Chapter 9
The Importance of Scale in Land Use Models: Experiments in Data Conversion, Data Resampling, Resolution and Neighborhood Extent

J. Díaz-Pacheco, H. van Delden and R. Hewitt

Abstract The investigation and modeling of land use dynamics can be conducted at different scales based on the objective of the study. However, few studies have looked at comparing various scale aspects, such as spatial resolution and the related neighborhood effect, for practical case study applications. In this chapter, we contribute to this under-explored area with a detailed study of how changes in the data preparation procedures and the scale decisions made in setting up a land use model can affect its performance. For these purposes we used a Cellular Automata (CA) based land use model, which we applied to the Madrid region in Spain. In order to discover the most appropriate method for preparing input data, different vector-to-raster conversion and resampling strategies were tested with reference to 4 statistics. For vector-to-raster conversion, the cell center method was found to give the best results across all of the statistics. Furthermore, direct conversion from the original vector map to raster format at the desired cell size was found to give better results than resampling to the desired cell size from a different cell size. We also tested the effect of changing spatial resolution and cell neighborhood distance on a model's goodness-of-fit to real data using a range of location and pattern metrics. Although differences were noted in the simulations, all the applications fitted the

J. Díaz-Pacheco (✉)
Universidad Complutense de Madrid, Madrid, Spain
e-mail: jdiazpac@ull.es

J. Díaz-Pacheco
Cátedra RRD. Ciudades Resilientes, Universidad de La Laguna,
San Cristóbal de La Laguna, Spain

H. van Delden
Research Institute for Knowledge Systems, Maastricht, The Netherlands
e-mail: hvdelden@riks.nl

H. van Delden
The University of Adelaide, Adelaide, Australia

R. Hewitt
Observatorio para una Cultura del Territorio, Madrid, Spain
e-mail: richardjhewitt@hushmail.com

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_9

data satisfactorily. Nevertheless, the 50 × 50 m cell resolution applications were visually much more realistic, perhaps because this resolution was used in the initial calibration of the model. The results indicate that data conversion issues have a major effect on the quality of the input data. Additionally, models of this type appear to be much less sensitive to scale changes, either through cell resolution changes, neighborhood changes, or both, than is usually suggested by the literature.

Keywords Land use models · Land use change · Scale · Cellular automata · Data conversion

1 Introduction

Researchers in land use/land cover change (hereafter LUCC) modeling approach their research with different objectives for different regions, and as a result work at different cartographic spatial scales. The observation and modeling of spatial phenomena should be carried out at a scale appropriate to the phenomena in question (Woodcock and Strahler 1987). However, the same phenomenon can be modelled at various scales depending on the spatial context of the analysis. When investigating the urbanization processes in a specific city, a high level of spatial detail might be required, because urban land use units (facilities, parks, residential areas, etc.) are small in size and specific local information about socio-economic and physical characteristics and planning regulations can play a key role in understanding dynamics. However, for modeling the urban growth of Europe, the analysis should not be too detailed or the generality of the process at the continental scale could be missed. The geographical scale of models depends therefore on both the phenomenon to be modelled and the spatial scale (local, regional, national, global…) of the analysis. From a geographical point of view, variability of scale can be regarded as both a strength and a weakness of the discipline (Lam and Quattrochi 1992). LUCC models, as geographical tools to understand phenomena (Longley and Batty 2003 p. 5) and provide policy support (Van Delden et al. 2010), are clearly subject to the same considerations. In fact, the results of dynamic spatial models are strongly influenced by scale, and results derived from a model developed for one spatial scale may not be applicable at another. When moving from one scale to another, land use patterns may disappear or emerge (de Koning et al. 1998), significant processes may lose their significance, and rates of change may vary (Kok and Veldkamp 2001). The Modifiable Areal Unit Problem (MAUP), a concept which describes '*the variation in results that can often be obtained when data for one set of areal units are progressively aggregated into fewer and larger units for analysis*' (Openshaw and Taylor 1979; Openshaw 1983) helps to understand some of these key issues. On the other hand, variability of scale could be considered an advantage since data and information can be adapted to suit the context in which the analyzed process are occurring (i.e. operational scale), or to take into account the way humans may perceive different spaces at differing levels of detail in a hierarchical fashion according to their proximity (Van Vliet et al. 2009).

Although the importance of scale in LUCC models is recognized by many researchers (e.g. Jenerette and Wu 2001; Theobald and Hobbs 1998; Ménard and Marceau 2005; van Delden et al. 2011), experimental studies in which the implications of different scale options are directly compared, are generally lacking (Jantz and Goetz 2005). In this paper, we contribute to this under-explored topic. We assess the accuracy and quality of the data when they are converted from vector to raster or resampled by different methods, to be used on spatially-explicit land use models. In particular, we explore the data conversion and rasterization options, and the impact of spatial resolution on the calibration of a land use model for Madrid.

The paper is organized as follows. First, we present the background and the method applied for each of the two components of this study, data preparation and model resolution, followed by details of the application including the GIS and the land use model used in this study. Finally we present and discuss the results, draw the relevant conclusions and make recommendations for further research.

2 Finding the Right Scale

2.1 Background

Although several authors have proposed ways to find the best scale for setting up raster-based LUCC models (e.g. Tobler 1988; Lam and Quattrochi 1992), no single widely-agreed method has emerged, and the final decision is usually taken on the basis of the researcher's own specialist knowledge. Such a decision may not always be the result of a rigorous procedure, but is not usually arbitrary. For example, in policy-relevant LUCC models, the decision about which scale to use is often a trade-off between the scale required by intended users, the scale at which processes are best represented, and practical considerations like data availability or computational resources (van Delden et al. 2011).

The scale decisions a LUCC modeler needs to make include the spatial and temporal extent, the spatial level(s) and the hierarchy by which they are ordered, and the amount of detail incorporated. Levels refer to locations along a scale (Gibson et al. 2000) and detail relates to the spatial, temporal and thematic resolution(s) and the complexity by which processes are represented (van Delden et al. 2011). When focusing on CA based land use models, other important factors related to spatial resolution, such as neighborhood size and type, must also be taken into account (Ménard and Marceau 2005).

Previous work on the effect of changing scale in a LUCC model for Central America by Kok and Veldkamp (2001) found that coarsening the resolution from 15×15 km to 75×75 km led to improved model explanatory power (r^2), but did not significantly affect the explanatory variables (i.e. land change drivers identified were broadly the same at both resolutions tested). However, changing the extent of the model produced a strong variation in performance (poorer fit for all Central America, better fit for individual countries). Though these authors do not say so explicitly, this is an excellent example of the scale problem (Openshaw 1983) in a land use model, since the land change dynamics modelled by these authors relate to national, not supra-national drivers, and are not generalizable across borders. Though it is common practice in land use models to work at larger spatial extents, the findings of these authors are a clear warning of the perils that this may entail. However, various modeling approaches overcome this issue by dividing the modelled area into smaller subdivisions, as for example in the case of the regional model incorporated into the Metronamica Modeling framework (RIKS 2014).

Jantz and Goetz (2005) investigated the behavior of different types of urban growth rules at different cell sizes in the popular SLEUTH model, concluding that cell resolution was a major determinant in model performance and that some types of urban growth rule produced much more growth at coarser resolutions than at finer ones. Though their findings are quite specific to the SLEUTH model, the implication is that neighborhood effects for urban land, which are fundamental in all CA models, may vary non-linearly across scales.

Ménard and Marceau (2005) observed how changing the size of the neighborhood radius and the resolution produced a non-linear relationship between the spatial scale and the simulation results. Their work was based on a dataset derived from remote-sensing images for two time periods and focused on land cover change. The study area was dominated by forest and agriculture, so the phenomenon of urban expansion was not considered (Ménard and Marceau 2005). Samat (2006) undertook sensitivity analysis of a CA-based urban model with the aim of finding the appropriate scale for the modelled region (Seberang Perai, Malaysia). The study found that the model performed well at 30, 90, and 270 m cell resolution, but at coarser resolutions (810, 2430 m), accuracy declined rapidly. These findings appear to contradict the findings of Kok and Veldkamp. However, these studies are difficult to compare for a number of reasons. Firstly, Kok and Veldkamp compared only two resolutions, while Samat investigated five. Secondly, the studies do not compare the same cell resolutions and address different spatial extents. Thirdly, the statistical comparison methods used were quite different (Kok and Veldcamp used the coefficient of determination (R^2) of a regression model, whereas Samat used cell-by-cell map comparison techniques). Finally, Samat employed standard Kappa for comparing real and simulated maps, an approach which has since been found to be inadequate (Pontius and Millones 2011; Van Vliet et al. 2011).

As an aid to determining the appropriate spatial scale for the general case, Samat's work (2006) has some limitations. On the one hand, the analysis comprised only two land use classes (urban and non-urban), so the type of urban land use was not a determining factor for selecting the scales for the tests. Moreover, the land use dataset employed was drawn from different sources for each of the two time periods (1990 and 1998). In addition, the cartographic scale chosen for the smallest cell resolution tested (30×30 m) does not seem to respect, at least for 1990, the general rules for transformation of a scaled vector map (1:75,000) to a raster map (see Tobler 1988).

The various studies show that the choice of scale, and, in particular, of the spatial resolution, is key in setting up a land use model, as these can have a large impact on the model results. With limited work being carried out in urban environments, this paper aims to contribute to an enhanced understanding in this area by exploring the effects of spatial resolution and neighborhood extents on a land use model's capacity to simulate land use change.

3 Approach

3.1 Scale in Geography and Remote Sensing

Three techniques of land use data conversion from vector to raster, and two techniques of aggregation by resampling from a high cell resolution to a lower one are tested. The data conversion and resampling techniques used are those implemented inside the popular ArcGIS 10.0 software. ArcGIS was chosen because it is widely used and provides a detailed description of the procedures in the user manual. Testing was undertaken by developing a series of land use maps as input data for a LUCC model generated by each technique and then comparing the results using statistical map comparison algorithms. In the following section we describe the data conversion and cell aggregation methods used to obtain the most appropriate data for use in the LUCC applications at different resolutions, together with the metrics used to evaluate the maps generated by the various techniques.

3.1.1 Vector-to-Raster Conversion

In the vector-to-raster conversion, some loss of accuracy is unavoidable, due to classification errors where the irregular polygon boundary coincides with a regular grid (Carver and Brunsdon 1994). Three techniques implemented in ArcGIS 10.0 for direct conversion from a vector polygon coverage to a regular grid were analyzed, namely Cell Center, Maximum Area and Maximum Combined Area (the names used in the software) (Fig. 1).

Using the cell center (Cc) algorithm the final categorical value of every cell in the grid is the attribute value which coincides with the center of the cell. In the case of the maximum area algorithm (Ma) the final value of the cell is established by assigning the value of the largest polygon coincident with the cell. The maximum combined area algorithm (Mca) works in a similar way to the Cc algorithm, except that the value of the cell is taken from the total area of different polygons with the same attributes coincident with the cell.

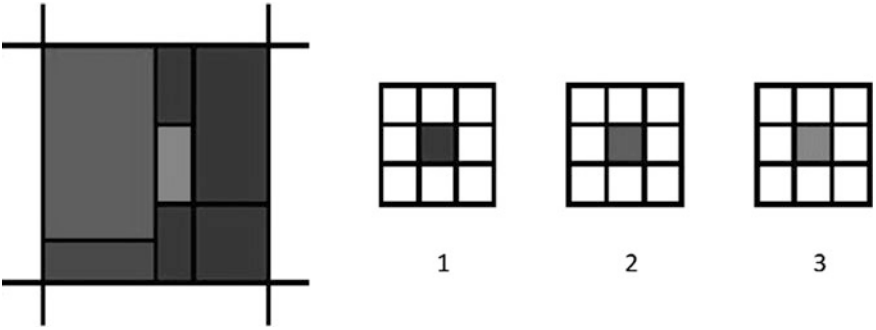
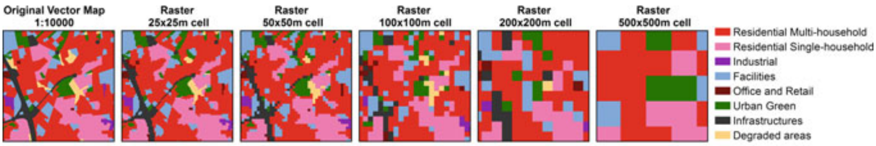


Fig. 1 Vector polygon to raster. (*1*) Mca. Maximum Combined Area; (*2*) Ma. Maximum Area; (*3*) Cc. Cell center. *Source* Adapted from ESRI (2010)



Absolute variation on urban land patches

Land Use	**Vector**	**raster 25**	**raster 50**	**raster 100**	**raster 200**	**raster 500**
Facilities	5887	4546	4188	2991	1360	346
Industrial	1931	1747	1778	1463	848	251
Office and Retail	540	555	547	381	198	55
Residential Multi-household	5512	1563	1552	1228	663	217
Residential Single-household	4538	3380	3620	2897	1695	575
Urban Green	2038	2214	2429	1624	678	195
Total land patches	20446	14005	14114	10584	5442	1639
% of missed patches		31.50	30.97	48.23	73.38	91.98

Relative variation on urban land patches (%)

Land Use	**raster 25**	**raster 50**	**raster 100**	**raster 200**	**raster 500**
Facilities	22.78	28.86	49.19	76.90	94.12
Industrial	9.53	7.92	24.24	56.08	87.00
Office and Retail	-2.78	-1.30	29.44	63.33	89.81
Residential Multi-household	71.64	71.84	77.72	87.97	96.06
Residential Single-household	25.52	20.23	36.16	62.65	87.33
Urban Green	-8.64	-19.19	20.31	66.73	90.43

Fig. 2 Variation on urban land patches after conversion vector information and resample up to a resolution of 500 × 500 m

Vector-to-raster conversions were performed from an original land use vector dataset to a grid of 25 × 25 m, 50 × 50 m, 100 × 100 m resolution successively. We considered that at lower resolution the general land use structure is missed (Fig. 2).

The most detailed resolution (25 × 25 m) was selected following recommendations given by Switzer (1975) in which 50% of the area of the cell should be larger than the smallest mapped polygon. In the MLU geodatabase the smallest mapped polygon is 30.4 m^2 and 50% of a 25 × 25 m cell is 312.5 m^2, which complies with the requirements of this rule.

3.1.2 Resampling

In the same way as for the vector-to-raster conversion analysis, for the resampling of grid maps, two techniques, Nearest Neighborhood Assignment and Majority, implemented in the software ArcGIS 10.0 were applied consecutively. The former assigns the categorical value to the new cell according to the value of the cell closest to the center of the new cell and the latter assigns the most popular values of the cells in the input map that fall inside the new cell in the output map. A simple example of both algorithms is shown in Fig. 3.

For the nearest neighborhood assignment method, the maximum spatial error must be one-half of the cell size, while for the majority method the results of the resampling tend to create higher compactness (ESRI 2010).

The techniques were applied to the 25 × 25 m raster map obtained from the vector polygon land use map using the vector to raster method that provided the best results. Aggregations were carried out into grids with 50 × 50 m and 100 × 100 m cell sizes, each one from the 25 × 25 m raster map.

3.1.3 Assessment Procedure for Vector to Raster Conversion and Resampling

Comparisons of the maps resulting from application of the various shape-to-raster and aggregation techniques were carried out at 25 m, 50 m and 100 m resolution. In order to compare 50 m and 100 m resolution maps, all the resulting maps were disaggregated to a 25 m resolution.

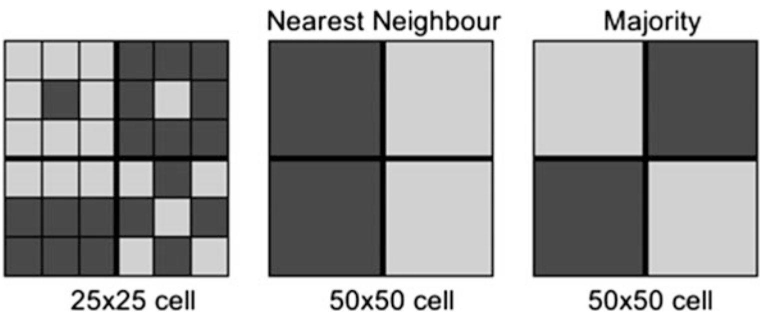


Fig. 3 Different techniques for resampling. Nearest Neighbour and Majority. *Source* Adapted from ESRI (2010)

The similarity of the land use maps resulting from the different conversion and resampling methods was analyzed using metrics to assess the similarity in location and the similarity in the resulting landscape pattern.

3.2 Assessing the Impact of Spatial Resolution and Neighborhood Extent on the LUCC Model

In order to evaluate the effects of the spatial resolution and the size of the neighborhood, a set of applications was developed, in which the cell size and neighborhood were varied while the extent remained constant. To keep the work manageable, it was decided to apply the model at three different spatial resolutions. Resolutions of 25 × 25 m, 50 × 50 m, and 100 × 100 m were selected based on the urban context and the authors' interest in investigating whether higher spatial resolutions, possible due to the availability of detailed land use data sets, would also result in improved model calibration and validation results.

We began by developing an application at 50 m, with a neighborhood of 8 cells (400 m). This application was calibrated over a first historic time period and validated over a second. Once this application was considered suitable for reproducing the (historic) land use dynamics, some of its scale characteristics were modified in order to evaluate their effects on the model results. To this end, two applications were developed with a modified resolution of the cells (25 m and 100 m), using the most appropriate methods found for data preparation, while maintaining the cell radius for the neighborhood effect at 8 cells. Next, two additional applications at 25 m and 100 m cell-resolution applications were created with respectively larger (16 cells) and smaller (4 cells) radii, so as to maintain the equivalent cell neighborhood distance as in the original 50 m application. All the applications were run using the same parameter settings employed in the original 50 m application (Table 1).

As with the assessment of the different conversion and resampling methods, the results of the calibration and validation have been analyzed using metrics for assessing similarity in location and in the resulting landscape pattern.

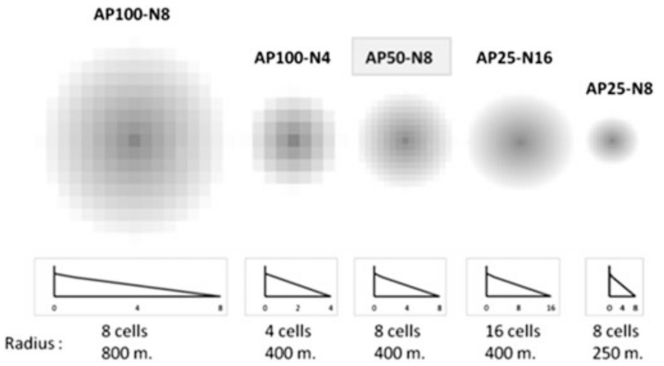
4 Applications

4.1 Study Area

The area selected for analysis is the Madrid region (Fig. 4), an area of around 6 million inhabitants. This region was chosen because of the large increase in urban development that it has experienced over recent decades (until the beginning of the

Table 1 Scale changes on neighborhood for each application

Resolution	100 × 100 Doubled resolution	100 × 100 Doubled resolution	50 × 50 Original	25 × 25 Halved resolution	25 × 25 Halved resolution
Feature of changes	Doubled resolution Equal radius in cells Unequal radius in meters	Doubled resolution Unequal radius in cells Equal radius in meters		Halved resolution Unequal radius in cells Equal radius in meters	Halved resolution Equal radius in cells Unequal radius in meters
Cell radius	8	4	8	16	8
Meter radius	800	400	400	400	200
Number of cells	196	48	196	796	196
Area in m^2	1,960,000	480,000	490,000	497,500	122,500



current economic crisis around 2008) and because a highly detailed land use database documenting this change has recently become available (Díaz-Pacheco and García Palomares 2014).

The expansion of urban land use in the Madrid metropolitan area during the 1990s was extraordinary, at least by European standards. According to CORINE land cover (EEA 2014), artificial land cover increased by more than 30,000 ha, an annual growth rate of 4.77%, while over the same period, the population of around 6 million grew by only 0.8% a year. Furthermore, over this decade the area under construction (mines, dumps, and construction sites) grew by 200% (Rocha et al. 2009; Hewitt and Escobar 2011). This growth in urban land, in a situation of demographic stability, produced a notable increase in the amount of artificial land per person, which in only 5 years (1996–2001) shot up from 153 to 179 m^2 per inhabitant (de Lucio 2011).

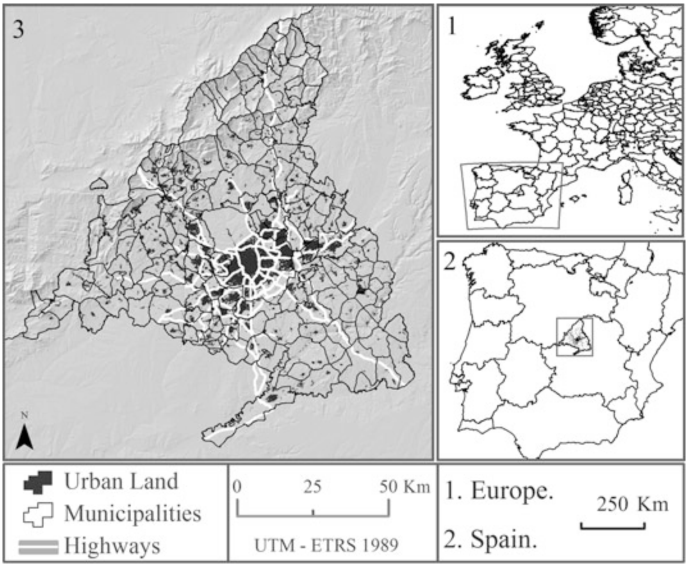


Fig. 4 Location of Madrid Region. *Source* Díaz-Pacheco and Gutiérrez (2013)

4.2 Land Use Data Set

Madrid Land Use (MLU) is a cartographic database with land use and land cover information for the Madrid Region, covering the time periods 2000, 2006 and 2009. The MLU dataset comprises 22 land use classes of which 7 are urban. Mapping was undertaken at a highly detailed basic reference scale of 1:10,000. The technical process did not include automatic or computer-assisted classification tasks, and the mapping work was undertaken entirely by photo-interpretation of high resolution (0.5 m) aerial orthophotographs, supported by large scale cartographic and cadastral information (1:5,000 and 1:1,000, respectively). Identical criteria were used for the digitization and thematic classification for each of the land use dataset periods. MLU clearly represents an excellent cartographic dataset for assessing urban land use in Madrid and outperforms CORINE land cover in this area in a number of respects (see Díaz-Pacheco and Gutiérrez 2013).

4.3 Land Use Model

The LUCC model applications were built using the well-known "Metronamica" framework, developed by RIKS (e.g. White and Engelen 1993, 2000; Van Delden and Hurkens 2011) and widely used around the world for simulating urban land transformation (Barredo et al. 2004; van Delden et al. 2005; Lajoie and Hagen-Zanker 2007).

In this model, the distribution of land use in a given area is represented as a raster map in which each cell has a value that represents a land use. Not all land uses are modelled in the same way and individual land use classes must be assigned to one of three land use states. They may be either active (dynamic, changing as a result of external demands), passive (dynamic, without an external demand), or static (inert throughout the model runtime).

Metronamica calculates land use changes over time according to a set of transition rules computed by simple equations in which the geographic effect of a cell over its neighbors (attraction or repulsion between land use cells, representing economic and political power to obtain locations of interest, inertia and ease of conversion) is the main driving force of change in the system. Three additional factors are included to reflect the heterogeneity of the area: accessibility and suitability drivers are introduced to align the model with the characteristics of the study region and zoning is included to incorporate the influence of policies or planning. The model includes a stochastic component to reflect uncertainty in the allocation process. Cells are allocated at each step of the model on the basis of the transition potential until cell demand (determined exogenously) is exhausted or all suitable and available cell space is used up (see the Metronamica documentation (RIKS 2014) for more information).

For the application to Madrid, we combined some of the MLU land use classes to create a set of 12 land use classes of which 7 are urban and 6 are actively simulated (Fig. 5). This permits the observation of the effects of the change of scale

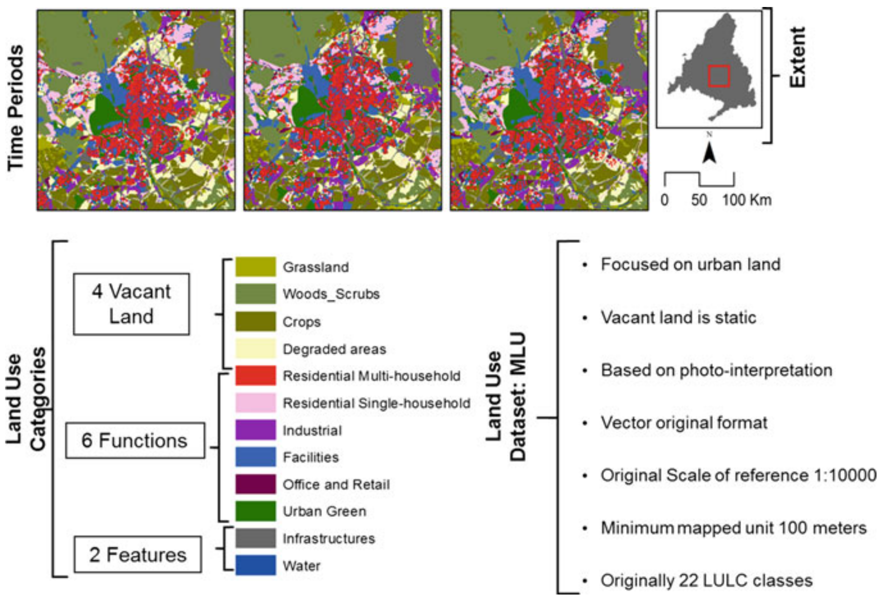


Fig. 5 Characteristics of the Madrid Model

on the model (cell size and neighborhood) for different urban land uses with dissimilar spatial behavior and dissimilar clustering.

The application for the Madrid region used in this research does not incorporate zoning so as to give the system much greater freedom. The only suitability factor included is the slope of the terrain, as this was found to be the only physical factor affecting urban land change in this region. Infrastructure networks and nodes (highways, roads, train stations and metro stations) are included and accessibility is empirically calibrated for each simulated land use through a distance decay function. The amount of randomness was set by trial-and-error during calibration.

A manual calibration was also performed on the 50 m resolution application using the common Metronamica calibration procedure (Van Delden et al. 2010, 2012). The transition rules were determined by trial-and-error, informed by previous analysis of land change processes, and by comparing the resulting simulations with historical data, until they achieved an acceptable goodness of fit (according to plausible parameters and map statistics).

The accessibility values were introduced in a similar way for each application, but in this case the values between the nearest and the furthest distance considered to the network (roads, rail, highway, metro stations…) were automatically computed by the software through a distance decay function. The only change made in this case was doubling or halving the distances in order to adapt the function for each application, e.g. if in the 50×50 m application a value for the road influence at 200 m to the residential land cells was considered, in the 100×100 m application this value was doubled to 400 m to respect the proportionality demanded by the size of the cell.

Following common practice, the transition rules thus obtained were tested for validation purposes by running the model over a different historical period than that over which the calibration was performed.

4.3.1 Metrics

Calibration and validation results were assessed through visual inspection of result maps and temporal dynamics, assessment of the plausibility of the parameters (structural validation) and a number of objective metrics to assess similarity between result maps and historic data (t_f data and t_f simulated).

The map comparison methods and techniques used during the calibration and validation processes are currently implemented in the software Map Comparison Kit (MCK), initially created by RIKS for the Netherland Environment Assessment Agency (Visser and De Nijs 2006). Three statistical tests were used to determine model accuracy, namely Kappa simulation (Ksim), clumpiness, and mass fractal dimension. The first of these, Ksim, is useful for determining the number of cells that have been correctly simulated, while the remaining two measures are used for determining the degree of spatial similarity between elements in the simulated map and the real map (White 2006). The extent (in cells) occupied by every land use on each map is also measured. In addition, a previous qualitative visual assessment based on the research criteria is generally included in the examination.

The Kappa coefficient of agreement (Cohen 1960) is a widely used index to calculate the rate of agreement between two images or two maps (categorical datasets). The Kappa simulation (Ksim) is a modification of the traditional Kappa coefficient, which is useful for evaluating simulations over short time periods (van Vliet et al. 2011). Most land use models usually simulate changes over years or decades during which time many locations do not undergo any land use change. Unfortunately, under standard Kappa, locations which do not change are also included in the calculation, which means that very high Kappa scores can be obtained regardless of the degree of accuracy of the simulation (Pontius and Millones 2011). Standard Kappa is therefore not a useful measure of the goodness of fit of simulations produced by land use models. Ksim takes values from −1, meaning total disagreement, to 1, for total agreement. The value 0 represents a special situation where the agreement is as good as can be expected by chance given a random distribution of the given class transitions (see van Vliet et al. 2011).

Clumpiness and mass fractal dimension are often employed in landscape ecology to analyze landscape structure. In this research, these metrics allow the pattern similarity of the simulated map and reference map to be assessed. Clumpiness is a measure of the degree of dispersion/aggregation of the patches in an image according to their type (McGarigal 1994). Mass fractal dimension measures the degree of "linearity" of elements in the map in which plane filling objects like circles or squares will have a value of 2.0 and a line will have a value of 1.0 (Gardner et al. 1987).

5 Results and Discussions

5.1 Results of Resample/Conversion Comparison

To examine the results (Table 2), five land use classes selected from the land use map for the year 2000 were analyzed. These classes were chosen in order to provide the greatest possible diversity of patch size for the experiment. The crops category has a very large mean patch size (107.70 ha) compared to the facilities category (3.07 ha). Residential multi-household (10.89 ha), industrial (7.56 ha) and urban green (5.57 ha) were selected to provide intermediate patch sizes between the two extremes.

Results of the vector to raster conversion and resampling tests are given in Table 2. It can be rapidly appreciated that the *CELL CENTER METHOD* gives the best results for direct conversion and the *NEAREST NEIGHBORHOOD METHOD* gives the best results for resampling. However, for the Crops and Urban Green category, the *MAXIMUM AREA* and *MAXIMUM COMBINED AREA* direct conversion methods give acceptable results, at least on the basis of the fractal

Table 2 Comparison of results from the conversion/resample operations

Land use class mean patch size	Index	Direct conversion to 50 m. Central cell	Direct conversion to 50 m. Maximum area	Direct conversion to 50 m. Maximum combined area	Resample to 50 m. Nearest neighbourhood	Resample to 50 m. Majority
Crops	Clumpiness difference	0.0060	0.0089	0.0092	0.0060	0.0165
107.70	Fractal dimension difference	0.0090	0.0212	0.0193	0.0236	0.0271
	Area difference ha	36.7394	4.2394	4.2394	19.2394	870.0106
	Kappa index	0.9535	0.9555	0.9556	0.9463	0.9500
R. Multihousehold	Clumpiness difference	0.0117	0.0179	0.0193	0.0113	0.0345
10.89	Fractal dimension difference	0.0004	0.0118	0.0137	0.0022	0.0485
	Area difference ha	20.7797	176.0297	177.2797	7.5297	325.7797
	Kappa index	0.9136	0.9168	0.9177	0.9003	0.9082
Industrial	Clumpiness difference	0.0063	0.0102	0.0100	0.0064	0.0179

(continued)

Table 2 (continued)

Land use class mean patch size	Index	Direct conversion to 50 m. Central cell	Direct conversion to 50 m. Maximum area	Direct conversion to 50 m. Maximum combined area	Resample to 50 m. Nearest neighbourhood	Resample to 50 m. Majority
7.56	Fractal dimension difference	0.0038	0.0107	0.0103	0.0005	0.0271
	Area difference ha	5.5083	108.5083	108.5083	5.2417	447.9917
	Kappa index	0.9217	0.9236	0.9077	0.9088	0.9146
Urban green	Clumpiness difference	0.0121	0.0158	0.0170	0.0128	0.0307
5.57	Fractal dimension difference	0.0266	0.0375	0.0427	0.0247	0.0642
	Area difference ha	17.3125	9.1875	9.1875	3.8125	212.0625
	Kappa index	0.9048	0.9074	0.9077	0.8907	0.8986
Facilities	Clumpiness difference	0.0120	0.0168	0.0177	0.0122	0.0313
3.07	Fractal dimension difference	0.0116	0.0178	0.0203	0.0120	0.0341

(continued)

Table 2 (continued)

Land use class mean patch size	Index	Direct conversion to 50 m. Central cell	Direct conversion to 50 m. Maximum area	Direct conversion to 50 m. Maximum combined area	Resample to 50 m. Nearest neighbourhood	Resample to 50 m. Majority
	Area difference ha	4.3011	69.9489	69.9489	6.6989	494.6989
	Kappa index	0.8930	0.8961	0.8966	0.8773	0.8852

dimension index, and the area difference in hectares. This could be related to the larger clusters found in this category (the mean patch size is 107.7 ha), but this argument does not apply to Urban Green land. The could be because Madrid contains unusually large, non-parcelled green areas such as the ‘Casa de Campo’ or ‘Dehesa de la Villa’, whose geometry is more like agricultural or natural areas than classical urban green land (parks, squares, gardens…).

For the resampling operations, the *MAJORITY METHOD* produces the lowest degree of similarity with the original data. As this method is widely used by researchers, this is a key finding.

For both clumpiness and mass fractal dimension, the calibrated application achieved similar values to the validated application and both outperformed a random land use map used as a benchmark (Table 3).

5.2 Results of Calibration and Validation of the Initial 50 × 50 m Application

Calibration was considered to be complete once values of 0.144 had been obtained for Ksim. The values considerably outperform a null model. The model was considered to have been acceptably validated at 0.113 (Table 3). These values are comparable with published values considered acceptable in other applications of the model (e.g. Hewitt et al. 2014).

5.3 Results of Testing the Changes on the Scale of the Applications

The results of the comparison of data for 2006 with simulations for the same year are shown in Table 4. According to the map comparison indices in use, the simulation results from all the different applications (apps) for 2006 (2000–2006) could be considered acceptable. Both the 25 m app with the 8 cell neighborhood radius and the 100 m app with the 4 cell neighborhood radius actually improve on the original 50 m, 8 cell neighborhood radius app (Table 4). If we look at the values for clumpiness, the difference between the clumpiness of the data and the clumpiness of the simulations is comparable across all the applications, and better than the random simulation used as a benchmark. The same is true for the fractal dimension index. In some cases, the scale-modified apps achieve slightly better values than the initial 50 m app (e.g. AP100-N4 clumpiness for multi-household and facilities classes). However, better performance of some categories tends to be compensated by poorer performance of others. Taken overall, the differences between the scale-modified apps and the original app are not large enough to be able to claim

Table 3 Values of the used indices for calibration and validation of the 50 m.application

Index	AP50 00-06	RAMD50 00-06	AP50 06-09	RAMD50 06-09
Kappa simulation	0.144	–	0.113	–
R. Multi-household Clumpiness difference	−0.0226	−0.0811	−0.0023	−0.0509
R. Single-household Clumpiness difference	−0.0018	−0.0754	0.0071	−0.0281
Industrial Clumpiness difference	0.0235	0.1196	0.0029	−0.0382
Facilities Clumpiness difference	−0.0183	−0.0856	−0.0093	−0.0331
Office and Reatil Clumpiness difference	0.0112	−0.2593	0.0081	−01271
Urban Green Clumpiness difference	−0.0361	−1115	−0.0081	−0.0352
Fractal Dimension difference	0.0070	−0.0268	0.0013	0.0040
	CALIBRATIONData06-Sim06	BENCHMARK	VALIDATIONData09-Sim09	BENCHMARK

AP = application; 50 = 50 m.; 00–06 = 2000–2006; 06–09 = 2006–2009; RAMD = random simulated map that any of the modified applications are significantly worse or better than the original 50 m app.

If we visually compare the land use data maps for 2006 and the simulations produced by the different applications, there are obvious differences that can be quickly detected, even though it is not really possible to specify the precise degree of similarity between the maps using this method. Figure 6 shows simulated (right) and real (left) land use for an enlargement of a highly urbanized (mainly residential) area in 2006. It is clear that the visual appearance of the maps reinforces the results of the statistical comparisons, i.e., none of the scale-modified apps looks significantly better or worse than any other for the land use changes simulated for the year 2006, despite the modifications in the scale. The middle column of Fig. 6 shows the apps with different resolutions (25, 50 and 100 m) and the same neighborhood radius in cells (N8). In these cases, despite the difference in resolution, all three simulations are quite alike, something that can be confirmed by consulting the results of the statistical indices (Table 4). Some classes, e.g. Residential Multi-household are not simulated very successfully in any application. This is probably because all locations in this residential area were equally favorable, being close to existing urban areas, on suitable land and close to transport networks. In such cases, identification of the "real" location tends to be difficult and is effectively made at random. Further discussion of this interesting topic is, however, beyond the scope of this paper.

The relationship between cell-size and the size of the land parcels is also clearly shown. Nonetheless, as the statistics do not show remarkable differences between the results of the apps at different resolution, a visual analysis of the 50 m resolution simulation seems to provide more realistic-looking results than the 25 m and 100 m resolution simulations, probably because the cell size is a closer match to the actual size of the land parcels, although the fact that the original calibration focused on this resolution could also be a factor. This emphasizes the importance of visual inspection when choosing the right resolution for a given application. It also suggests that pattern-based map comparison measures like clumpiness and fractal dimension have their limitations, as do all statistical measures.

This is a rather surprising result. Since the scale modifications were only applied to the maps themselves, and not to the neighborhood rules, neighborhood influence is different in all three applications. The maximum cell neighborhood of 8 cells corresponds to a distance of 400 m (8×50) away from the central cell in the original 50 m app, 200 m away from the central cell in the 25 m app, and 800 m away from the central cell in the 100 m app. Three possible explanations can be provided for this; (1) the cell neighborhood is not the key change driver (contrary to most known studies of urban change); (2) the neighborhood influence declines very steeply and all important interactions take place at close distances, or (3) the distance in cells is more important than the actual distance (in meters) in the calculation of the neighborhood effect. Further experimental work (see, e.g. Hewitt and Díaz-Pacheco 2017) would be needed to confirm or reject these hypotheses.

Table 4 4. Map comparison results for applications. Abbreviations: AP50-25-100: applications and resolutions; N8-4: neighborhood and radius in cells; RAMD50-100-25: random simulations and resolutions

	Applications					Benchmarks		
Index	AP50-N8	AP25-N8	AP100-N8	AP25-N16	AP100-N4	RAMD50	RAMD100	RAMD25
Kappa simulation	0.144	0.149	0.116	0.146	0.158	–	–	–
R. Multi-household Clumpiness difference	0.0226	0.0430	0.0050	0.0400	0.0007	0.0811	0.0602	0.1017
R. Single-household Clumpiness difference	0.0018	0.0079	0.0111	0.0192	0.0250	0.0754	0.0643	0.0867
Industrial Clumpiness difference	0.0235	0.0020	0.0409	0.0111	0.0409	0.1196	0.0991	0.1352
Facilities Clumpines Difference	0.0183	0.0290	0.0065	0.0567	0.0044	0.0856	0.0609	0.1054
Office and Reatil Clumpiness difference	0.0112	0.0156	0.0731	0.0284	0.0835	0.2593	0.1981	0.3007
Urban Green Clumpiness difference	0.0361	0.0642	0.0108	0.0849	0.0099	0.1115	0.0752	0.1550
Fractal Dimension difference	0.0070	0.0108	0.0053	0.0157	0.0003	0.0268	0.0182	0.0299

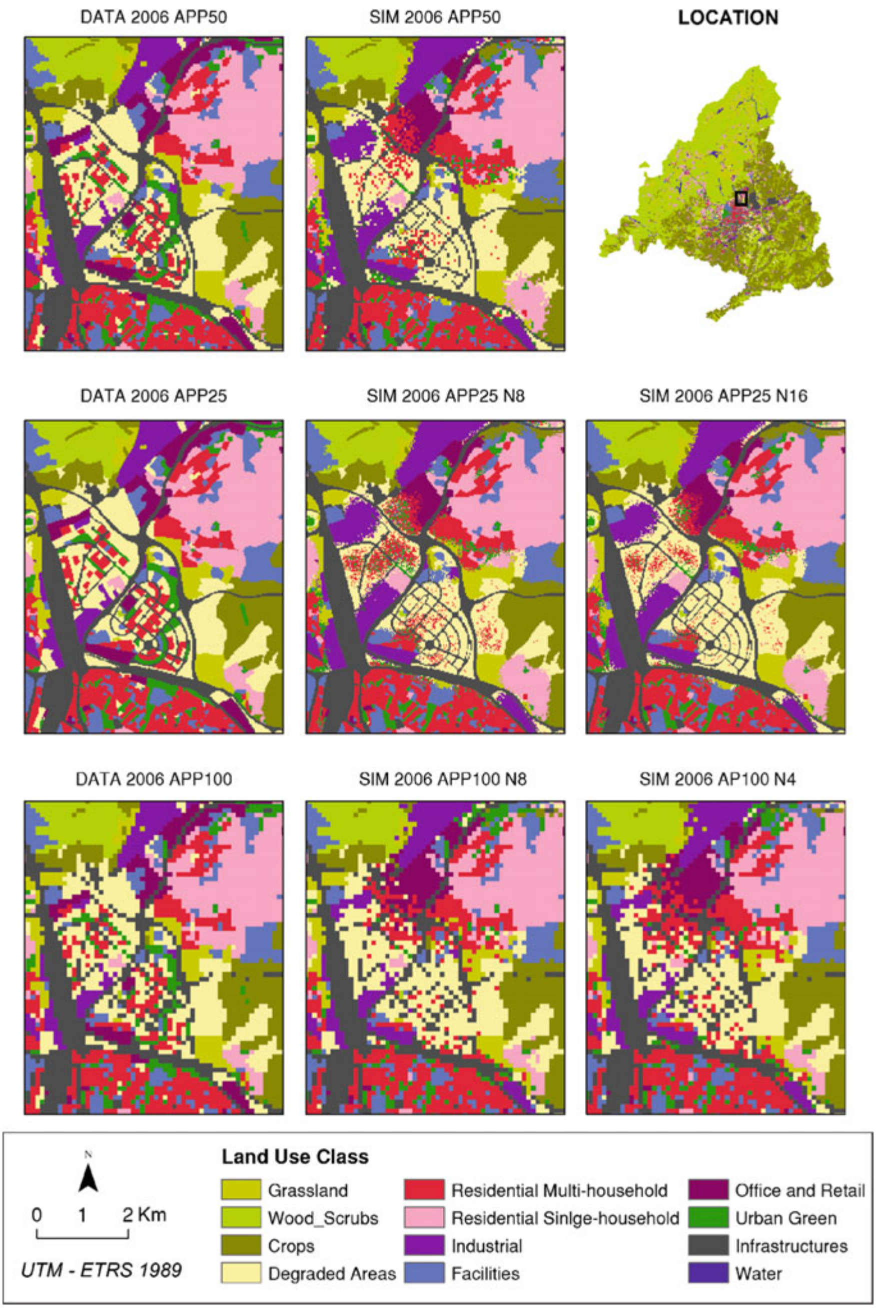


Fig. 6 Comparison of data 2006 and simulations for 2006 from the different apps. Abbreviations: SIM: simulation; APP25-50-100: application and resolution; N4-8: neighborhood and radius

6 Conclusions and Outlook

Standard GIS operations like vector-to-raster conversion and raster resampling have considerable influence on scale in models of land use and land cover change, and the MAUP (Openshaw 1983) would seem to be relevant. In CA models, which are highly dependent on the cell neighborhood for simulating land use conversions, both operations have a significant effect on the initial land use map and hence the cell neighborhood. The work presented in this paper has examined the influence of these operations, the first referred to the resolution transformation of the input data and the second to the land use model's capacity to simulate land use change for a case study application based on a large and detailed land use database for Madrid, Spain. Some important conclusions can be drawn that are likely to be extremely useful for researchers working with cell-based land use models. It is clear from this work that the use of one particular data preparation method over another can produce quite different results, both for vector to raster conversion operations and for raster resampling from one resolution to another. In the underlying research, for urban patch types (smaller mean patch sizes), better results (a closer match to the original land use dataset) are obtained by converting directly from the original vector coverage to a raster with the desired resolution than by converting to a scale equivalent to the original vector coverage and subsequently resampling up or down to obtain the desired resolution. Amongst the resampling methods themselves, the nearest neighbor technique gives improved agreement with regard to the original land use dataset than most other procedures. Future research could try to discover whether similar results would be found if the same methods were applied to different datasets.

Regarding the effects of changing the scale of a dynamic CA land use model, as reflected by the cell resolution and neighborhood radius, no significant variation was obtained in the accuracy of the final simulations measured by the metrics applied, at least in the urban context considered and for the range of resolutions tested (25, 50, and 100 m). A calibrated and validated land use model based on a 50 m resolution raster gave very similar results to applications with identical transition parameter settings but mapped at higher (25 m) and lower (100 m) resolutions.

The goodness-of-fit evaluation techniques (cell statistics, pattern comparison, visual inspection) showed that all of the applications acceptably reproduced the relevant land use change patterns. Despite this result, the 50 m resolution model looked more realistic than 25 m or 100 m resolution applications. This is likely to be because the 50 m cell size is a better fit to the size of the real land parcels, although the fact that the original calibration focused on this resolution may also be a factor.

The most surprising discovery is that doubling or halving the neighborhood distance radius did not produce any significant variation in the model's performance over the validation period. This indicates that for the applications we investigated the transition rules are rather insensitive to neighborhood distance effects. For future research it would be useful to investigate whether similar results are obtained for applications to different regions and datasets and if so, whether the following possible hypotheses could be confirmed or denied: the cell neighborhood is not the key driver for change, neighborhood effects all occur at close distances, the distance in cells is more important than the actual distance (in meters) in the calculation of the neighborhood effect.

Acknowledgements This work has been supported in part by project SIGEOMOD-II, BIA2013-43462-P (Spanish Ministry of Economy and Competitivity and European Regional Development Fund FEDER). The authors also gratefully acknowledge funding received under remit of the EU FP7 project COMPLEX (project no. 308601). The MLU geodatabase was built by researchers from the Department of Human Geography, Universidad Complutense de Madrid with financial support from the Ministerio de Ciencia e Innovación (project TRA2008-06682). Finally, we thank the anonymous reviewers for their helpful comments.

References

Barredo JI, Demicheli L, Lavalle C, Kasanko M, McCormick N (2004) Modelling future urban scenarios in developing countries: an application case study in Lagos, Nigeria. Environ Plann B 31(1):65–84

Carver S, Brunsdon C (1994) Vector to raster conversion error and feature complexity: an empir-ical study using simulated data. Int J Geogr Inf Syst 8(3):261–270

De Lucio RL (2011) Transformaciones territoriales recientes en la región urbana de Madrid. Urban 8:124–161

De Koning GHJ, Veldkamp A, Fresco LO (1998) Land use in Ecuador: a statistical analysis at different aggregation levels. Agric Ecosyst Environ 70:231–247

Díaz-Pacheco J, Gutiérrez J (2013) Exploring the limitations of CORINE land cover for monitoring urban land use dynamics in metropolitan areas. J Land Use Sci 9(3):243–259. doi:10.1080/1747423X.2012.761736

EEA (2014) Report of Content of CORINE Land Cover.European Environmental Agency. http://www.eea.europa.eu/publications/COR0-landcover

ESRI (2010) Manual ArcGIS [software GIS]. Versión 10.0. Redlands, CA: Environmental Sys-tems Research Institute, Inc

Gardner RH, Milne BT, Turnei MG, O'Neill RV (1987) Neutral models for the analysis of broad-scale landscape pattern. Landsc Ecol 1(1):19–28

Hewitt R, Escobar F (2011) The territorial dynamics of fast-growing regions: unsustainable land use change and future policy challenges in Madrid, Spain. Appl Geogr 31(2):650–667

Hewitt R, Van Delden H, Escobar F (2014) Participatory land use modelling, pathways to an integrated approach. Environ Model Softw 52:149–165

Hewitt R, Díaz-Pacheco J (2017) Stable models for metastable systems? Lessons from sensitivity analysis of a Cellular Automata urban land use model. Comput Environ Urban Syst 62:113–124.

Jantz C, Goetz SJ (2005) Analysis of scale dependencies in an urban land-use-change model. Int J Geogr Inf Sci 19(2):217–241

Jenerette GD, Wu J (2001) Analysis and simulation of land-use change in the central Arizona-Phoenix region, USA. Landsc Ecol 16(7):611–626

Kok K, Veldkamp A (2001) Evaluating impact of spatial scales on land use pattern analysis in Central America. Agr Ecosyst Environ 85(1):205–221

Lajoie G, Hagen-Zanker AH (2007) Modeling urban sprawl on the island La Réunion: contribu-tion of the Metronamica® cellular automata for territorial prospect. Cybergeo 405

Lam SN, Quattrochi DA (1992) On the issues of scale, resolution, and fractal analysis in the mapping sciences. The Professional Geographer 44(1):88–98

Longley P, Batty M (2003) Advanced spatial analysis: the CASA book of GIS. ESRI, Inc

McGarigal K, Marks BF (1994) Spatial pattern analysis program for quantifying landscape structure reference manual. Forest Science Department, Oregon State University, Corvallis Oregon

Ménard A, Marceau DJ (2005) Exploration of spatial scale sensitivity in geographic cellular automata. Environ Plan 32(5):693–714

Openshaw S, Taylor P (1979) A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: Wrigley N (ed) Statistical applications in the spatial sciences. Pion, London, pp 127–144

Openshaw S (1983) The modifiable areal unit problem, vol 38. In: Geo books, Norwich

Plata Rocha W, Gómez Delgado M, Bosque Sendra J (2009) Cambios de usos del suelo y expansión urbana en la Comunidad de Madrid (1990–2000). Scripta Nova: revistaelectrónica de geografía y cienciassociales 13

Pontius JRG, Millones M (2011) Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. Int J Remote Sens 32(15):4407–4429

RIKS (2014) Metronamica Documentation.Research Institute for Knowledge Systems, The Netherlands. http://www.riks.nl/

Samat N (2006) Characterizing the scale sensitivity of the cellular automata simulated urban growth: A case study of the SeberangPerai Region, Penang State, Malaysia. Computers, environment and urban systems 30(6):905–920

Switzer P (1975) Estimation of the accuracy of qualitative maps. Wiley, London

Theobald DM, Hobbs NT (1998) Forecasting rural land-use change: a comparison of regression-and spatial transition-based models. Geogr Environ Modell 2:65–82

Tobler W (1988) Resolution, resampling, and all that. Building databases for global science, vol 12, pp 9–137

Van Delden H, Escudero JC, Uljee I, Engelen G (2005) METRONAMICA: A dynamic spatial land use model applied to Vitoria-Gasteiz. In: Virtual Seminar of the MILES Project. Centro de Estudios Ambientales, Vitoria-Gasteiz

Van Delden H, Stuczynski T, Ciaian P, Paracchini ML, Hurkens J, Lopatka A, Shi Y, Gomez Prieto O, Calvo S, Van Vliet J, Vanhout R (2010) Integrated assessment of agricultural poli-cies with dynamic land use change modelling. Ecol Model 221(18):2153–2166

Van Delden H, Hurkens J (2011) A generic integrated spatial decision support system for urban and regional planning. Keynote presented at MODSIM11 international congress on modelling and simulation, Perth, Australia

Van Delden H, Van Vliet J, Rutledge DT, Kirkby MJ (2011) Comparison of scale and scaling issues in integrated land-use models for policy support. Agric Ecosyst Environ 142(1):18–28

Van Delden H, Díaz-Pacheco J, Shi Y, Van Vliet J (2012) Calibration of cellular automata based land use models: lessons learnt from practical experience, CAMUSS (Cellular Automata Modeling for Urban and Spatial Systems) conference, November 8-10. Oporto, Portugal

Van Vliet J, White R, Dragicevic S (2009) Modeling urban growth using a variable grid cellular automaton. Comput Environ Urban Syst 33(1):35–43

Van Vliet J, Bregt AK, Hagen-Zanker A (2011) Revisiting Kappa to account for change in the accuracy assessment of land-use change models. Ecol Modell 222(8):1367–1375

Visser H, De Nijs T (2006) The map comparison kit. Environ Model Softw 21(3):346–358

White R (2006) Pattern based map comparisons. J Geogr Syst 8(2):145–164

White R, Engelen G (1993) Cellular automata and fractal urban form: A cellular modelling approach to the evolution of urban land-use patterns. Environ Plann A 25:1175

White R, Engelen G (2000) High-resolution integrated modelling of the spatial dynamics of urban and regional systems. Comput Environ Urban Syst 24(5):383–400

Woodcock CE, Strahler AH (1987) The factor of scale in remote sensing. Remote Sens Environ 21(3):311–332