This paper has been published online in the South African Geographical Journal. Readers can access the edited version of this manuscript through the following link: https://doi.org/10.1080/03736245.2021.1872413

Sensitivity of a standard Land Use Cover Change Cellular Automata Model to resample input Land Use Cover maps

David García-Álvarez[a]* and María Teresa Camacho Olmedo[a]

[a]*Departmento de Análisis Geográfico Regional y Geografía Física, University of Granada, Granada, Spain*

*Departamento de Análisis Geográfico Regional y Geografía Física, Facultad de Filosofía y Letras, Campus de Cartuja SN, 18071, Granada, Spain – dagaral@ugr.es

Provide short biographical notes on all contributors here if the journal requires them.

Sensitivity of a standard Land Use Cover Change Cellular Automata Model to resample input Land Use Cover maps

Input data is one of the main sources of uncertainty in Land Use Cover Change (LUCC) modelling. Research has focused on the sensitivity of LUCC models to the spatial resolution of Land Use Cover (LUC) maps. However, little attention has been paid to the way that spatial resolution is changed. Both the spatial resolution and the resampling method change the modelled landscape composition and configuration. This may affect the way Cellular Automata LUCC models behave and, accordingly, the landscapes they simulate. This paper analyses the sensitivity of a standard LUCC model (Metronamica) to changes in the spatial resolution and resampling method of input LUC maps. Results prove how the model is more sensitive to changes in the spatial resolution than to variations of the resampling method. This last component has not much influence in the simulated landscape, although it alters the landscape composition.

Keywords: uncertainty; Metronamica; majority rule; nearest neighbour; scale

Subject classification codes: include these here if the journal requires them

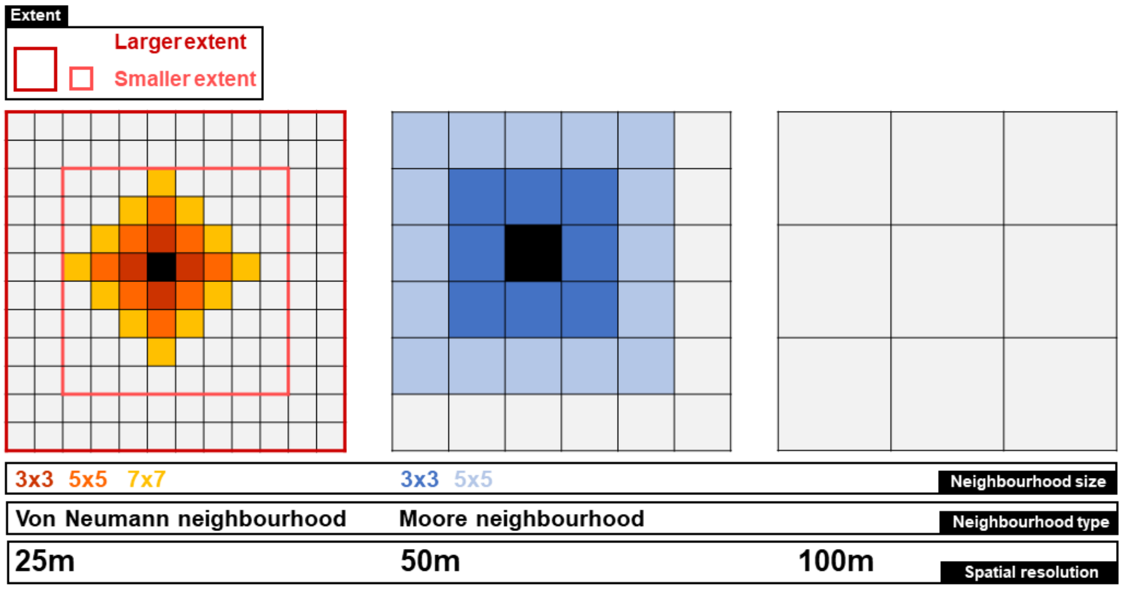
1. Introduction

Characterization of the uncertainty associated to any Land Use Cover Change (LUCC) model is a required step to increase the confidence in LUCC modelling exercises (Li and Wu 2006; Yeh and Li 2006). Only in this way audience and stakeholders can be totally aware of the limits and possibilities of these tools (Warmink et al. 2010).

Studies comparing several LUC sources have revealed important disagreements between them (Bach et al. 2006; McCallum et al. 2006; Ran et al. 2010; García Martínez et al. 2015). Each source shows a different landscape and LUC dynamics. Thence, the modelling exercise is source specific. However, even when employing the same source, important variations are measured because of changes in the scale (Dietzel and Clarke 2004; Chen and Pontius Jr. 2011; Kim 2013; Blanchard et al. 2015).

Scale is a complex term that, according to Wu (2007), can refer to different dimensions (spatial, temporal, organizational hierarchies), types (instrinsic, observation, experimental, analysis/modelling and policy scale) and components (grain, extent, coverage, spacing, catographic scale). When talking about CA models and only focusing on the spatial dimension of the scale, authors mainly refer to the extent, neighbourhood type and size and spatial resolution (Fig. 1) (Ménard and Marceau 2005; Pan et al. 2010; Wu et al. 2019). There is wide research focusing on the sensitivity of CA models to changes in the last three elements of the scale (Ménard and Marceau 2005; Kocabas and Dragicevic 2006; Samat 2006; Zhao 2013; Morais Viana 2014; Altartouri et al. 2015; Hewitt and Díaz-Pacheco 2017). Fewer are the analyses that study the sensitivity of CA models to changes in the extent (Pan et al. 2010).

Figure 1. Representation of the four components of the spatial dimension of the scale that we usually refer to in CA models.



All studies usually agree on the sensitivity of CA models to changes in the spatial resolution, type and extent of the employed neighbourhood. Disagreement arises when choosing which is the element that introduces more variability, that is, bigger uncertainty. Morais Viana (2014) found her exercises more sensitive to the neighbourhood configuration than to the cell size. On the contrary, Ménard and Marceau (2005) found their exercises more sensitive to changes in the spatial resolution than to different neighbourhood configurations. Notwithstanding, most authors agree on the interdependence between cell size and spatial neighbourhood (Benenson 2007; Pan et al. 2010; Altartouri et al. 2015).

Depending on the study case, the spatial resolution of the exercise that showed the best performance was different (Samat 2006; Morais Viana 2014; Hewitt and Díaz-Pacheco 2017). Accordingly, no rules can be established when choosing the proper cell size (Chen and Pontius Jr. 2011). Depending on the pattern and complexity of the modelled landscape, among other criteria, one or another resolution will be more suitable (Ménard and Marceau 2005; Altartouri et al. 2015; Blanchard et al. 2015).

Although some agreement has been achieved regarding that point, there is still a gap on the criteria we must follow to select the required resampling or rescaling method to vary that cell size. This is usually a required step given that the original spatial resolution of input maps does not commonly fit with the user desires or limitations.

For categorical data, there are two main resampling methods commonly used and implemented in GIS software: nearest neighbour and majority rule (Fig. 2). The nearest neighbour (NN) method assigns to the new cell in the resampled raster the value of the cell in the original raster that is nearest to the centre of the new cell (ESRI). The majority rule (MR) method assigns to the new cell the most repeated value between those cells of the original raster that fall inside the new one (ESRI). Depending on the chosen method, the obtained landscape will be different (Díaz-Pacheco et al. 2018).

Figure 2. Graphic representation of Nearest Neighbour (NN) and Majority Rule (MR) resampling methods for an original raster of 30m resampled to 50 and 100m.

Up to date, there is not any study that have addressed the influence of the selected resampling method in LUCC modelling practice. Díaz-Pacheco et al. (2018) compared resampled LULC maps with the original one. They concluded that the NN method produced the closest results to the original dataset. In addition, they, as well as Dendoncker et al. (2008), assessed the influence of the way vector data is rasterized (method and cell size) in the maps used as input for a modelling exercise. They focused on rasterization, but not on resampling. Neither any of the cited studies assessed the sensitivity of a LUCC model to those variations in input maps.

For hydrological modelling, papers addressing the sensitivity of a model to different resampling methods and spatial resolutions of the input Digital Elevation Model (DEM) proved the important influence that these decisions play on the obtained results (Le Coz et al. 2009; Ficklin et al. 2015). Similar conclusions were obtained for the modelling of carbon dynamics when employing and resampling LUC maps (Zhao and Liu 2014).

Through this paper, we aim to test the sensitivity of a LUCC model to resample input maps. We assume that maps produced through different resampling methods at different spatial resolutions generate different landscape compositions (quantity of cells of each category) and configurations (distribution of cells of a given category).

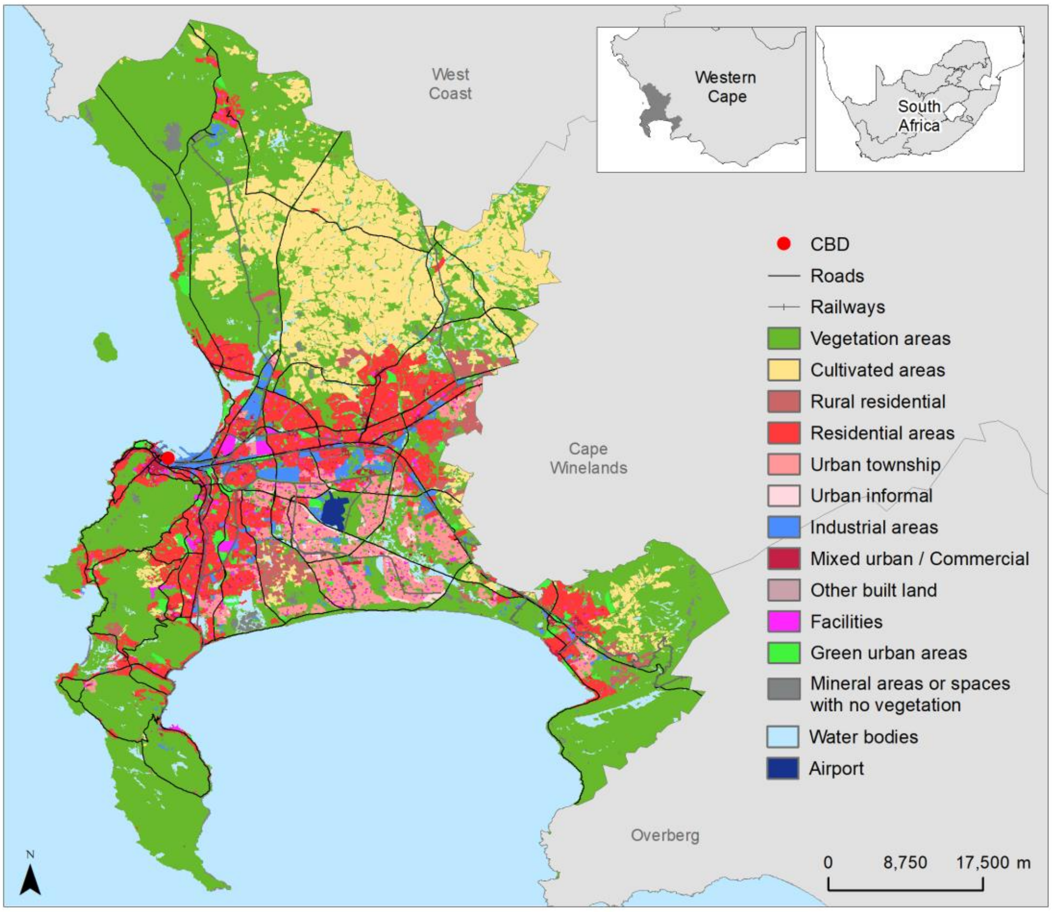
For different landscape compositions and configurations, changes in the neighbourhood interactions are expected. Transition rules, that explicit these neighbourhood interactions, are one of the key components of most CA models (Altartouri et al. 2015). Therefore, cell size and resampling method may have a strong impact on the landscape simulated by these models. This paper aims to specify the extent of this influence and to set out, according to the obtained results, a guideline about the criteria we must follow to decide about the spatial resolution and resampling method of input LUC maps.

To that end, the paper is structured as follows: first, we describe the study area we modelled. Second, we provide a short description of the materials and model employed, pointing out how the input LUC maps were resampled. Third, details of the model calibration and the way the sensitivity analysis was carried out are provided. Finally, we analyse and discuss the results in the next two sections. The conclusions summarize the main findings of this research.

2. Study area

The City of Cape Town (CCT) is the modelled area in this study (Fig. 3). After its constitution in 2000, it embraces the Cape Town metropolitan area (Wilkinson 2004). It is made of built-up surfaces, agricultural areas and an important series of natural assets. In this regard, 17.7% of the CCT is under protection (Rebelo et al. 2011).

Figure 3. City of Cape Town location map.



The urban footprint is mostly shaped by the Apartheid policy and its consequences (City of Cape Town 2018). The Group Areas Act of 1950 promoted the segregation of racial communities (Western 1997). New neighbourhoods were built in places further from the city centre and urban footprint to host Black Africans and Coloureds. With the creation of those new neighbourhoods, the city sprawled at fast speed following the infrastructure network (City of Cape Town 2018). However, lack of housing to host rural immigrants led to the emergence of informal housing (Wilkinson 2000; Town and Turok 2001).

After the end of Apartheid regime, that trend has intensified. South Africa democratic government policy of housing has tried to grant a house to all those without a formal roof (Pieterse 2009). This has been only possible in those areas created to host black and coloured communities, which has promoted the city sprawl. In addition, despite those efforts, housing supply is still largely behind housing demands. Consequently, the number of population inhabiting informal houses has been steadily increasing (City of Cape Town 2012).

According to those dynamics, three types of residential land uses may be differentiated in Cape Town: formal private developed housing; formal state subsidised housing (townships); and informal dwellings. They relate differently with industrial and commercial areas. Because of the Apartheid legacy, there is a dissociation between places to work and places to live in the case of low income communities: formal state subsidised housing and informal dwellings (City of Cape Town 2012). Activity concentrates in the north of the CCT and around the major interchanges on the transport network (Wilkinson 2000). Although some policies have been implemented to overcome that polarization of activity, it still remains concentrated in a few activity centres (City of Cape Town 2012).

3. Materials

3.1 Land Use Cover (LUC) maps

Two Land Use and Cover (LUC) maps for the years 1990 and 2013 were generated from the South African National Land Cover Dataset (SANLCD) (GeoTerra Image 2015a, b). A third map for the yar 2002 was obtained by photointerpretation of historical satellite imagery. Figure 4 summarizes the processes followed to obtain the input LUC maps.

The SANLCD is a dataset obtained from Landsat imagery with a spatial resolution of 30m. It is available for the years 1990 and 2013. Its thematic resolution is 72 classes, which were simplified to only 15. Those classes whose detail was not needed for the purposes of our modelling exercise (simulating urban sprawl) were grouped. Additionally, we mapped four extra categories from auxiliary data because of their importance to better explain the modelling processes (Table 1). The Cape Town Airport was extracted from the National Topographic Map provided by the National Geo-Spatial

Information (NGI) of South Africa. Cemeteries, military areas and informal settlements were mapped from spatial data provided by the City of Cape Town.

Table 1. Legend of the LUC maps produced for this study. In *italics* are the categories we manually mapped from auxiliary data.

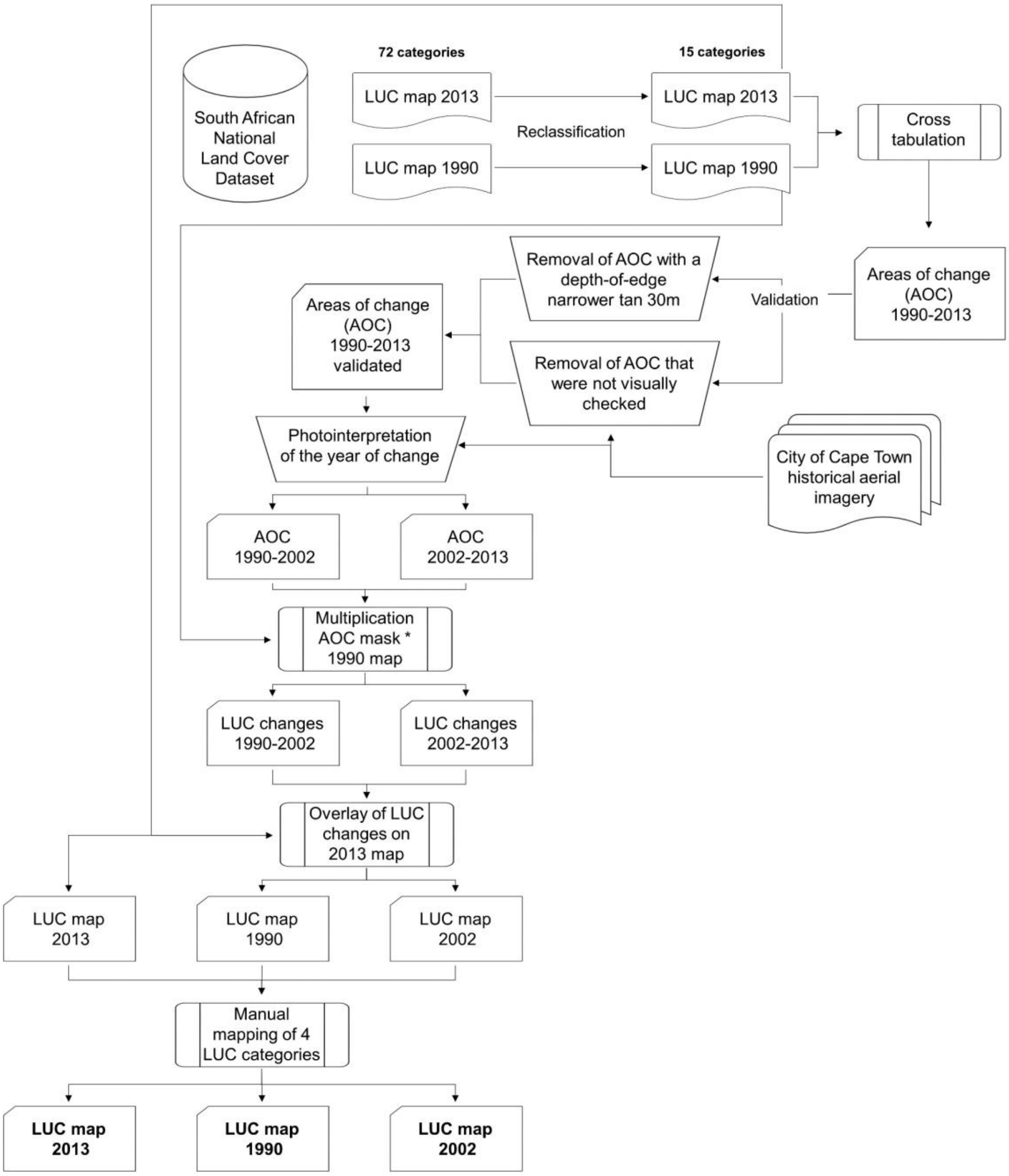
Land use cover categories
Vegetation areas
Other cultivated areas
Cultivated vine areas
Rural residential
Residential areas
Urban township
Urban informal
Industrial areas
Mixed urban / commercial
Other built land
Facilities
Recreational areas
Military areas
Cementeries
Mineral extraction sites
No vegetation areas
Wetlands
Water bodies
Airport

Raw LUC maps were cross tabulated to obtain the areas of LUC change (AOC) for the period 1990-2013. These areas of change were further validated to rule the technical changes out. All AOC with a depth-of-edge no longer than the pixel size (30m) were discarded. We also eliminated those AOC whose land use or cover change we could not validate visually. Historical imagery provided by the City of Cape Town Council for the years 1988 and 2012 was used to this end.

From the validated layer of AOC, we photointerpreted those changes happening between 1990 and 2002 from those happening between 2002 and 2013. Historical imagery for the year 2002 was the basis for this photointerpretation task. In the end, two layers of AOC (1990-2002; 2002-2013) were produced.

To obtain the final LUC maps, the AOC layers were used as mask of the raw 1990 LUC map to produce two LUC layers of changes. These were furtherly superimposed on the 2013 raw map.

Figure 4. Flowchart of the procedure followed to obtain the input LUC maps of our analysis from the South African National Land Cover Dataset.

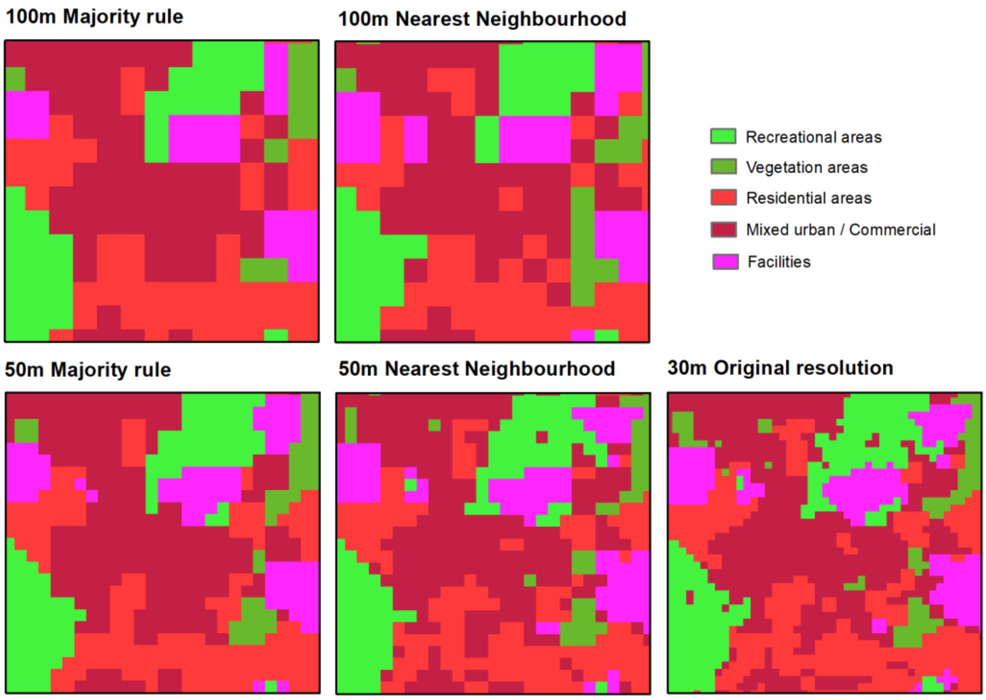


3.2 Resampled maps

The original maps at 30m were resampled at 50 and 100m (Fig. 5). These spatial resolutions reflect common practice in LUCC urban modelling (Santé et al. 2010; Van Vliet et al. 2016). In addition, they comply with the minimum level of detail that is required in the modelling of urban dynamics. Resolutions coarser than 100m would be far from the scale at which agents play and stakeholders work.

Maps were resampled by means of two rasterization methods: nearest neighbour (NN) and majority rule (MR). They are the common methods employed when changing the resolution of categorical data. In addition, they are the methods usually implemented in standard and wide used GIS software like ArcGis (Díaz-Pacheco et al. 2018).

Figure 5. Example area of the City of Cape Town at the original resolution of input LUC maps (30m) and resampled at 50 and 100m through NN and MR methods. Readers may found the full maps as supplementary material to the online version of this paper



3.3 Land Use Cover Change (LUCC) model

Metronamica (RIKS 2012) is a Cellular Automata (CA) constrained model built on the theory proposed by White and Engelen (1993, 1997) and White et al. (1997), over which relies an important part of the CA LUCC modelling practice (Santé et al. 2010). The model simulates all land uses at every time step based on four different factors (accessibility, neighbourhood, suitability and zoning) plus a random component that accounts for the uncertainty of human action.

The model distinguishes between three types of classes: features, vacants and functions. Features are those categories that do not change through the simulation. Vacants are categories modelled passively. They are allocated after all function classes based on a suitability map and an inertia/conversion factor. Functions are modelled actively based on the combination of all above-mentioned factors (accessibility, neighbourhood, suitability, zoning and a random component). These factors define the transition potential of each cell to allocate every function. According to their demands, function classes allocate progressively in those cells with the highest potential. Vacant classes allocate next in the remaining cells following the same procedure.

3.4 Model factors

In addition to the LUC maps described above, we employed other spatial data from different sources to build the factors that drive the model behaviour in Metronamica. Details about these data, their sources and the factors for whose construction have been used can be checked in Table 2.

Table 2. Data sources employed to build the factors that drive the model we set up.

Data	**Source**	**Factors where they have been used**
Raw land use maps	South African National Land Cover Dataset (Department of Environmental Affairs)	Suitability / Zoning
Road and rail infrastructure	Topographic Data (NGI, Department of Rural Development and Land Reform)	Accessibility / Suitability
Water bodies		Accessibility

Activity centres	Economic Areas Management Programme (ECAMP)	Accessibility
Median land economic value per suburb	Freehold residential property valuations (median values) (City of Cape Town)	Suitability
Slope	Digital elevation model 10m (City of Cape Town)	Suitability
Zoning map	Integrated zoning land parcel (City of Cape Town)	Suitability / Zoning
Protected areas (conservation)	South African Protected Areas Database (SAPAD) (Department of Environmental Affairs)	Zoning

4. Methods

4.1 Model calibration and validation

We set up five modelling exercises, one for each series of input maps (section 3.2). An initial exercise was calibrated at the original resolution of input maps (30m). Then, four extra exercises, one for each pair of resampled input maps, were further calibrated based on those initial parameters.

Calibration was in all cases manual, as defined by Van Vliet et al. (2016). Demands were manually introduced from measured changes by each pair of input maps. Residential areas, urban informal, urban township and industrial areas were the function categories. They lead the urban sprawling process we want to simulate. Vegetation areas, other cultivated areas, cultivated vine areas and rural residential were the vacant categories. Urban areas sprawl on the basis of these uses.

An initial calibration was made in each case for the period 1990-2002. Next, this calibration was improved in a trial and error basis according to the information given by the following indices: Kappa (Cohen 1960), Kappa simulation (KSim) (Van Vliet et al. 2011), Fuzzy Kappa Simulation (FKSim) (Van Vliet et al. 2013a), Figure of Merit (FOM) (Pontius Jr. et al. 2008) and the clumpiness spatial metric (McGarigal et al. 2015). The simulated landscape was also evaluated through visual inspection and compared to a random benchmark to assess its goodness. Once we achieved the best possible simulation, each model was validated against reference data for the year 2013.

4.2 Sensitivity analysis

The calibrated exercises were compared qualitatively and quantitatively through the methods described below. Differences between calibrations allow to see how the model behaviour changes because of the different input maps. Differences between simulations explicit how those input maps affect the modelled landscape.

To compare calibrations, we examined the calibrated parameters and processing time of each exercise. In addition, we calculated the Pearson correlation coefficient (Pearson 1895) between the reference LUC map of each modelled category and its corresponding factors for the year 2002.

Simulated landscapes were compared visually and quantitatively in terms of allocation, quantity and pattern (dis)agreement. Quantity and allocation (dis)agreements were calculated through the matrix proposed by Pontius Jr. (2018). Quantity disagreement refers to the different number of cells that make up a category in a pair of input maps. Allocation disagreement refers to the cells of the same class allocated in different positions in a pair of input maps.

Pattern disagreement was assessed by means of a series of spatial metrics calculated through the software Fragstats. From all spatial metrics calculated, we selected the two that better describe the different pattern simulated by each exercise: number of patches and largest patch index. The first refers to the number of patches that make up a category in a LUC map, and the second to the size of the largest patch in a LUC map with respect to its total area.

Kappa (K), Kappa Simulation (KS), Fuzzy Kappa Simulation (FKS) and the Figure of Merit (FOM) were also employed as a reference to assess the similarity between each simulation and its corresponding reference map. K expresses the agreement between two maps, corrected for the agreement expected by chance (Cohen 1960). KS expresses the agreement between changes of two maps with respect to a third one, corrected for the agreement expected by chance (Van Vliet et al. 2011). FKS expresses the same agreement but accounting for the degree of spatial mismatch (Van Vliet et al. 2013a). FOM expresses the spatial overlapping between simulated and reference changes, meaning 0% perfect disagreement and 100% perfect agreement (Pontius Jr. et al. 2008).

5. Results

5.1 Model calibration

The spatial resolution affects the processing time the model needs to perform the simulation. and the explanatory power of the factors. On the contrary, the resampling method employed to obtain the maps does not show a meaningful influence in any of those aspects.

Finer spatial resolutions exponentially increase the processing time of the model. From only 13 seconds of the 100m exercise to 3 minutes and 30 seconds of the 50m exercise. The 30m exercise also required 3 minutes and 30 seconds, but because the extent of the landscape was much smaller than in the other two cases. The model was not able to perform the simulation at the original landscape extent for the 30m maps. Therefore, landscape extent was cut off to the exact limits of the simulated area to avoid the model crash. Thus, despite the different extent, the simulated area was the same at the three resolutions.

Explanatory power of factors varies because of the spatial resolution, but not as consequence of the resampling method followed. Factors at finer spatial resolutions relate more with the land uses than factors at coarser spatial resolutions (Table 3).

Table 3. Correlation coefficient betweenn Transition Potential (TP) maps for each function category and its corresponding land use binary map, where 1 means presence of that category and 0 absence. TP maps are the result of the combination of the four factors of a modelling exercise and define the suitability of each cell to hold a land use category.

	Residential areas	**Urban township**	**Urban informal**	**Industrial areas**
30m	0.88	0.86	0.68	0.90
50m NN	0.87	0.83	0.24	0.81
50m MR	0.87	0.83	0.22	0.82
100m NN	0.86	0.80	0.02	0.75
100m MR	0.86	0.80	0.02	0.76

At the category level, these trends are very contrasted. Whereas the correlation between factors and land uses is similar for all spatial resolutions and resampling methods in the case of the residential areas' category, it varies to a great extent in the case of the urban informal (Table 3). In this case, correlations vary much more because of the spatial resolution than because of the resampling method employed.

From the four factors considered, neighbourhood is always the one that correlates the most with the land use maps. It is also the factor more affected by changes in the spatial resolution and resampling methods. Suitability shows some correlation with land use maps at the original resolution, but does not show any correlation at all at 50 and 100 meters. The rest of factors do no show very different correlation coefficients between different resolutions and resampling methods. Readers may find the correlation coefficients disaggregated for each factor as supplementary material to the online version of this paper.

Each exercise was manually calibrated on a trial error basis. Accordingly, parameters may be different depending on the considered exercise, so as to better fit the factors with the simulated landscape. Accessibility, suitability and zoning factor parameters are very similar between exercises at 30 and 50m of spatial resolution. Bigger differences may be pointed out between exercises at 100m of spatial resolution and the ones at 30 and 50m for these factors.

Neighbourhood rules are more affected by changes in the spatial resolution and resampling method. The weight of attraction / repulsion rules between land uses varies in all cases, with a greater variation when changing the spatial resolution from 50m to 100m and, for the coarser resolution exercise, when changing the method used to resample the input maps. Although the weights change, the cell extent of neighbourhood influence does not change between simulations, always around a maximum of 6 cells of influence. This means a very different neighbourhood influence in meters when employing maps at different spatial resolutions: 180m (30m exercise), 300m (50m exercises) or 600m (100m exercises).

5.2 Simulation

According to Kappa indices and the Figure of Merit, global model performance was slightly better in exercises at 50m run with maps resampled through MR (Table 4). Exercises with maps at 100m perform better than the ones at the original resolution and also show the highest FKSim. This makes sense because of their coarser resolution, which increases the fuzzy radius over which hits are searched.

Table 4. Kappa and Figure of Merit scores obtained for each of the calibrated modelling exercises. Figure of Merit (FOM), Kappa Simulation (Ksim), Fuzzy Kappa Simulation (FKsim) and Standard Kappa (Kappa) are provided globally. At the category level we only show the main metric: Ksim. Ksim and Kappa scores for all simulated classes may be found as supplementary material to the online version of this paper.

	Global				Residential areas	Urban township	Urban informal	Industrial areas
	FOM	Ksim	Fksim	Kappa	Ksim	Ksim	Ksim	Ksim
30m	23.5%	0.36	0.38	0.96	0.40	0.48	0.25	0.20
50m NN	24.5%	0.38	0.41	0.96	0.41	0.49	0.23	0.23
50m MR	25.1%	0.40	0.43	0.96	0.43	0.50	0.15	0.29

100m NN	23.3%	0.37	0.42	0.96	0.40	0.46	0.18	0.23
100m MR	24.9%	0.39	0.44	0.96	0.41	0.50	0.18	0.27

Differences of model performance between spatial resolutions and resampling methods may be pointed out at the category level. Residential areas, urban township and industrial areas show the same pattern of behaviour than the global one: performance is the best at 50m and in exercises run with maps resampled through MR. However, the simulation of urban informal is the best at the original resolution and decreases according to the level of generalization of the resampled maps: the coarser the resolution, the worse is the model performance; between resampling methods, MR leads to lower scores than NN, at least in maps resampled at 50m.

Simulations obtained from maps at the same spatial resolution but resampled through different methods show the highest agreement between simulated changes (Table 5). More than 7 of each 10 ha that change are simulated in the same location when only the resampling method is changed. When changing the spatial resolution, this agreement is much lower, without meaningful differences because of the resampling method employed.

Resampling the maps through a different method does make a difference in the simulated quantities of change (Table 5). Whereas simulated quantities show a significative difference when using maps resampled through MR, simulations run with maps resampled through NN do not show meaningful differences respect to the quantities simulated in the exercise run with maps at the original resolution.

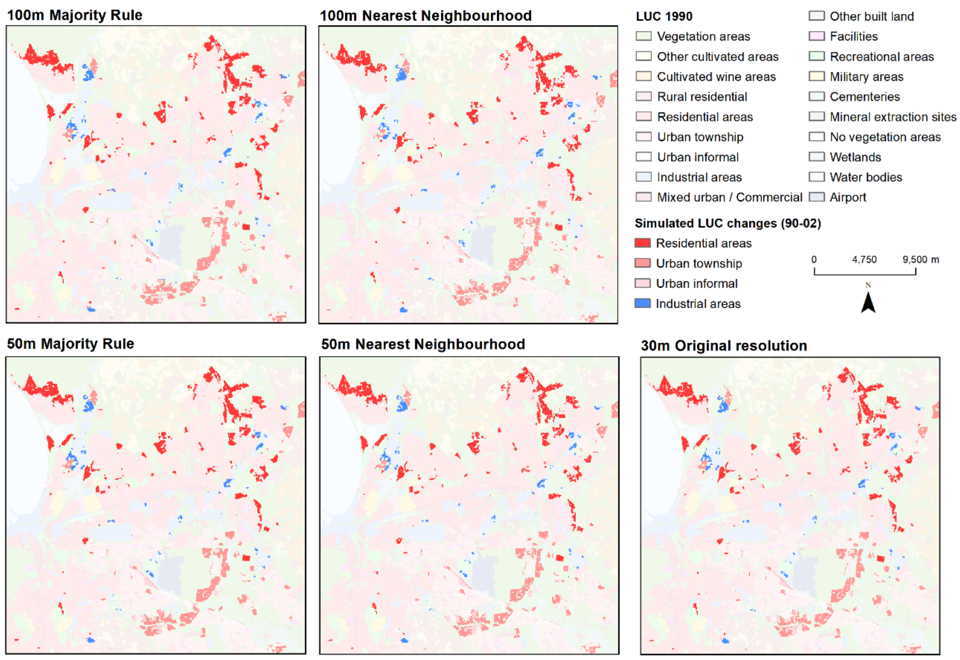
Table 5. Quantity and allocation (dis)agreement between simulated changes. Simulated changes are compared in pairs: first changes simulated from maps at the same spatial resolution but through different resampling methods (Nearest Neighbour, NN; Majority Rule, MR) and then from maps simulated at different spatial resolutions and through

different methods. Result at the category level may be found as supplementary material to the online version of this paper.

	Agreement	Quantity disagreement	Allocation disagreement
50m NN vs 50 MR	74.52	5.74	19.75
100m NN vs 100 MR	72.86	5.38	21.76
100m NN vs 50 NN	49.59	0.46	49.95
100m MR vs 50 MR	51.63	0.43	47.94
50m NN vs 30m	62.12	0.32	37.56
50m MR vs 30m	60.39	5.55	34.07
100m NN vs 30m	42.81	0.45	56.74
100m MR vs 30m	42.81	4.37	52.82

The simulated pattern is visually very similar independent of the spatial resolution and resampling method. No visual important differences are observed between simulations run with maps at different spatial resolutions and obtained through different resampling methods (Fig. 6). However, when making this analysis in detail, spatial metrics inform us about some differences between simulations.

Figure 6. Simulated changes in exercises using maps at different spatial resolutions (30, 50 and 100m) and obtained through different resampling methods for an example area of the City of Cape Town. Readers may find the full simulations for all the modelling area as supplementary material to the online version of this paper.



The fragmentation of input maps is always lower at coarser spatial resolutions and in maps resampled through MR (Table 6). After the simulation, the map´s fragmentation (number of patches) varies more at finer resolutions and, to a lesser extent, when using NN resample. The map comes usually more compact because of the action of the CA model component, with then finer resolutions maps and the ones obtained through NN being relatively more affected by this behaviour.

Some differences may be pointed out between categories. Urban residential fragmentation increases at finer resolutions and decreases in simulations at 100m. The other three function categories always come more compact. Urban informal is the class more affected by this compactness effect, given it is the category with the input pattern more fragmented. In this regard, most of the changes simulated for urban informal allocate next to the largest patch of urban informal, especially in the exercises at finer resolutions, as the LPI metric shows (Table 6).

Table 6. Spatial metrics of the input and simulated maps at different spatial resolutions and obtained through different resampling methods: Nearest Neighbour (NN) and Majority Rule (MR). Values of the metrics refer to a standard scale where 100 means the value of the metric in the input map (2002) at the original resolution (30m). Values above or below 100 mean a value bigger or smaller than the one used as reference. E.g. the number of patches of the simulation of the exercise calibrated at 30m of spatial resolution is 11% bigger than the number of patches of the LUC map at 30m for the year 2002. Readers may found all calculated metrics, standardized and with their original values, in the supplementary material attached to the online version of this paper.

	Number of patches				**Largest Patch Index**			
	Residential	**Township**	**Informal**	**Industrial**	**Residential**	**Township**	**Informal**	**Industrial**
Input LUC maps (2002)								
30m O	100	100	100	100	100	100	100	100
50m NN	86	84	78	92	32	32	38	31
50m MR	60	60	38	71	38	33	37	33
100m NN	58	61	47	72	39	35	36	30
100m MR	48	50	31	60	41	69	31	31
Simulated LUC maps (2002)								

30m O	111	84	79	92	103	105	289	101
50m NN	89	74	57	86	33	33	96	32
50m MR	66	59	30	68	39	33	101	33
100m NN	53	54	39	69	43	78	75	30
100m MR	44	47	25	58	45	78	56	31

6. Discussion

Resampling the input maps of a modelling exercise impacts the calibration of a LUCC model and the landscape it simulates. In our study case, this impact was mostly driven by the spatial resolution, with little variations because of changes in the resampling method employed.

Section 6.1 addresses the impact of the chosen resampling method on the modelling exercise. Section 6.2 specifically focuses on the impact of changes in the spatial resolution. The effects of the interaction between the spatial resolution and resampling method are not further discussed in any section as the impact of the resampling method was similar at the two spatial resolutions considered. Thence, we do not expect it to be dependent on the cell size.

6.1 The impact of the resampling method on the model calibration and simulation

Modelling exercises at the same spatial resolution but run with maps obtained through different resampling methods were very similar, with only slight differences between them. The most important change refers to the different neighbourhood rules defined in each case, which however did not end in meaningful differences between simulations.

Performance of exercises run with maps obtained through the MR method was usually better. However, the pattern is simpler in maps resampled through this method compared to the ones resampled through the NN method. In this regard, NN resampling preserves the original pattern of the resampled map and do not alter meaningfully the proportions of the map's categories. MR resampling simplifies the pattern and alter these proportions. In fact, quantity disagreement between simulations was bigger because of the resampling method employed than because of changes in the spatial resolution. However, focusing on the simulated pattern, the one modelled through exercises run with MR maps fitted better with the pattern of the reference maps than the one modelled through exercises run with NN maps, as we will see in the next section.

According to Hall et al. (2014), MR resampling favours the dominant classes over the rare ones. They recommend to choose NN resampling when rare categories are important to explain the processes of change and MR resampling if dominant categories are the most important ones for that purpose. Oyana et al. (2014) make similar appreciations when assessing the impact of the spatial resolution and resampling method in LUC change analysis for other resampling schemes, including the MR one. Accordingly, the selection of the resampling method, together with the selection of the proper spatial resolution, must agree with the landscape properties (Le Coz et al. 2009).

When choosing between one or another resampling method, the modeller must strike a balance between model performance, adequacy of the simulated pattern and realism. In our case, model performance was only a bit better for the exercises with maps obtained through the MR method. Explanatory power of factors was almost the same in both cases. On the contrary, differences were detected between simulations obtained with MR and NN maps in terms of pattern and realism. Consequently, NN maps are preferred.

MR maps should be chosen in those cases when the simplification that this resampling method introduces in the rasterized landscape comes with significative changes in the model performance, which was not our case. Notwithstanding, previous studies analysing maps at several spatial resolutions and obtained through diverse resampling methods have proved important differences between maps because of these decisions (He et al. 2002; Christman 2010; Díaz-Pacheco et al. 2018).

6.2 The impact of the spatial resolution on the model calibration and simulation

Most of the detected differences between calibrations and simulated landscapes were caused by the spatial resolution. It is one of the main drivers behind the computer requirements a model needs. In addition, together with the resampling method employed, affects the composition and configuration of the rasterized landscape.

At finer resolutions, models need to handle more information, which increases their computing needs. This may make the model crash or slow down its performance. Our application did not work at the original resolution and extent. Kim and Batty (2011) reported the crash of the model SLEUTH when setting up an application at 50m. Van Vliet et al. (2013b) and Jantz and Goetz (2005) resampled their input maps at 500m and 45m respectively to fit them with the model requirements. On the other hand, when the model performance is slower, calibration on a trial and error basis is more difficult. Also the engagement of audience and stakeholder in model calibrations if it takes a long time to run (Van Delden et al. 2011). Nonetheless, modellers may work with simpler models to this end.

Changes in the spatial resolution deeply affect the landscape configuration, as proved by Fassnacht et al. (2006), Ozdogan and Woodcock (2006) or Dendoncker et al. (2007), among others. Changes in the landscape composition are less significative, being the resampling method more important to this end.

LUCC models usually try to fit the simulated landscape to the one used as reference (Mas et al. 2014; National Research Council 2014). The more different is the reference landscape to be simulated, the more likely is that the model parameters will vary to a greater extent. Different to the resampling method, the spatial resolution substantially changed our reference landscapes configurations. That is why, in our case, model parameters and simulated pattern varied much more because of the spatial resolution than because of the resampling method employed.

Variation in model parameters do not have to be the same at all resolutions, neither for all categories. Whereas at some resolutions, the landscape configuration keeps similar (*scale domain*), once a *scale threshold* is exceeded, landscape configuration or properties change meaningfully (Ménard and Marceau 2005). Consequently, its processes and dynamics must be explained by different parameters. In our analysis, at 30 and 50m of resolution, exercises were calibrated through similar parameters. At 100m, we crossed the scale threshold that changed the landscape properties. Accordingly, at that resolution, model parameters were much different.

At the category level, each category has a specific pattern and, therefore, is differently affected by variations in the spatial resolution. For our analysis, at 30 and 50m the urban informal pattern of change is partially explained by the contiguity of new urban informal to previous developments of the same use, with this factor being much more important at the finest resolution. At 100m, this factor does not explain the new urban informal growth. Thus, when choosing the correct spatial resolution, we must not only consider the sensitivity of the global landscape to this change, but also how each category is individually affected by it.

Despite of the different spatial resolutions, all exercises considered similar neighbourhoods' extents in terms of cells: 6 cells of maximum neighbourhood influence. This means that the neighbourhood extent in meters is very different depending on the spatial resolution of the exercise: from a maximum of 180m in the exercise at the original resolution of input maps (30m) to 600m in the exercises run with maps at 100m. Accordingly, we can confirm one of the hypotheses laid out by Díaz-Pacheco et al.

(2018), who checked the same behaviour for another application: "cell is more important than the actual distance (in meters) in the calculation of the neighbourhood effect". Thus, although neighbourhood influence in CA models may be generally explained by real interactions between land uses in terms of distances, when detailing this influence, the way the simulated landscape is conceptualized through a raster plays a key role.

The simulated pattern is not very different between exercises at different spatial resolutions. In all cases, the landscape comes more compact because of the action of the CA model component. One of the main rules that usually drives its behaviour is the self-attraction of the class actively modelled. Accordingly, the simulated changes always tend to fill existing patches or put some of them together, making its shape more compact.

When compared to the reference landscape, the exercises at finer resolutions are the ones that simulate a more different landscape. The CA compactness logic previously explained fits better with more generalized maps than with the ones more fragmented. That is why the pattern of simulated changes is more similar to the pattern of reference changes in the exercises at coarser resolutions and with maps resampled through MR.

In addition, changes usually take place as a group of pixels at fine resolutions, whereas they are just made up of a few pixels in maps at coarser resolutions. Thence, simulation of the correct pattern is usually easier at coarser resolutions, as checked by Díaz-Pacheco et al. (2018) in a modelling exercise for Madrid (Spain). In that case, the coarser pixel size fit better with the size of cadastral parcels in this city and, accordingly, with its urbanization pattern. In this regard, cadastral parcels are usually the unit over which change operates in urban environments.

When choosing the proper spatial resolution, the modeler must strike a balance between model performance, computing needs of the model, desired pattern and realism. Exercises at 30 and 50m provided a similar pattern and realism. On the other hand, exercises at 100m, although very fast, got a less realistic pattern, very compacted. In all cases, model performance was very similar. Accordingly, 50m is the optimum resolution according to our criteria, because it deeply lightened the computing resources that the model demanded while producing an adequate pattern and degree of realism. However, the best resolution will be different for each case study, according to the properties of the modelled landscape (Ozdogan and Woodcock 2006).

7. Conclusions

This study has revealed how, for our study case, the modelling exercise was not very sensitive to the method employed to resample input Land Use Cover Maps. Calibration parameters and computational requirements were much more different because of changes in the spatial resolution. However, in all cases, the obtained simulations were similar. Differences in calibration parameters, but not in simulated patterns and changes, prove that user intervention is behind the similarity of simulated landscapes. Accordingly, manual calibration may be considered a good tool to deal with the sensitivity of Land Use Cover Change models to changes in the spatial resolution and, to a lesser extent, resampling method.

Studies like this one aid to understand the way users must analyse the sensitivity of their exercises to key decisions like the spatial resolution and resampling method. They may be also used as a reference for those exercises with similar landscape characteristics. However, they will never be able to provide a unique answer to the problem. In this regard, there is not a unique and universal answer to aid in the selection of the proper spatial resolution and resampling method. These decisions are case dependent. Users must be fully aware of these uncertainties and their complexity when calibrating their models. Depending on the balance between model performance, simulated pattern, realism and audience needs, one or another resolution and resampling method must be selected in each case.

Acknowledgements

This work has been supported by the project INCERTIMAPS PGC2018-100770-B-100 (Spanish Ministry of Science, Innovation and Universities and the Feder European Regional Development Fund). The first author is also grateful to the Spanish Ministry of Economy and Competitiveness, the European Social Fund and the University of Granada for the funding of his research activity (Ayudas para contratos predoctorales para la formación de doctores 2014; Contratos Puente 2018). We also thank the support of the University of Cape Town to obtain the data and set up the application on which the analysis is built upon.

References

Altartouri A, Nurminen L, Jolma A (2015) Spatial neighborhood effect and scale issues in the calibration and validation of a dynamic model of Phragmites australis distribution - A cellular automata and machine learning approach. Environ Model Softw 71:15–29 . doi: 10.1016/j.envsoft.2015.04.010

Bach M, Breuer L, Frede HG, et al (2006) Accuracy and congruency of three different digital land-use maps. Landsc Urban Plan 78:289–299 . doi: 10.1016/j.landurbplan.2005.09.004

Benenson I (2007) Warning! The scale of land-use CA is changing! Comput Environ Urban Syst 31:107–113 . doi: 10.1016/j.compenvurbsys.2007.01.001

Blanchard SD, Pontius Jr. RG, Urban KM (2015) Implications of Using 2 m versus 30 m Spatial Resolution Data for Suburban Residential Land Change Modeling. J Environ Informatics 25:1–13 . doi: 10.3808/jei.201400284

Chen H, Pontius Jr. RG (2011) Sensitivity of a Land Change Model to Pixel Resolution and Precision of the Independent Variable. Environ Model Assess 16:37–52 . doi: 10.1007/s10666-010-9233-3

Christman ZJ (2010) Land change in central Mexico: Landscape heterogeneity, natural variability, and classification uncertainty. CLARK UNIVERSITY

City of Cape Town (2018) Municipal Spatial Development Framework

City of Cape Town (2012) Cape Town Spatial Development Framework. Statutory report. Cape Town

Cohen J (1960) A Coefficient of Agreement for Nominal Scales. Educ Psychol Meas 20:37–46

Dendoncker N, Rounsevell M, Bogaert P (2007) Spatial analysis and modelling of land use distributions in Belgium. Comput Environ Urban Syst 31:188–205

Dendoncker N, Schmit C, Rounsevell M (2008) Exploring spatial data uncertainties in land-use change scenarios. Int J Geogr Inf Sci 22:1013–1030 . doi: 10.1080/13658810701812836

Díaz-Pacheco J, Van Delden H, Hewitt R (2018) The importance of scale in land use models: experiments in data conversion, data resampling, resolution and neighbourhood extent. In: Camacho Olmedo MT, Paegelow M, Mas J-F, Escobar F (eds) Geomatic Approaches for Modeling Land Change Scenarios. Springer, Cham, Switzerland, pp 163–186

Dietzel C, Clarke KC (2004) Replication of Spatio-temporal Land Use Patterns at Three Levels of Aggregation by an Urban Cellular Automata. In: Sloot PMA, Chopard B, Hoekstra AG (eds) 6th International Conference on Cellular Automata for Research and Industry, ACRI 2004. Springer, Amsterdam, pp 523–532

ESRI Resample. In: ArcGIS Pro Help. https://pro.arcgis.com/en/pro-app/tool-reference/data-management/resample.htm. Accessed 22 Feb 2019

Fassnacht KS, Cohen WB, Spies TA (2006) Key issues in making and using satellite-based maps in ecology: A primer. For Ecol Manage 222:167–181 . doi: 10.1016/j.foreco.2005.09.026

Ficklin DL, Tan ML, Chaplot V, et al (2015) Impacts of DEM resolution, source, and resampling technique on SWAT-simulated streamflow. Appl Geogr 63:357–368 . doi: 10.1016/j.apgeog.2015.07.014

García Martínez ED, Chas-Amil ML, Touza J (2015) Assessment of the Spanish land cover information to estimate forest area in Galicia. Boletín la Asoc Geógrafos Españoles 69:333–350

GeoTerra Image (2015a) 1990 South African National Land Cover Dataset (including 1990-2013/14 land-cover change comments). Data User Report and MetaData

GeoTerra Image (2015b) 2013-2014 South African National Land-Cover Dataset. Data User Report and MetaData. 2:1–55

Hall JM, Staub CG, Marsik MP, et al (2014) Scaling categorical spatial data for earth systems models. Glob Chang Biol 21:1–3 . doi: 10.1111/gcb.12708

He HS, Ventura SJ, Mladenoff DJ (2002) Effects of spatial aggregation approaches on classified satellite imagery. Int J Geogr Inf Sci 16:93–109 . doi: 10.1080/13658810110075978

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 . doi: 10.1016/j.compenvurbsys.2016.10.011

Jantz CA, Goetz SJ (2005) Analysis of scale dependencies in an urban land-use-change model. Int J Geogr Inf Sci 19:217–241 . doi: 10.1080/13658810410001713425

Kim D, Batty M (2011) Calibrating Cellular Automata Models for Simulating Urban Growth: Comparative Analysis of SLEUTH and Metronamica

Kim JH (2013) Spatiotemporal scale dependency and other sensitivities in dynamic land-use change simulations. Int J Geogr Inf Sci 27:1782–1803 . doi: 10.1080/13658816.2013.787145

Kocabas V, Dragicevic S (2006) Assessing cellular automata model behaviour using a sensitivity analysis approach. Comput Environ Urban Syst 30:921–953 . doi: 10.1016/j.compenvurbsys.2006.01.001

Le Coz M, Delclaux F, Genthon P, Favreau G (2009) Assessment of Digital Elevation Model (DEM) aggregation methods for hydrological modeling: Lake Chad basin, Africa. Comput Geosci 35:1661–1670 . doi: 10.1016/j.cageo.2008.07.009

Li H, Wu J (2006) Uncertainty analysis in ecological studies: an overview. In: Wu J, Jones KB, Li H, Loucks OL (eds) Scaling and Uncertainty Analysis in Ecology: Methods and Applications. Springer, Dordrecht, pp 45–66

Mas J-F, Kolb M, Paegelow M, et al (2014) Inductive pattern-based land use/cover change models: A comparison of four software packages. Environ Model Softw 51:94–111 . doi: 10.1016/j.envsoft.2013.09.010

McCallum I, Obersteiner M, Nilsson S, Shvidenko A (2006) A spatial comparison of four satellite derived 1 km global land cover datasets. Int J Appl Earth Obs Geoinf 8:246–255 . doi: 10.1016/j.jag.2005.12.002

McGarigal K, Cushman SA, Neel MC, Ene E (2015) FRAGSTATS: Spatial Pattern Analysis Program for Categorical and Continuous Maps

Ménard A, Marceau DJ (2005) Exploration of spatial scale sensitivity in geographic cellular automata. Environ Plan B Plan Des 32:693–714 . doi: 10.1068/b31163

Morais Viana C (2014) A influência do efeito de escala nos modelos de simulação baseados em autómatos celulares. Universidade de Lisboa

National Research Council (2014) Advancing Land Change Modeling: Opportunities and Research Requirements. National Academies Press, Washington, D.C.

Oyana TJ, Johnson SJ, Wang G (2014) Landscape metrics and change analysis of a national wildlife refuge at different spatial resolutions. Int J Remote Sens 35:3109–3134 . doi: 10.1080/01431161.2014.903443

Ozdogan M, Woodcock CE (2006) Resolution dependent errors in remote sensing of cultivated areas. Remote Sens Environ 103:203–217 . doi: 10.1016/j.rse.2006.04.004

Pan Y, Roth A, Yu Z, Doluschitz R (2010) The impact of variation in scale on the behavior of a cellular automata used for land use change modeling. Comput Environ Urban Syst 34:400–408 . doi: 10.1016/j.compenvurbsys.2010.03.003

Pearson K (1895) Note on regression and inheritance in the case of two parents. Proc R Soc London 58:240–242 . doi: 10.1098/rspl.1895.0041

Pieterse E (2009) Post-Apartheid Geographies in South Africa: why are urban divides so persistent? In: Interdisciplinary Debates on Development and Cultures: Cities in Development - Spaces, Conflicts and Agency. p 15

Pontius Jr. RG (2018) Pontius Matrix

Pontius Jr. RG, Boersma W, Castella J-C, et al (2008) Comparing the input, output, and validation maps for several models of land change. Ann Reg Sci 42:11–37 . doi: 10.1007/s00168-007-0138-2

Ran Y, Li X, Lu L (2010) Evaluation of four remote sensing based land cover products over China. Int J Remote Sens 31:391–401 . doi: 10.1080/01431160902893451

Rebelo AG, Holmes PM, Dorse C, Wood J (2011) Impacts of urbanization in a biodiversity hotspot: Conservation challenges in Metropolitan Cape Town. South African J Bot 77:20–35 . doi: 10.1016/j.sajb.2010.04.006

RIKS (2012) Metronamica Documentation. Maastricht

Samat N (2006) Characterizing the scale sensitivity of the cellular automata simulated urban growth: A case study of the Seberang Perai Region, Penang State, Malaysia. Comput Environ Urban Syst 30:905–920 . doi: 10.1016/j.compenvurbsys.2005.11.002

Santé I, García AM, Miranda D, Crecente R (2010) Cellular automata models for the simulation of real-world urban processes: A review and analysis. Landsc Urban Plan 96:108–122 . doi: 10.1016/j.landurbplan.2010.03.001

Town C, Turok I (2001) Persistent Polarisation Post-Apartheid? Progress towards Urban Integration in Cape Town. Urban Stud 38:2349–2377 . doi: 10.1080/00420980120094551

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 142:18–28 . doi: 10.1016/j.agee.2011.03.005

Van Vliet J, Bregt AK, Brown DG, et al (2016) A review of current calibration and validation practices in land-change modeling. Environ Model Softw 82:174–182 . doi: 10.1016/j.envsoft.2016.04.017

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 222:1367–1375 . doi: 10.1016/j.ecolmodel.2011.01.017

Van Vliet J, Hagen-Zanker A, Hurkens J, Van Delden H (2013a) A fuzzy set approach to assess the predictive accuracy of land use simulations. Ecol Modell 261–262:32–42 . doi: 10.1016/j.ecolmodel.2013.03.019

Van Vliet J, Naus N, Van Lammeren RJA, et al (2013b) Measuring the neighbourhood effect to calibrate land use models. Comput Environ Urban Syst 41:55–64 . doi: 10.1016/j.compenvurbsys.2013.03.006

Warmink JJ, Janssen JAEB, Booij MJ, Krol MS (2010) Identification and classification of uncertainties in the application of environmental models. Environ Model Softw

25:1518–1527 . doi: 10.1016/j.envsoft.2010.04.011

Western J (1997) Outcast Cape Town

White R, Engelen G (1993) Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. Environ Plan A 25:1175–1199 . doi: 10.1068/a251175

White R, Engelen G (1997) Cellular automata as the basis of integrated dynamic regional modelling. Environ Plan B Plan Des 24:235–246 . doi: https://doi.org/10.1068/b240235

White R, Engelen G, Uljee I (1997) The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. Environ Plan B Plan Des 24:323–343 . doi: 10.1068/b240323

Wilkinson P (2004) Renegotiating local governance in a post-apartheid city: The case of Cape Town. Urban Forum 15:213–230 . doi: 10.1007/s12132-004-0001-9

Wilkinson P (2000) City profile: Cape Town. Cities 17:195–205 . doi: 10.1016/S0264-2751(99)00059-1

Wu H, Li Z, Clarke KC, et al (2019) Examining the sensitivity of spatial scale in cellular automata Markov chain simulation of land use change. Int J Geogr Inf Sci 00:1–22 . doi: 10.1080/13658816.2019.1568441

Wu J (2007) Scale and scaling: a cross-disciplinary perspective. In: Wu J, Hobbs RJ (eds) Key Topics in Landscape Ecology. Cambridge University Press, Cambridge, New York, pp 115–142

Yeh AG-O, Li X (2006) Errors and uncertainties in urban cellular automata. Comput Environ Urban Syst 30:10–28 . doi: 10.1016/j.compenvurbsys.2004.05.007

Zhao G (2013) Effects of Spatial Scale in Cellular Automata Model for Land Use Change. In: Gaol FL (ed) Recent Progress in Data Engineering and Internet Technology. Springer, Berlin, Heidelberg, pp 101–106

Zhao S, Liu S (2014) Scale criticality in estimating ecosystem carbon dynamics. Glob Chang Biol 20:2240–2251 . doi: 10.1111/gcb.12496