of non trend-based scenarios (also known as contrasted scenarios) is even more difficult. Secondly, we centered on pattern-based models (PBM) while the large panoply of agent-based models (ABM) require their own particular tools, especially when we go beyond purely operational validation to consider conceptual realism as well (Rykiel 1996). Finally, model output accuracy depends above all on the quality of data and its conscientious use, as countless studies have proved.

Acknowledgements This work was supported by the BIA2013-43462-P project funded by the Spanish Ministry of Economy and Competitiveness and by the Regional European Fund FEDER. This study was also supported by the Consejo Nacional de Ciencia y Tecnología (CONACYT) and the Secretaría de Educación Pública through the project ¿Puede la modelación espacial ayudarnos a entender los procesos de cambio de cobertura/uso del suelo y de degradación ambiental? Fondos SEP-CONACyT 178816.

References

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, pp 141–166

Ahlqvist O (2008) 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 Sens Environ 112(3):1226–1241

Aldwaik SZ, Pontius RG Jr (2012) Intensity analysis to unify measurements of size and stationaity of land changes by interval, category, and transition. Landsc Urban Plann 106:103–114

Alcamo J (2008) Environmental futures: the practice of environmental scenario analysis. Developments in integrated environmental assessment, vol 2. Elsevier, Amsterdam

Alcamo J, Schaldach R, Koch J, Kölking C, Lapola D, Priess J (2011) Evaluation of an integrated land use change model including a scenario analysis of land use change for continental Africa. Environ Model Softw 26(8):1017–1027. doi:10.1016/j.envsoft.2011.03.002

Allen THF, Starr B (1982) Hierarchy: Perspectives for ecological complexity. University of Chicago Press, Chicago 310 p

Burnicki AC, Brown DG, Goovaerts P (2007) Simulating error propagation in land-cover change analysis: The implications of temporal dependence. Comput Environ Urban Syst 31:282–302

Bone C, Dragicevic S, White R (2011) Modeling-in-the middle: bridging the gap between agent-based modeling and multi-objective decision-making for land use change. Int J Geogr Inf Sci 25(5):717–737. doi:10.1080/13658816.2010.495076

Botequilha A, Miller J, Ahern J, McGarigal K (2006) Measuring landscapes. A planner´s handbook. Island Press, Washington

Brown D et al (2005) Path dependence and the validation of agent-based spatial models of land use. Int J Geogr Inf Sci 19:153–174

Camacho Olmedo MT, Paegelow M, Mas JF (2013) Interest in intermediate soft-classified maps in land change model validation: suitability versus transition potential. Int J Geogr Inf Sci 27(12):2343–2361

Camacho Olmedo MT, Pontius RG Jr, Paegelow M, Mas JF (2015) Comparison of simulation models in terms of quantity and allocation of land change. Environ Modell Softw 69:214–221 (Elsevier)

Colas R (2016) Comprendre les changements d'utilisation des terres en France pour mieux estimer leurs impacts sur les émissions à effet de serre. De l'observation à la modélisation. PhD thesis, University Paris 7, 530 p

Conway TM, Wellen CC (2011) Not developed yet? Alternative ways to include locations without changes in land use change models. Int J Geogr Inf Sci 25(10):1613–1631. doi:10.1080/13658816.2010.534738

Coquillard P, Hill DRC (1997) Modélisation et simulation d'écosystèmes. Des modèles déterministes aux simulations à événements discrets, Paris, Masson

Croks AT, Heppenstall AJ (2012) Introduction to Agent-based modelling. In: Heppenstall et al (eds) Agent-based models of geographical systems. Springer, pp 85–108

Eastman JR, Jin W, Kyem PAK, Toleando J (1995) Raster procedures for multi-criteria/multi-objective decisions. Photogramm Eng Remote Sens 61(5):539–547

Eastman JR, Solorzano LA, Van Fossen ME (2005) Transition potential modeling for landcover change. In: Maguire DJ, Batty M, Goodchild MF (eds) GIS, spatial analysis, and modeling. ESRI, Redland CA, pp 357–385

Eastman JR (2009) Idrisi Taiga, Guide to GIS and Image Processing, manual version 16.02, Clark University 342 pp

Engelen G, White R (2007) Validating and calibrating integrated cellular automata based models of land use change. In: Albeverio S et al (eds) The dynamics of complex urban systems. An interdisciplinary approach, Physica, pp 185–211

Forman RTT (1995) Land Mosaics: the ecology of landscapes and regions. EEUU, Cambridge

Gaucherel C (2007) Multiscale heterogeneity map and associated scaling profile for landscape analysis. Landsc Urban Plann 82(3):95–102

Gaucherel C, Alleaume S, Hély C (2008) The Comparison Map Profile method: a strategy for multiscale comparison of quantitative and qualitative images. IEEE Trans Geosci Remote Sens 46(9):2708–2719

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

Gómez Delgado M, Tarantola S (2006) Global sensitivity analysis, GIS and multi-criteria evaluation for a sustainable planning of hazardous waste disposal site in Spain. Int J Geogr Inf Sci 20:449–466

Hagen A (2003) Fuzzy set approach to assessing similarity of categorical maps. Int J Geogr Inf Sci 17(3):235–249

Hagen-Zanker A, Straatman B, Uljee I (2005) Further developments of a fuzzy set map comparison approach. Int J Geogr Inf Sci 19:769–785

Hagen-Zanker A (2009) An improved Fuzzy Kappa statistic that accounts for spatial autocorrelation. Int J Geogr Inf Sci 23(1):61–73. doi:10.1080/13658810802570317

Houet T, Gourmelon F (2014) La géoprospective—Apport de la dimension spatiale aux approches prospectives. Cybergéo 8 pp. http://cybergeo.revues.org/26194

Houet T, Aguejdad R, Doukari O, Battaia G, Clarke K (2016) Description and validation of a "non path-dependent" model for projecting contrasting urban growth futures. Cybergeo, document 759, 6 Jan 2016. http://cybergeo.revues.org/27397

Huang J, Pontius RG Jr, Li Q, Zhang Y (2012) Use of intensity analysis to link patterns with process of land change from 1986 to 2007 in a coastal watershed of southeast China. Appl Geogr 34:371–384

Jansen LJM (2006) Harmonisation of land-use class sets to facilitate compatibility and comparability of data across space and time. In: 12th CEReS international symposium, 13–14 Dec 2005, Japan 29 pp

Jackson DA (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, Springer Theses

Jansen LJM, Veldkamp TA (2011) Evaluation of the variation in semantic contents of class sets on modelling dynamics of land-use changes. Int J Geogr Inf Sci. doi:10.1080/13658816.2011.609989

Kamusoko C, Aniya M, Adi B, Manjoro M (2009) Rural sustainability under threat in Zimbabwe-Simulation of future land use/cover changes in the Bindura district based on the Markov-cellular automata model. Appl Geogr 29:435–447. doi:10.1016/j.apgeog.2008.10.002

Kim JH (2013) Spatiotemporal scale dependency and other sensitivities in dynamic land-use change simulations. Int J Geogr Inf Sci 27:1782–1803

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. Elsevier, London

Krüger C, Lakes T (2015) Revealing uncertainties in land change modeling using probabilities. Trans GIS 20(4):526–546

Kuhnert M, Voinov A, Seppelt R (2005) Comparing raster map comparison algorithms for spatial modelling and analysis. Photogramm Eng Remote Sens 71(8):975–984

Lee YJ, Lee JW, Chai DJ, Hwang BH, Ryu KH (2009) Mining temporal interval relational rules from temporal data. J Syst Softw 82:155–167

Li X, Chen Y, Liu X, Li D, He J (2011) Concepts, methodologies and tools of an integrated geographical simulation and optimization system. Int J GIS 25(4):633–655

Lin YP, Chu HJ, Wu CF, Verburg PH (2011) Predictive ability of logistic regression, auto-logistic regression and neural network models in empirical land-use change modeling—a case study. Int J Geogr Inf Sci 25(1):65–87. doi:10.1080/13658811003752332

Liu J, Deng X (2010) Progress of the research methodologies on the temporal and spatial process of LUCC. Chin Sci Bull 55:1354–1362

Mas JF, Kolb M, Houet T, Paegelow M, Camacho Olmedo MT (2011) Eclairer le choix de modèles de simulation des changements des modes d'occupation et d'usages des sols. Revue Internationale de Géomatique 21(3):405–430

Mas JF, Soares Filho B, Pontius RG Jr, Farfan Gutierrez M, Rodrigues H (2013) A suite of tools for ROC analysis of spatial models. ISPRS Int J Geo-Inf 2(3):869–887. doi:10.3390/ijgi2030869

Mas JF, Kolb M, Paegelow M, Camacho Olmedo MT, Houet T (2014) Inductive pattern-based land use/cover change models: a comparison of four software packages. Environ Model Softw 51:94–111

McGarigal K, Marks BJ (1995) FRAGSTATS: spatial pattern analysis program for quantifying landscape structure. USDA For Serv Gen Tech Rep

Mobaied S, Riera B, Lalanne A, Baguette M, Machon N (2011) The use of diachronic spatial approaches and predictive modelling to study the vegetation dynamics of a managed heathland. Biodivers Conserv 20:73–88. doi:10.1007/s10531-010-9947-1

Ngo TA, See L (2012) Calibration and validation of agent-based models of land cover change. In: Heppenstall et al (eds) Agent-based models of geographical systems. Springer, pp 181–198

North MJ, Macal CM (2007) Managing business complexity: Discovering strategic solutions with agent-based modelling and simulation. Oxford University Press, New York

Oreskes N, Shrader-Frechette K, Belitz K (1994) Verification, validation and confirmation of numerical-models in the earth-sciences. Science 263:641–646

Paegelow M, Camacho Olmedo MT (2005) Possibilities and limits of prospective GIS land cover modelling—a compared case study: Garrotxes (France) and Alta Alpujarra Granadina (Spain). Int J Geogr Inf Sci 19(6):697–722

Paegelow M, Camacho Olmedo MT (eds) (2008) Modelling environmental dynamics. Advances in geomatic solutions. Springer, series: Environmental science and engineering

Paegelow M, Camacho Olmedo MT, Mas JF, Houet T, Pontius RG Jr (2013) Land change modelling: moving beyond projections. Int J Geogr Inf Sci 27(9):1691–1695

Paegelow M, Camacho Olmedo MT, Mas JF, Houet T (2014) Benchmarking of LUCC modelling tools by various validation techniques and error analysis. Cybergeo 701, 22 Dec 2014. http://cybergeo.revues.org

Perica S, Foufoula-Georgiou E (1996) Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions. J Geophys Res 101(D21):26347–26361
Pérez-Vega A, Mas JF, 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. Environ Model Softw 29(1):11–23
Pontius RG Jr (2000) Quantification error versus location error in comparison of categorical maps. Photogramm Eng Remote Sens 66(8):1011–1016
Pontius RG Jr (2002) Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. Photogramm Eng Remote Sens 68:1041–1049
Pontius RG Jr, Pacheco P (2004) Calibration and validation of a model of forest disturbance in the Western Ghats, India 1920–1990. GeoJournal 61:325–334
Pontius RG Jr, Schneider LC (2001) Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agric Ecosyst Environ 85:239–248
Pontius RG Jr, Huffaker D, Denman K (2004a) Useful techniques of validation for spatially explicit land-change models. Ecol Model 179(4):445–461
Pontius RG Jr, Shusas E, McEachern M (2004b) Detecting important categorical land changes while accounting for persistence. Agric Ecosyst Environ 101:251–268
Pontius RG Jr, Malanson J (2005) Comparison of the structure and accuracy of two land change models. Int J Geogr Inf Sci 19:243–265
Pontius RG Jr, Cheuk ML (2006) A generalized cross-tabulation matrix to compare soft-classified maps at multiple resolutions. Int J Geogr Inf Sci 20(1):1–30. doi:10.1080/13658810500391024
Pontius RG Jr, Lippitt CD (2006) Can error explain map differences over time? Cartogr Geogr Inf Sci 33(2):159–171
Pontius RG Jr, Verluis AJ, Malizia NR (2006) Visualizing certainty of extrapolations from models of land change. Landsc Ecol 21(7):1151–1166
Pontius RG Jr, Walker RT, Yao-Kumar R, Arima E, Aldrich S, Caldas M, Vergara D (2007) Accuracy assessment for a simulation model of Amazonian deforestation. Ann Assoc Am Geogr 97:677–695
Pontius Jr RG, LI X (2010) Land transition estimates from erroneous maps. J Land Use Sci 5 (1):31–44
Pontius RG Jr, Petrova SH (2010) Assessing a predictive model on land change using uncertain data. Environ Model Softw 25:299–309
Pontius RG Jr, Boersma W, Castella JC, Clarke K, de Nijs T, Dietzel C, Duan Z, Fotsing E, Goldstein N, Kok K, Koomen E, Lippitt CD, McConnell W, Sood AM, Pijankowski B, Pidhadia S, Sweeney S, Trung TN, Veldkamp AT, Verburg PH (2008) Comparing the input, output, and validation maps for several models of land change. Anna Region Sci 42(1):11–27
Pontius RG Jr, Millones M (2011) Death to kappa and to some of my previous work: a better alternative. Int J Remote Sens 32(15):4407–4429
Pontius RG Jr, Gao Y, Giner NM, Kohyama T, Osaki M, Hirose K (2013) Design and interpretation of intensity analysis illustrated by land change in Central Kalimantan, Indonesia. Land 2(3):351–369. doi:10.3390/land2030351
Pontius Jr RG, Si K (2013) The total operating characteristic to measure diagnostic ability for multiple thresholds. Int J Geogr Inf Sci
Rodrigues HO, Soares Filho BS, de Souza Costa WL (2007) Dinamica EGO, uma plataforma para modelagem de sistemas ambientais. Anais XIII Simposio Brasileiro de Sensoriamento Remoto, INPE 3089–3096
Runfola DM, Pontius RG Jr (2013) Measuring the temporal instability of land change using the flow matrix. Int J Geogr Inf Sci 26(9):1696–1716
Rykiel EJJ (1996) Testing ecological models: the meaning of validation. Ecol Model 90:229–244
Sargent RG (2009) Verification and validation of simulation models. In: Rosetti MDD, Hill RR, Johansson B, Dunkin A, Ingalls RG (eds) Proceedings of the 2009 winter simulation conference. IEEE, Syracuse, NY, pp 162–176

Schneider LC, Pontius RG Jr (2001) Modeling land-use change in the Ipswich watershed, Massachusetts, USA. Agric Ecosyst Environ 85:83–94
Shirley LJ, Battaglia LL (2008) Projecting fine resolution land cover dynamics for a rapidly changing terrestrial-aquatic transition in Terrebone basin, Louisiana, USA. J Coastal Res 246:1545–1554
Silva TS, Tagliani PRA (2012) Environmental planning in the medium littoral of the Rio Grande do Sul coastal plain—Southern Brazil: elements for coastal management. Ocean Coast Manag 59:20–30
Soares-Filho BS, Pennachin CL, Cerqueira G (2002) DINAMICA—a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. Ecol Model 154(3):217–235
Sohl TL et al (2012) Spatially explicit land-use and land-cover scenarios for the Great Plains of the United States. Agric Ecosyst Environ 153:1–15. doi:10.1016/j.agee.2012.02.019
Steiner F, McSherry L, Cohen J (2000) Land suitability analysis for the upper Gila River watershed. Landsc Urban Plann 50(4):199–214
Syphard AD, Clarke KC, Franklin J, Regan HM, Mcginnis M (2011) Forecasts of habitat loss and fragmentation due to urban growth are sensitive to source of input data. J Environ Manage 92:1882–1893
Takada T, Miyamoto A, Hasegawa SF (2010) Derivation of a yearly transition probability matrix for land-use dynamics and its applications. Landsc Ecol 25(4):561–572
Torrens PM (2011) Calibrating and validating cellular automata models of urbanization. In: Yang X (ed) Urban remote sensing. Monitoring, synthesis and modelling in the urban environment. Wiley, New York, pp 335–345
Turner MG, Constanza R, Sklar FH (1989) Methods to evaluate the performance of spatial simulation models. Ecol Model 48:1–18
Villa N, Paegelow M, Camacho Olmedo MT, Cornez L, Ferraty F, Ferré L, Sarda P (2006) Various approaches for predicting land cover in Mediterranean mountains. Commun Stat 36 (1):73–86
Visser H, de Nijs T (2006) The map comparison kit. Environ Model Softw 21:346–358
van Vliet J, Bregt AK, Brown DG, van Delden H, Heckbert S, Verburg PH (2016) A review of current calibration and validation practices in land change modelling. EMS 82:174–182
Wang J, Mountrakis G (2011) Developing a multi-network urbanization model: A case study of urban growth in Denver, Colorado. Int J Geogr Inf Sci 25(2):229–253. doi:10.1080/13658810903473213
White R, Uljee I, Engelen G (2012) Integrated modelling of population, employment, and land use change with a multiple activity based variable grid cellular automaton. Int J GIS 26:1251–1280
White R, Engelen G, Uljee I (1997) The use of constrained cellular automata for high resolution modelling of urban land use dynamics. Environ Plan 24:323–343
Yu J, Chen Y, Wu J, Khan S (2011) Cellular automata-based spatial multi-criteria land suitability simulation for irrigated agriculture. Int J Geogr Inf Sci 25(1):131–148. doi:10.1080/13658811003785571