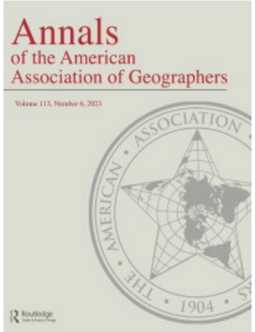

Annals of the American Association of Geographers

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/raag21

Four Fundamental Questions to Evaluate Land Change Models with an Illustration of a Cellular Automata–Markov Model

Cláudia M. Viana, Robert Gilmore Pontius Jr. & Jorge Rocha

To cite this article: Cláudia M. Viana, Robert Gilmore Pontius Jr. & Jorge Rocha (2023): Four Fundamental Questions to Evaluate Land Change Models with an Illustration of a Cellular Automata–Markov Model, Annals of the American Association of Geographers, DOI: 10.1080/24694452.2023.2232435

To link to this article: https://doi.org/10.1080/24694452.2023.2232435

Published online: 25 Aug 2023.

Four Fundamental Questions to Evaluate Land Change Models with an Illustration of a Cellular Automata–Markov Model

Cláudia M. Viana,[*] Robert Gilmore Pontius, Jr.,[†] and Jorge Rocha[*]

[*]Centre for Geographical Studies, Institute of Geography and Spatial Planning, University of Lisbon, Portugal
[†]School of Geography, Clark University, USA

Numerous models exist for users to simulate land change to communicate with an audience concerning future land change. This article raises four fundamental questions to help model users decide whether to use any model: (1) Can the user understand the model? (2) Can the audience understand the model? (3) Can the user control the model? (4) Does the model address the goals of the specific application? This article applies these questions to the popular cellular automata–Markov (CA–Markov) model as IDRISI's CA–Markov module expresses. Sensitivity analysis examines 120 ways to set the module's parameters. Verification compares the module's behavior to the software's documentation. Results show that the cellular automata's allocation fails to follow the quantity of change that the Markov module computes. The module's behavior is likely to cause users to misinterpret the validation metrics and to miscommunicate with audiences. Thus, the answers to the four questions were not satisfactory for this article's case study. This article's framework helps users to judge a model's appropriateness for a specific application by combining sensitivity analysis with verification in a manner that helps to interpret validation. Users should answer the four questions as they decide whether to use any software's modules. *Key Words: CA–Markov, IDRISI software, sensitivity analysis, validation, verification.*

Dynamic simulation models have been the predominant tools used to support the analysis of future changes among land categories (Houet, Verburg, and Loveland 2010; Noszczyk 2019). Scientists have developed numerous land change models to depict possible land changes and to support landscape planning and environmental management (Verweij et al. 2018; Roodposhti et al. 2019; Sankarrao et al. 2021). Therefore, it can be difficult for a prospective model user to decide which model(s) to use because many models exist. Also, there is a need to encourage the use of more quantitative and qualitative methods to assess the quality of software and algorithms (Nüst and Pebesma 2021; Tullis and Kar 2021; Wilson et al. 2021). This article raises four fundamental questions to guide users to decide whether to use any model to address the goals of the user's specific application: (1) Can the user understand the model? (2) Can the audience understand the model? (3) Can the user control the model? (4) Does the model address the goals of the specific application? We encourage users to ask themselves these questions repeatedly when working with geographic information systems (GIS) software and algorithms to study geographical phenomena. The answers to these questions must be satisfactory for a user to be able to control the model's parameters to create various scenarios of landscape change in a manner that will communicate helpful insights. A model must be well documented for the user to have any hope to answer satisfactory to Question 1. A model must be sufficiently straightforward to answer satisfactory to Question 2. A model must have parameters that allow the user to control the output to answer satisfactory to Question 3. The answer to Question 4 depends on the alignment between the goals and the model. If the answers to Questions 1 through 3 are not satisfactory, then the likely answer to Question 4 is also not satisfactory. We recommend prospective users apply our four fundamental questions to evaluate the appropriateness of any model for their specific goals.

We apply our four questions to evaluate a cellular automata–Markov (CA–Markov) model as the Selva version of the IDRISI software expresses in its CA–Markov module (Eastman 2006, 2012; Eastman and Toledano 2018). We select IDRISI's CA–Markov

Annals of the American Association of Geographers, 0(0) 2023, pp. 1–15
Initial submission, June 2022; revised submissions, January and April 2023; final acceptance, May 2023

module for several reasons. Cellular automata (CA) and Markov matrices are general mathematical concepts that exist in other software, so our analysis can offer insights that might apply to other software. CA is popular and makes theoretical sense when a cell's neighbors influence the likelihood that the cell experiences change through time. The Markov process is popular in mathematics and land change modeling software. We suspect Markov's popularity derives from its mathematical convenience because we have neither theory nor evidence that landscapes change according to a Markov process. CA–Markov is integrated into the IDRISI software, which makes CA–Markov easily accessible to IDRISI's 100,000 users worldwide. IDRISI's modules are better documented than some other software packages so we have a better chance of having the information necessary to answer the four questions than we would have with some other software packages. The creators of the IDRISI software introduced the CA–Markov module as experimental, which warrants its testing. The creators of IDRISI have considered removing CA–Markov from the software, but users wanted CA–Markov to remain in the software because they have frequently claimed that CA–Markov is helpful. CA–Markov has been popular in the literature for many years. Examples include Aksoy and Kaptan (2021), Aliani et al. (2019), and Ghosh et al. (2017).

Any model that simulates transitions among land categories across time and space must perform two tasks. First, the model must specify the size of the area that transitions from one category to another category during each time interval, which is a concept known as *quantity*. The Markov part specifies the quantity in a CA–Markov model. Second, the model must specify the spatial distribution of the transitions, which is a concept known as *allocation*. The CA part controls the allocation in the CA–Markov model (Eastman 2012; Mas et al. 2014; Camacho Olmedo et al. 2015). If a module's parameters allow the user to specify the quantity separately from the allocation, then the user can control two important components to create various scenarios of simulated change, in which case the user might be able to answer satisfactory to Question 3.

The behavior and output of a module derive from three factors: (1) the user's decisions concerning how to format the input data, (2) the user's selection of the module's parameters, and (3) the software's design (Dahal and Chow 2015; Liao et al. 2016; Lin et al. 2020). For the evaluation of the first two factors, we apply sensitivity analysis to measure how the model's output varies depending on eight ways to format the input data and fifteen ways to set the model's parameters. Thus, the number of combinations of data format and parameter selection is eight multiplied by fifteen, which generates 120 runs in the sensitivity analysis. For the evaluation of the third factor, we cannot apply sensitivity analysis because we do not control the software's design.

Some users are tempted to evaluate a model based on its predictive power, which can be an unfair criterion to evaluate a module for several reasons. First, there are many ways to format the input data, and each one can influence the module output. Second, there are many ways to select the module's parameters and each one can influence the output. More important, the purpose of many models is to extrapolate from the calibration time interval, not necessarily to predict accurately during the extrapolation time interval. If the patterns in the reference data are not stationary from the calibration interval to the extrapolation interval, then an extrapolation from the calibration interval will not have predictive power because the patterns in the landscape are not consistent through time. Thus, low measures of validation might be due more to the landscape's nonstationarity than to the model failing to do what the software developers intended. Nevertheless, validation is an important consideration for many users (van Vliet et al. 2016). For example, Memarian et al. (2012) applied IDRISI's CA–Markov to an application where validation revealed more errors than correctly simulated change, which is typical for land change models (Pontius et al. 2008; Pontius et al. 2018). Memarian et al. (2012) concluded that "CA–Markov shows poor performance for land use and cover change simulation due to uncertainties in the source data, the model, and future land use and cover change processes in the study area" (542). The CA–Markov algorithm might have behaved exactly how the software developers intended, but the validation results derived from poor data quality, inappropriate selection of model parameters, or nonstationarity of the processes, in which case it is inappropriate to blame the algorithm for the validation results. Therefore, this article illustrates an insightful method to evaluate 120 runs in a manner that distinguishes validation from verification.

Validation measures predictive power; verification tests whether a module behaves as its documentation leads the user to believe it will.

We illustrate our four fundamental questions by applying IDRISI's CA–Markov module to a 10,000 km^2 region in Beja District, Portugal, which is a mixed agro-silvo-pastoral environment with low urban density (Viana, Girão, and Rocha 2019; Viana and Rocha 2020). Our innovation is to combine sensitivity analysis with verification in a manner that helps to interpret validation. This is important because errors that the validation reveal might derive from an inappropriate conceptual model or an unintended behavior of a software's module. Our purpose is to judge a model's appropriateness for a specific application and to describe how the CA–Markov module behaves for the application. Our results give insights into the general CA–Markov concept and also into IDRISI's implementation in its CA–Markov module.

Material and Methods

Study Area and Data

The study area is the Beja District in southeastern Portugal, which covers 11 percent of Portugal's mainland and had 144,000 inhabitants in 2021 (Statistics Portugal 2021). Beja is flat in its southeast and has extensive plains cut by tiny hills in its north and west. The landscape is a mixed agro-silvo-pastoral land-use system. Figure 1 shows the reference maps of land categories at three time points: 1995, 2007, and 2018. We calibrate the model using the reference change between 1995 and 2007, CA–Markov simulates change from 2007 to 2018, and then validation compares the simulated change to the reference change between 2007 and 2018.

Our study uses *Carta do uso e ocupação do solo* (COS) maps from 1995, 2007, and 2018 produced by the Portuguese General Directorate for Territorial Development (DGT). DGT used photo interpretation, vector polygon data, cartographic accuracy of 0.5 m, minimum mapping unit of 1 ha, and a hierarchical nomenclature system (DGT 2018). We converted the COS maps into raster format at four spatial resolutions: 10×10 m, 20×20 m, 50×50 m, and 100×100 m. We aggregated the COS categories in two ways to produce either two or five categories. The two categories are the agricultural category and the aggregation of all nonagricultural categories. The five categories are the aggregated nonagriculture category and four types of agricultural categories based on the second hierarchical level: arable land, permanent crops, pastures, and heterogeneous areas. Table 1 specifies how the two and five land-use categories derive from the COS levels.

IDRISI's CA–Markov Model

We used the CA–Markov module in the Selva version of the IDRISI software (Eastman 2006, 2012; Eastman and Toledano, 2018). The software's documentation describes three stages: (1) calculation of the transition area matrix, (2) development of the suitability images, and (3) allocation of change. The software reads land-use maps at two time points and the duration of the extrapolation to compute a transition area matrix, which expresses the extrapolated quantity of each transition from one category to another category. The software also reads a collection of suitability images that express the suitability of each pixel for each of the land categories. The suitability maps and a spatial filter influence the allocation of simulated change.

Figure 2 presents an overview of the modeling process for our application. The upper left of Figure 2 shows Stage 1, which computes the Markov transition area matrix that derives from the land-use map at 1995, the land-use map at 2007, and the eleven-year duration for extrapolation. The upper right of Figure 2 shows Stage 2, where the user converts the driver maps to suitability maps and sets the parameters that control the neighborhood configuration of CA's spatial filter. The middle of Figure 2 is Stage 3, which allocates the simulated change from 2007 to 2018. Our experimental design repeats the simulation for the two thematic scales, four spatial resolutions, three neighborhood shapes, and five neighborhood sizes. This generates 120 combinations; thus, we ran the CA–Markov module 120 times. For each run, verification compares the size of change in the extrapolated area transition matrix to the size of change for the simulated maps to see whether the module performs as its documentation describes. The bottom of Figure 2 shows validation to compare each run of simulated change to the reference change from 2007 to 2018. The following subsections describe the modeling procedures during each stage.

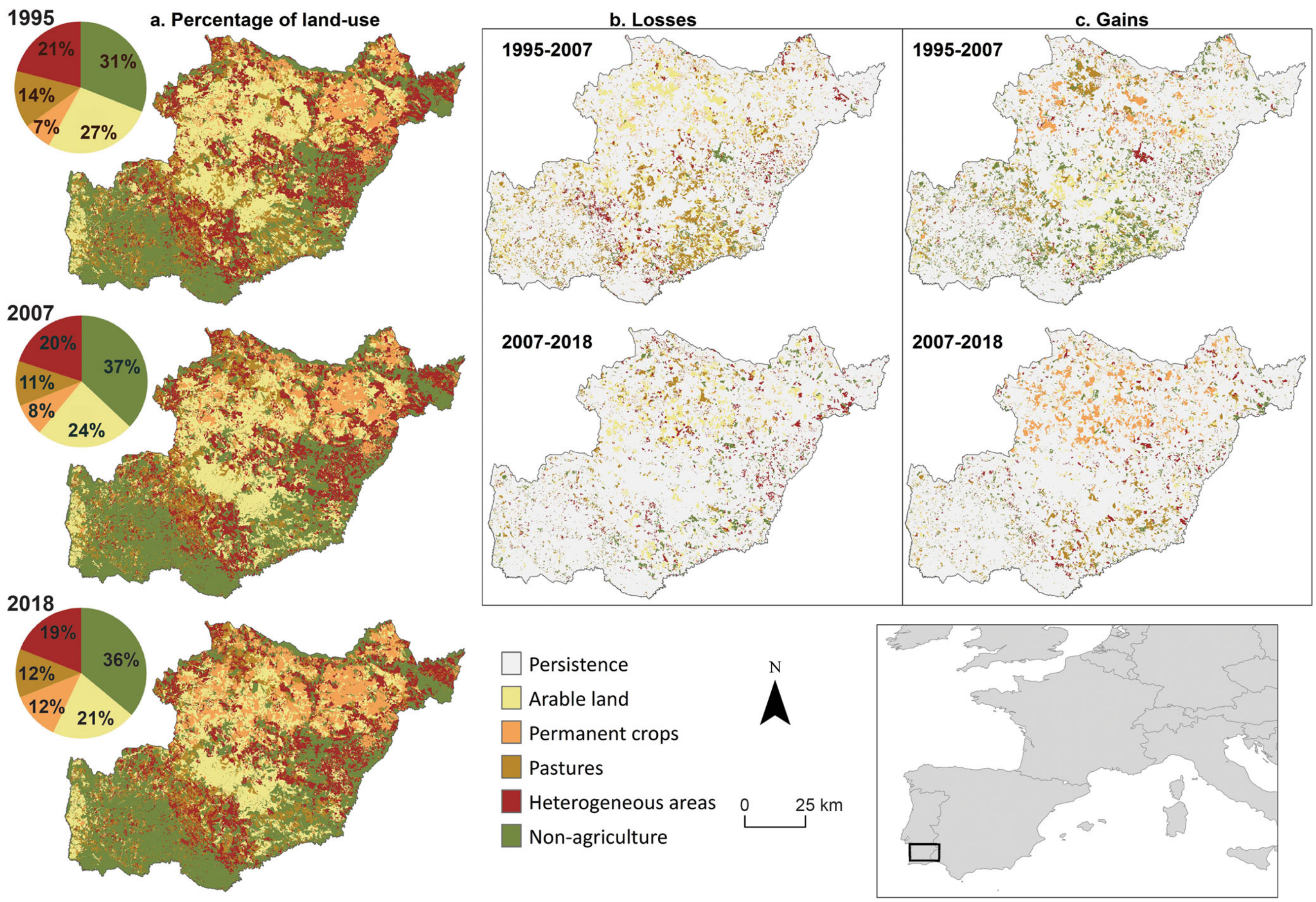

Figure 1. Reference land use in 1995, 2007, and 2018 and changes from 1995 to 2007 and 2007 to 2018 for the Beja District with a five-category classification.

Table 1. Land-use category description

Category in model	Category in COS [level]	Description
Nonagriculture	Artificial surfaces, forest and seminatural areas, wetlands, and water bodies [1]	Urban fabric; artificial nonagricultural vegetated areas; forests; open spaces with little or no vegetation; inland wetlands
Agricultural areas	Agricultural areas [1]	Areas principally occupied by agriculture, interspersed with significant natural or seminatural areas
Arable land	Arable land [2]	Lands that are rain-fed or irrigated under a rotation system used for annually harvested plants and fallow lands
Permanent crops	Permanent crops [2]	Lands not under a rotation system, includes fruit orchards, olive groves, and shrub orchards such as vineyards
Pastures	Pastures [2]	Lands that are used for at least 5 years for fodder production
Heterogeneous areas	Heterogeneous areas [2]	Landscapes in which permanent crops on the same parcel, meadows, and/or pastures are intimately mixed with natural vegetation or natural areas

Note: COS = *Carta do uso e ocupação do solo*.

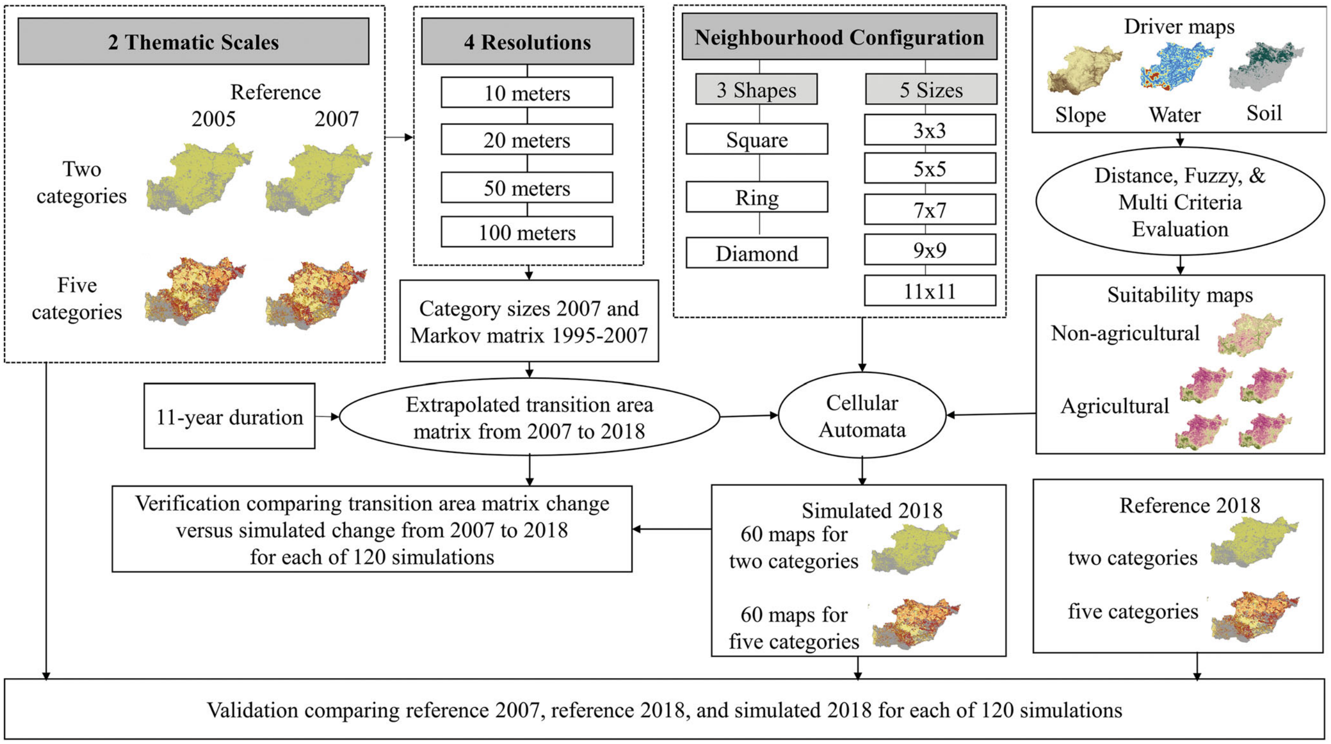

Figure 2. Flow diagram of the modeling framework. The dark infilled rectangles represent the various ways we formatted the input data and set the neighborhood configuration parameters, which generate 120 runs. Ovals indicate software procedures.

Calculation of the Transition Area Matrix. The first stage computes the transition area matrix, which extrapolates land change in terms of quantity. IDRISI's Markov module controls this first stage. The Markov module computes Markov proportions during the calibration time interval then multiplies each loss proportion by 11/12 to account for the fact that the extrapolation interval is eleven years and the calibration interval is twelve years. We used the default proportional error of zero where the Markov module allows the user to set a proportional error. The resulting matrix of proportions is multiplied by the category sizes at 2007 to compute the size of each extrapolated transition from 2007 to 2018 (Muller and Middleton 1994; Eastman 2006, 2012). Matrix **A** in Equation 1 expresses the size of each extrapolated transition, which IDRISI's CA–Markov module then reads as input for the allocation.

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1J} \\ \vdots & \ddots & \vdots \\ a_{J1} & \cdots & a_{JJ} \end{bmatrix} \quad (1)$$

Entry a_{ij} in row i and column j of matrix **A** is the area that is category i at the start of the extrapolation and category j at the end of the extrapolation. The number of categories is J.

Development of the Suitability Maps. The user must convert driver maps into suitability maps in the second stage. Each suitability map portrays the appropriateness of each cell for a particular land-use category. Slope, water bodies, and soil type are the driver maps that determine our suitability maps. The slope data are from *Instituto Geográfico Português*, water bodies are from the *Agência Portuguesa do Ambiente*, and soil type data are from *Cartas de Solos e de Capacidade de Uso do Solo* collected from the *Direção-Geral de Agricultura e Desenvolvimento Rural.* We developed one suitability map for the nonagriculture category and another for the agricultural category for the two-category case. For the five-category case, the four agricultural categories use the same agricultural suitability map. Flatter slopes have higher suitability for agriculture. Closer distances to water bodies have higher suitability for agriculture. We rescaled the soil type map onto the [0, 1] interval where larger values indicate higher suitability for agriculture. A weighted linear combination multicriteria evaluation combined the suitability maps by using weights for slope as 0.4, distance to water bodies as 0.4, and soil type as 0.2. Data availability influenced variable selection and expert opinion determined the suitability values and the weights.

Change Allocation. In the third stage, the CA–Markov module allocates the extrapolated quantities in the transition area matrix **A**. CA consists of cells, states, time, neighborhoods, and transition rules (Torrens 2000). Equation 2 expresses conceptually how the simulated change derives from a cell's category, the number of cells that change from one category to another, the suitability maps, and the cell's neighbors (Wolfram 1984; White and Engelen 2000).

$$^{t+1}C_m = f\left(^{t}C_m, a_{ij}, S_{mj}, {}^{t}N_m\right) \tag{2}$$

where $^{t+1}C_m$ denotes the category at time $t+1$ of cell m; $^{t}C_m$ denotes the category at time t of cell m; a_{ij} is the matrix entry that gives the extrapolated size of transition from category i to category j; S_{mj} is the suitability of cell m for category j, and $^{t}N_m$ denotes the condition at time t of the neighborhood of cell m.

IDRISI's CA–Markov requires the land-use map at 2007 and the Markov transition area matrix **A** for the extrapolation from 2007 to 2018. The CA module allows the user to define the neighborhood configuration of a spatial filter. The spatial filter defines a cell's neighborhood in terms of shape and size (Verburg et al. 2004). Figure 3 shows three filter shapes: diamond, ring, and square. The diamond shape includes pixels that are within the same Manhattan distance from the central cell. The ring shape is the outermost cells at a distance identical to the square type, where the neighborhood size is the number of cells on one side. The square shape is a square neighborhood that fills the ring (White and Engelen 2000; Pan et al. 2010). Figure 3 shows how the spatial filter has fifteen combinations that derive from three shapes and five sizes.

Model Validation and Verification

A three-way cross-tabulation compares the simulated change to the reference change for validation. We compared the reference map at 2007, the reference map at 2018, and each of the 120 simulated maps at 2018. This comparison generates five components: misses, hits, wrong hits, false alarms, and

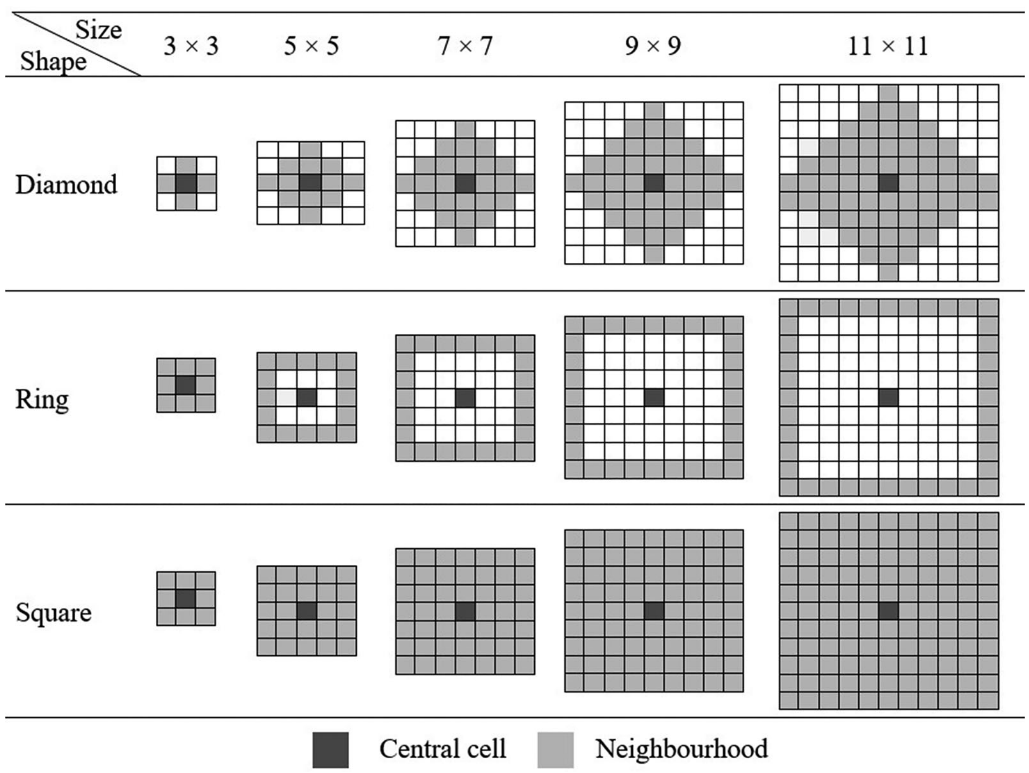

Figure 3. Fifteen combinations of the spatial filter's shape and size.

correct rejections. Misses are reference changes simulated as persistence. Hits are reference changes simulated as changes to the correct gaining category. Wrong hits are reference changes simulated as change to the wrong gaining category, which requires more than two categories. False alarms are reference persistence simulated as changes. Correct rejections are reference persistence simulated as persistence. Quantity error is the absolute difference between misses and false alarms. Allocation error is two times the minimum of misses and false alarms. Equation 3 computes the figure of merit (FOM), which is a measure of the accuracy of change, where 0 percent means no intersection between simulated change and reference change and 100 percent means perfect intersection between simulated change and reference change (Pontius et al. 2007; Pontius et al. 2008; Pontius, Peethambaram, and Castella 2011; Varga et al. 2019).

$$\text{FOM} = \frac{\text{Hits}}{\text{Misses} + \text{Hits} + \text{Wrong Hits} + \text{False Alarms}} 100\% \tag{3}$$

In addition, we computed the square kilometers per year of four concepts. Equation 4 computes the annual reference change during the calibration interval. Equation 5 computes the annual extrapolated change according to a Markov process. Equation 6 computes the annual simulated change according to a CA–Markov output map. Equation 7 computes the annual reference change during the extrapolation interval.

$$\text{Reference change during calibration interval} = \frac{\text{Change area from 1995 to 2007}}{12\text{years}} \tag{4}$$

$$\text{Extrapolated change via Markov} = \frac{\text{Markov change area from 2007 to 2018}}{11 \text{ years}} \tag{5}$$

$$\text{Simulated change via CA–Markov} = \frac{\text{Hits} + \text{Wrong Hits} + \text{ False Alarms from 2007 to 2018}}{11 \text{ years}} \tag{6}$$

$$\text{Reference change during extrapolation interval} = \frac{\text{Change area from 2007 to 2018}}{11 \text{ years}} \tag{7}$$

If the result from Equation 4 is greater than the result from Equation 5, then the Markov process extrapolates decelerating change, which a Markov process frequently does. If the result from Equation 5 equals the result from Equation 6, then IDRISI's CA–Markov is verified concerning how the module simulates the quantity of change. If the result from Equation 5 equals the result from Equation 7, then the reference data demonstrate stationarity during the calibration and extrapolation intervals concerning Markov proportions. If the result from Equation 4 equals the result from Equation 7, then the reference data demonstrate stationarity during the calibration and extrapolation intervals concerning the size of annual change.

Results

Figure 4 uses Equations 4 through 7 to generate a bar for each of the 120 runs. In Figure 4, each row of bars shows one data format for fifteen filters, and each column of bars shows one filter for eight data formats. Each bar has brown, blue, and orange segments. The sum of these segments is the annual reference change during the calibration interval, which Equation 4 expresses. The difference between Equations 4 and 5 is the orange segment, which is the deceleration of annual change that the Markov extrapolation implies. Equation 5 gives the sum of the brown and blue segments, which is the annual change according to matrix **A** from IDRISI's Markov module. Equation 6 computes each brown segment, which is the annual change in each simulation map. The difference between Equations 5 and 6 generates each blue segment, which indicates a deficit where the simulated change is less than the change in matrix **A**. A deficit implies a lack of verification of the CA–Markov module concerning the size of change. Figure 4 shows larger deficits with smaller filter sizes.

Figure 4 uses dotted rectangles to show the annual reference change during the extrapolation interval, which Equation 7 computes. Reference change during the extrapolation interval is slower than reference change during the calibration interval because change to the landscape decelerated. A Markov process implies deceleration, which the orange segments indicate. The dotted rectangles in Figure 4 are below the orange segments, which indicates that the deceleration in the landscape is more severe than in a

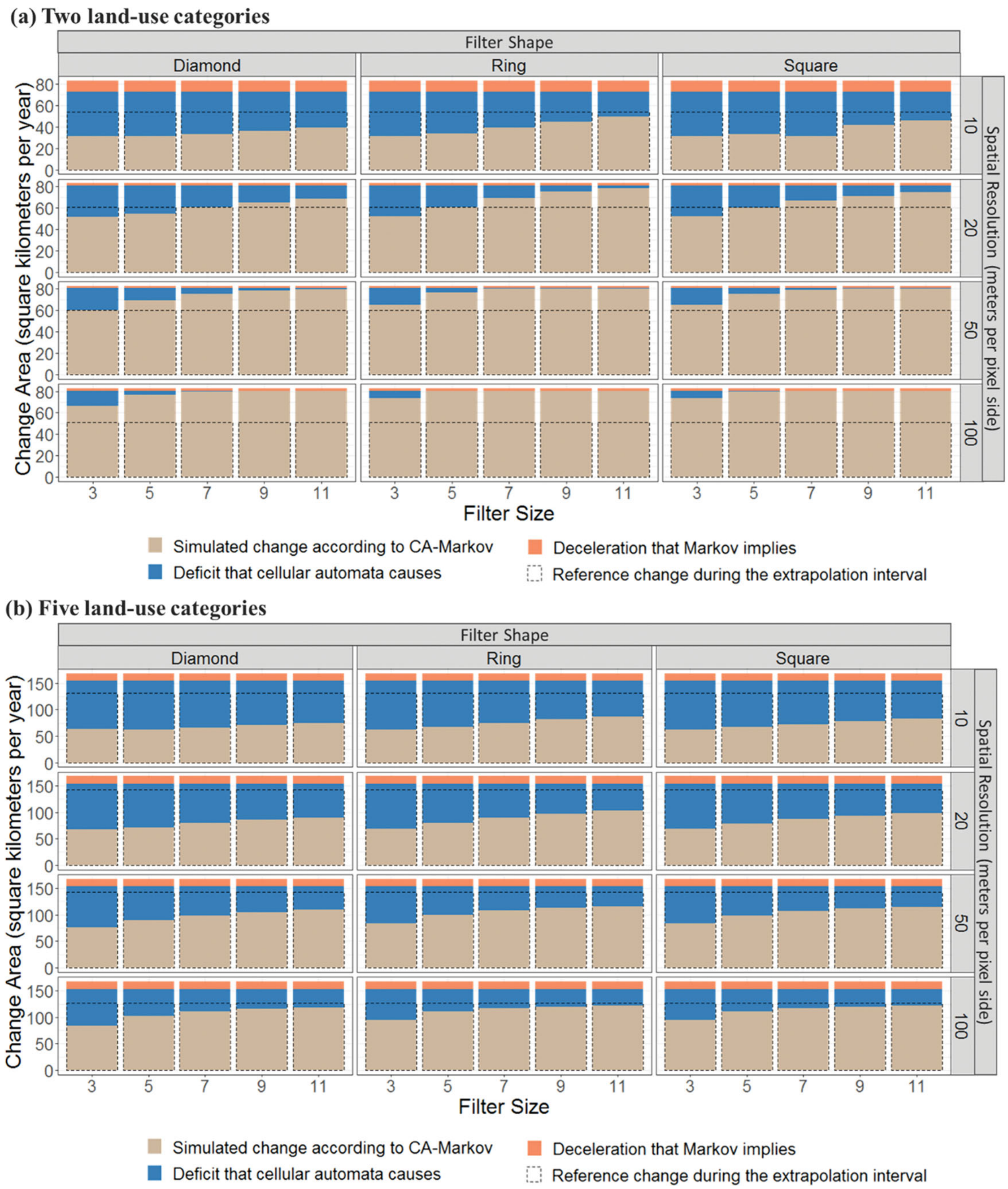

Figure 4. Annual change for 120 runs. Each row is one of eight data formats and each column is one of fifteen spatial filters. The height of each stacked bar is the change during the calibration interval.

Markov process. Verification revealed this conclusion because verification generated the orange and blue segments. If we had assumed that the module behaved as its documentation describes, then we would have compared the dotted rectangles to the brown segments. The brown segments in Figure 4B and the top row of Figure 4A are smaller than the dotted rectangles; therefore we would have concluded that the deceleration in a Markov process is more severe than in the landscape, which is the opposite of the truth.

More variation occurs down the columns of bars that reflect the variation in data format than across the rows of bars that reflect the variation in the filter. This indicates that a user's decisions concerning the data format are more influential than the decisions concerning the filter. Comparison across each row of graphs shows how the filter influences the size of the simulated change, but the filter is supposed to influence the allocation, not the quantity. This further reveals that the module does not behave as its documentation describes concerning the quantity of simulated change.

Figure 5 presents the misses, hits, wrong hits, and false alarms generated from the comparison of the reference map for 2007, the reference map for 2018, and each of the 120 simulated maps for 2018. Each run produces one bar. The layout in Figure 5 is the same as in Figure 4, meaning each row of bars shows one data format for fifteen filters and each column of bars shows one filter for eight data formats. Results vary most by the number of categories, then by the spatial resolution. Aggregation from five categories to two categories shrank all the components of validation. With five categories, FOMs ranged from 4.11 percent to 7.66 percent, implying the size of the overall error is approximately twenty-three to twelve times the size of hits. With two categories, FOMs ranged from 3.80 percent to 8.79 percent, implying the size of the overall error is twenty-five to ten times the size of hits. Coarser resolutions caused an increase in simulated change, thus an increase in false alarms.

If false alarms are smaller than misses, then the simulated change is smaller than the reference change during the extrapolation interval, which is the case for most of the runs in Figure 5B. The reason for the difference is CA's deficit that caused the simulated change to be less than the change that matrix **A** dictated, which the verification revealed in Figure 4B. If a user were to see Figure 5B without performing verification, then the user would conclude that the quantity error derives from how the Markov process dictates less change than the reference change during the extrapolation interval, but Figure 4B shows that the proper conclusion is just the opposite. If the simulation would have followed matrix **A** of the Markov extrapolation, then the simulation would have simulated more change than what occurred according to the reference data during the extrapolation time interval. Thus, the verification revealed information that prevents misinterpretation of Figure 5.

Figure 6 shows the spatial allocation of misses (12 percent), hits (2 percent), wrong hits (1 percent), false alarms (10 percent), and correct rejections (75 percent) where the percentages in parentheses indicate the percentage of the spatial extent. These results are for the run with five land categories at the 50 m resolution and a square filter of size 11×11. The filter causes the CA–Markov model to allocate the gain of a category around the patches where the category existed at the start of the simulation. The filter prevents CA–Markov from simulating a leapfrog pattern, where a category would gain at a place that is not spatially connected to the category's previous place. The FOM is 8 percent for the result in Figure 6.

Discussion

This section answers the four questions of our proposed evaluation framework to determine whether IDRISI's CA–Markov model is appropriate for our specific application. The first question is this: Can the user understand the model?

Our understanding required verification to give insights that we did not get from the software's documentation, despite CA–Markov in IDRISI being better documented than many models. If we did not test the module via verification, then we would not have realized that the model simulates a different quantity of change than what the Markov transition area matrix dictates. If users trust the documentation, then they are likely to misinterpret the results, especially the validation metrics. The validation's overall error is several times larger than the hits and the allocation error is more than twice the size of the quantity error for all the runs. Validation results tend to inspire users to modify the model's parameters to perform additional runs to reduce the error. If the user wants to reduce the allocation error, then the software's documentation would likely inspire the user to modify the filter. The filter's modification, however, influences both the allocation and quantity, so the filter's modification could influence the error in unpredictable ways. If the user wants to reduce the quantity error, then the software's documentation would likely inspire the user to modify the Markov matrix. The matrix's modification will not necessarily have the desired effect, though, because the spatial filter also influences the simulated quantity. For example, the 20-m resolution row in Figure 4A shows CA–Markov simulates less than the reference change for small filters but more than the reference change for large filters. Some of

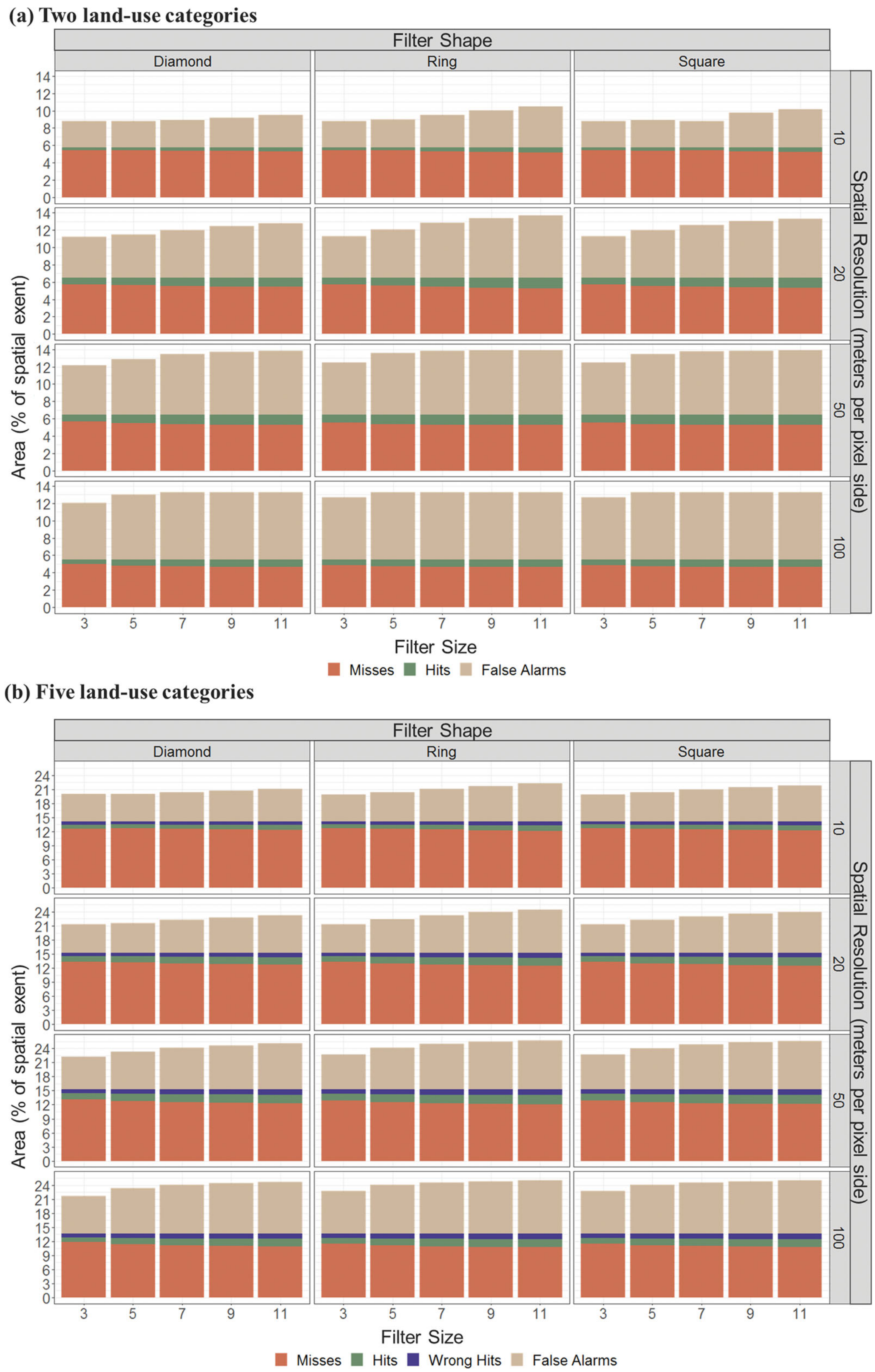

Figure 5. Components of validation for 120 runs.

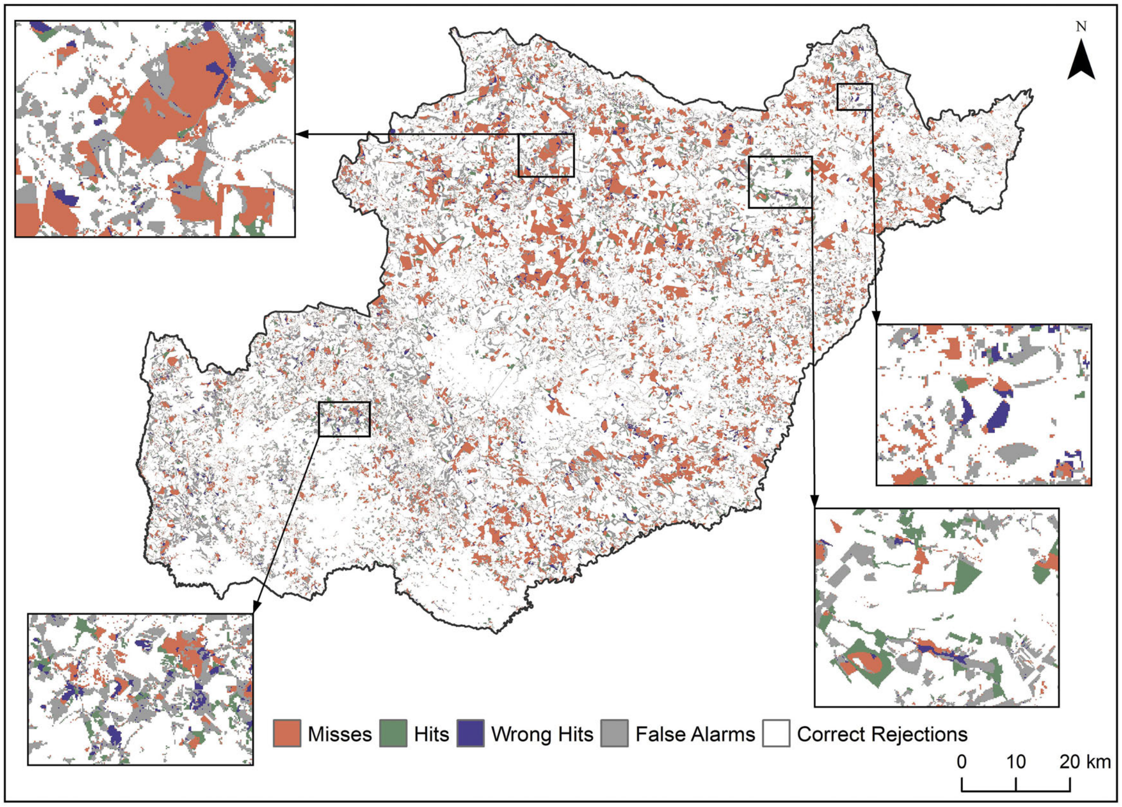

Figure 6. Map of the components of validation for one of the 120 runs.

our runs show zero quantity error, such as the run in the lower right corner of Figure 4B where the brown bar matches the dotted rectangle; but the reason for the match derives from the confusing deficit. If the user were to see zero quantity error in the validation, then the user would likely trust how the model extrapolates the quantity. Such trust would derive from a misunderstanding of the module's counterintuitive behavior, however. We do not know why CA–Markov fails to follow its Markov transition area matrix. The software dictates this characteristic of the module's behavior while users cannot see the computer code. Consequently, the answer to our first question is that users might think they understand the model, but their actual understanding of the module is likely to be unsatisfactory.

The second question is this: Can the audience understand the model? The answer is probably not, given the answer to the first question. In the vast literature that has used IDRISI's CA–Markov module, we have neither seen verification nor read a discussion that CA–Markov fails to follow the quantities that the Markov transition area matrix dictates, besides two manuscripts (Camacho Olmedo et al. 2015; Varga et al. 2019). If the user misinterprets the results, then the audience is also likely to misinterpret the results.

The third question is this: Can the user control the model? Our experience was not satisfactory. Users are likely to expect that Markov exclusively controls the quantity and CA exclusively controls the allocation. We found, though, that the CA influences both the allocation and the quantity. If the user wanted to use one quantity to portray scenarios of various allocations of change, then the user would likely modify the CA's parameters, and thus become frustrated or not realize that the CA's parameters influence the quantity of change. If the user cannot specify the quantity independently from the allocation, then the user cannot control the two important components necessary to create various scenarios of simulated change.

The last question is this: Does the model address the goals of the specific application? The answer to this question depends on the user's goals. We began the CA–Markov modeling to integrate knowledge regarding agricultural land system dynamics using GIS. We choose the CA–Markov model because authors routinely claimed that the simulation model was helpful to provide insights for effective environmental planning and well-informed policy decisions. After using the CA–Markov module, however, we found that we could not communicate helpful insights for landscape management, due mainly to the fact that the answers to the first three questions were not satisfactory. If we could control the quantity of simulated change, then we might have been able to use the model to express various scenarios, which could be potentially helpful to inspire discussions concerning management. We could not use the software to portray various scenarios, however, because the spatial filter influences the quantities in ways that we cannot control and that the software's documentation does not describe. Therefore, the model did not address the goals of our specific application. Our results expose how CA–Markov behaved for our application, which is similar to the behavior found by Camacho Olmedo et al. (2015) and Varga et al. (2019; Varga et al. 2020).

Given our findings, we wonder why so many authors continue to use CA–Markov (Guan et al. 2011; Aksoy and Kaptan 2021; Nyamekye et al. 2021) and claim that the model is successful and useful for sustainable management of environmental systems. We suspect those authors did not ask themselves the four questions. The results from the evaluation of the CA–Markov module reinforce the statements of other authors who have expressed concern for the reproducibility and replicability of model algorithms and geospatial data (Chrisman 1986; Tullis and Kar 2021). We wonder whether the CA–Markov behaved according to the users' interpretations in other applications integrated with other models (Gomes et al. 2019; Gharaibeh et al. 2020).

Future research should focus on the effectiveness and limitations of our framework. Mas et al. (2014) found our framework helpful to formulate their discussion, but we have not yet solicited feedback from modelers concerning whether or how they would modify their actions based on their answers to our questions. These four questions apply to any model, so a next step is to apply the questions to models other than IDRISI's CA–Markov model.

Conclusions

Not satisfactory is the answer to all four questions that our framework posed to evaluate a CA–Markov simulation module for a case study. This article combined sensitivity analysis with verification to find that the module's behavior did not match the description in its documentation. We could neither understand nor control the model because the CA–Markov module simulates a quantity of change different from what the Markov transition area matrix dictates. If the purpose of the model is to allow the ability to portray various scenarios of the quantity of change, then we would have had difficulty because we could not control the quantity of simulated change. If the purpose of the model is to simulate change accurately, then validation is important, but the module's confusing behavior is likely to cause users to misinterpret the validation metrics.

CA–Markov remains popular despite its challenges. Clark Labs introduced the CA–Markov module decades ago as an experimental module. CA–Markov's most recent documentation encourages users to switch to the newer Land Change Modeler (LCM), which does not have the spatial filter that influences how the CA–Markov module loses control of the simulated quantity. LCM has more detailed documentation and a more sophisticated implementation of the Markov matrix. LCM does not generate the deficits in the quantity of simulated change that CA–Markov sometimes does. The next steps should be to evaluate LCM and other models concerning the quality of the documentation, verification, and reaction to our four questions. We encourage users to ask themselves our four questions to determine the appropriateness of any model.

Acknowledgments

We thank the GEOMODLAB—Laboratory for Remote Sensing, Geographical Analysis and Modelling—of the Centre for Geographical Studies (CEG) and Institute of Geography and Spatial Planning (IGOT) for providing the required equipment and software.

Funding

The Portuguese Foundation for Science and Technology (FCT) supported this research via Grant No. SFRH/BD/115497/2016 to Cláudia M. Viana. The Centre for Geographical Studies—Universidade de Lisboa and FCT supported this research via Grant No. UIDB/00295/2020 + UIDP/00295/2020. The National Science Foundation of the United

States supported this research via its Long-Term Ecological Research network via Grants OCE-1637630 and OCE-2224608 for Plum Island Ecosystems.

Disclosure Statement

No potential conflict of interest was reported by the authors.

ORCID

Cláudia M. Viana http://orcid.org/0000-0001-6858-4522
Robert Gilmore Pontius, Jr. http://orcid.org/0000-0001-7287-5875
Jorge Rocha http://orcid.org/0000-0002-7228-6330

References

Aksoy, H., and S. Kaptan. 2021. Monitoring of land use/land cover changes using GIS and CA-Markov modeling techniques: A study in northern Turkey. *Environmental Monitoring and Assessment* 193 (8):1–21. doi: 10.1007/s10661-021-09281-x.

Aliani, H., M. Malmir, M. Sourodi, and S. B. Kafaky. 2019. Change detection and prediction of urban land use changes by CA–Markov model (Case study: Talesh County). *Environmental Earth Sciences* 78 (17):546. doi: 10.1007/s12665-019-8557-9.

Chrisman, N. 1986. Obtaining information on quality of digital data. In *AutoCarto: Proceedings of the International 1986 Symposium*. https://cartogis.org/docs/proceedings/archive/auto-carto-london-vol-1/pdf/obtaining-information-on-quality-of-digital-data.pdf.

Dahal, K. R., and T. E. Chow. 2015. Characterization of neighborhood sensitivity of an irregular cellular automata model of urban growth. *International Journal of Geographical Information Science* 29 (3):475–97. doi: 10.1080/13658816.2014.987779.

DGT. 2018. *Especificações Técnicas da Carta de Uso e Ocupação Do Solo (COS) de Portugal Continental Para 1995, 2007, 2010 e 2015* [Technical specification for land use land cover data for mainland Portugal of 1995, 2007, 2010 and 2015]. Lisbon, Portugal: DGT. http://mapas.dgterritorio.pt/atom-dgt/pdf-cous/COS2015/ET-COS-1995-2007-2010-2015.pdf.

Eastman, J. R. 2006. *Guide to GIS and image processing*, vol. 1. Worcester, MA: Clark University.

Eastman, J. R. 2012. *IDRISI Selva manual*. Worcester, MA: Clark University.

Eastman, J. R., and J. Toledano. 2018. A short presentation of CA_MARKOV. In *Geomatic approaches for modeling land change scenarios: Lecture notes in geoinformation and cartography*, ed. M. T. C. Olmedo, M. Paegelow, J. F. Mas, and F. Escobar, 481–84. Cham, Switzerland: Springer International. doi: 10.1007/978-3-319-60801-3_33.

Gharaibeh, A., A. Shaamala, R. Obeidat, and S. Al-Kofahi. 2020. Improving land-use change modeling by integrating ANN with cellular automata-Markov chain model. *Heliyon* 6 (9):e05092. doi: 10.1016/j.heliyon.2020.e05092.

Ghosh, P., A. Mukhopadhyay, A. Chanda, P. Mondal, A. Akhand, S. Mukherjee, S. K. Nayak, S. Ghosh, D. Mitra, T. Ghosh, et al. 2017. Application of cellular automata and Markov-chain model in geospatial environmental modeling—A review. *Remote Sensing Applications: Society and Environment* 5:64–77. doi: 10.1016/j.rsase.2017.01.005.

Gomes, E., P. Abrantes, A. Banos, and J. Rocha. 2019. Modelling future land use scenarios based on farmers' intentions and a cellular automata approach. *Land Use Policy* 85 (June):142–54. doi: 10.1016/j.landusepol.2019.03.027.

Guan, D., H. Li, T. Inohae, W. Su, T. Nagaie, and K. Hokao. 2011. Modeling urban land use change by the integration of cellular automaton and Markov model. *Ecological Modelling* 222 (20–22):3761–72. doi: 10.1016/j.ecolmodel.2011.09.009.

Houet, T., P. H. Verburg, and T. R. Loveland. 2010. Monitoring and modelling landscape dynamics. *Landscape Ecology* 25 (2):163–67. doi: 10.1007/s10980-009-9417-x.

Liao, J., L. Tang, G. Shao, X. Su, D. Chen, and T. Xu. 2016. Incorporation of extended neighborhood mechanisms and its impact on urban land-use cellular automata simulations. *Environmental Modelling & Software* 75:163–75. doi: 10.1016/j.envsoft.2015.10.014.

Lin, J., X. Li, S. Li, and Y. Wen. 2020. What is the influence of landscape metric selection on the calibration of land-use/cover simulation models? *Environmental Modelling & Software* 129:104719. doi: 10.1016/j.envsoft.2020.104719.

Mas, J.-F., M. Kolb, M. Paegelow, M. Olmedo, and T. Houet. 2014. Inductive pattern-based land use/cover change models: A comparison of four software packages. *Environmental Modelling & Software* 51:94–111. doi: 10.1016/j.envsoft.2013.09.010.

Memarian, H., S. Balasundram, J. Talib, C. Sung, A. Sood, and K. Abbaspour. 2012. Validation of CA-Markov for simulation of land use and cover change in the Langat Basin, Malaysia. *Journal of Geographic Information Systems* 4 (6):542–54. doi: 10.4236/jgis.2012.46059.

Muller, M. R., and J. Middleton. 1994. A Markov model of land-use change dynamics in the Niagara Region, Ontario, Canada. *Landscape Ecology* 9 (2):151–57. doi: 10.1007/BF00124382.

Noszczyk, T. 2019. A review of approaches to land use changes modeling. *Human and Ecological Risk Assessment: An International Journal* 25 (6):1377–1405. doi: 10.1080/10807039.2018.1468994.

Nüst, D., and E. Pebesma. 2021. Practical reproducibility in geography and geosciences. *Annals of the American Association of Geographers* 111 (5):1300–10. doi: 10.1080/24694452.2020.1806028.

Nyamekye, C., S. Kwofie, E. Agyapong, S. A. Ofosu, R. Arthur, and L. B. Appiah. 2021. Integrating support vector machine and cellular automata for modelling land cover change in the tropical rainforest under equatorial climate in Ghana. *Current Research in Environmental Sustainability* 3:100052. doi: 10.1016/j.crsust.2021.100052.

Olmedo, M. T. C., R. G. Pontius, Jr., M. Paegelow, and J. F. Mas. 2015. Comparison of simulation models in terms of quantity and allocation of land change. *Environmental Modelling & Software* 69:214–21. doi: 10.1016/j.envsoft.2015.03.003.

Pan, Y., A. Roth, Z. Yu, and R. Doluschitz. 2010. The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. *Computers, Environment and Urban Systems* 34 (5):400–08. doi: 10.1016/j.compenvurbsys.2010.03.003.

Pontius, R. G., Jr., W. Boersma, J.-C. Castella, K. Clarke, T. de Nijs, C. Dietzel, Z. Duan, E. Fotsing, N. Goldstein, K. Kok, et al. 2008. Comparing the input, output, and validation maps for several models of land change. *The Annals of Regional Science* 42 (1):11–37. doi: 10.1007/s00168-007-0138-2.

Pontius, R. G., Jr., J.-C. Castella, T. Nijs, Z. Duan, E. Fotsing, N. Goldstein, K. Kok, E. Koomen, C. Lippitt, W. McConnell, et al. 2018. Lessons and challenges in land change modeling derived from synthesis of cross-case comparisons. In *Trends in spatial analysis and modeling: Geotechnologies and the environment*, vol. 19, ed. M. Behnisch and G. Meinel, 143–64. Cham, Switzerland: Springer International. doi: 10.1007/978-3-319-52522-8_8.

Pontius, R. G., Jr., S. Peethambaram, and J.-C. Castella. 2011. Comparison of three maps at multiple resolutions: A case study of land change simulation in Cho Don District, Vietnam. *Annals of the Association of American Geographers* 101 (1):45–62. doi: 10.1080/00045608.2010.517742.

Pontius, R. G., Jr., R. Walker, R. Yao-Kumah, E. Arima, S. Aldrich, M. Caldas, and D. Vergara. 2007. Accuracy assessment for a simulation model of Amazonian deforestation. *Annals of the Association of American Geographers* 97 (4):677–95. doi: 10.1111/j.1467-8306.2007.00577.x.

Roodposhti, M. S., J. Aryal, and B. A. Bryan. 2019. A novel algorithm for calculating transition potential in cellular automata models of land-use/cover change. *Environmental Modelling & Software* 112:70–81. doi: 10.1016/j.envsoft.2018.10.006.

Sankarrao, L., D. K. Ghose, and M. Rathinsamy. 2021. Predicting land-use change: Intercomparison of different hybrid machine learning models. *Environmental Modelling & Software* 145:105207. doi: 10.1016/j.envsoft.2021.105207.

Statistics Portugal (INE). 2021. *XVI Population Census, VI Housing Census*. Lisbon. I.

Torrens, P. M. 2000. *How cellular models of urban systems work (1. Theory). Centre for Advanced Spatial Analysis Working Paper* 28, University College, London.

Tullis, J. A., and B. Kar. 2021. Where is the provenance? Ethical replicability and reproducibility in GIScience and its critical applications. *Annals of the American Association of Geographers* 111 (5):1318–28. doi: 10.1080/24694452.2020.1806029.

van Vliet, J., A. K. Bregt, D. G. Brown, H. van Delden, S. Heckbert, and P. H. Verburg. 2016. A review of current calibration and validation practices in land-change modeling. *Environmental Modelling & Software* 82:174–82. doi: 10.1016/j.envsoft.2016.04.017.

Varga, O. G., R. G. Pontius, Jr., S. K. Singh, and S. Szabó. 2019. Intensity analysis and the figure of merit's components for assessment of a cellular automata–Markov simulation model. *Ecological Indicators* 101:933–42. doi: 10.1016/j.ecolind.2019.01.057.

Varga, O. G., R. G. Pontius, Jr., Z. Szabó, and S. Szabó. 2020. Effects of category aggregation on land change simulation based on corine land cover data. *Remote Sensing* 12 (8):1314. doi: 10.3390/rs12081314.

Verburg, P. H., T. C. M. de Nijs, J. R. van Eck, H. Visser, and K. de Jong. 2004. A method to analyse neighbourhood characteristics of land use patterns. *Computers, Environment and Urban Systems* 28 (6):667–90. doi: 10.1016/j.compenvurbsys.2003.07.001.

Verweij, P., A. Cormont, K. Kok, M. van Eupen, S. Janssen, J. Te Roller, W. de Winter, M. Pérez-Soba, and I. G. Staritsky. 2018. Improving the applicability and transparency of land use change modelling: The ICLUE model. *Environmental Modelling & Software* 108:81–90. doi: 10.1016/j.envsoft.2018.07.010.

Viana, C. M., I. Girão, and J. Rocha. 2019. Long-term satellite image time-series for land use/land cover change detection using refined open source data in a rural region. *Remote Sensing* 11 (9):1104. doi: 10.3390/rs11091104.

Viana, C. M., and J. Rocha. 2020. Evaluating dominant land use/land cover changes and predicting future scenario in a rural region using a memoryless stochastic method. *Sustainability* 12 (10):4332. doi: 10.3390/su12104332.

White, R., and G. Engelen. 2000. High resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems* 24 (5):383–400. doi: 10.1016/S0198-9715(00)00012-0.

Wilson, J. P., K. Butler, S. Gao, Y. Hu, W. Li, and D. J. Wright. 2021. A five-star guide for achieving replicability and reproducibility when working with GIS software and algorithms. *Annals of the American Association of Geographers* 111 (5):1311–17. doi: 10.1080/24694452.2020.1806026.

Wolfram, S. 1984. Cellular automata as models of complexity. *Nature* 311 (5985):419–24. doi: 10.1038/311419a0.

CLÁUDIA M. VIANA is a Postdoctoral Junior Researcher at the Center for Geographical Studies (CEG), University of Lisbon, Lisbon, Portugal. E-mail: claudiaviana@edu.ulisboa.pt. Her research interests include modeling the spatiotemporal changes of agricultural land use, as well as the spatial distribution over time of agricultural production. Key research interests include GIS packages, computer programming languages, geospatial data, historical data, time-series, artificial intelligence, computer-based modeling, and spatiotemporal analysis.

ROBERT GILMORE PONTIUS, JR. is a Professor in the School of Geography, Clark University, Worcester, MA 01610. E-mail: rpontius@clarku.edu. His areas of expertise include GIScience, land change science, simulation modeling, and spatial statistics.

JORGE ROCHA is an Assistant Professor in the Institute of Geography and Spatial Planning, University of Lisbon, Lisbon, Portugal. E-mail: jorge.rocha@campus.ul.pt. His research interests include geosimulation and geocomputation using, among other methods, machine learning, graphs theory, cellular automata, and multiagent systems.