Neutral models of landscape change as benchmarks in the assessment of model performance

Alex Hagen-Zanker and Gilles Lajoie

Under review

5

Abstract

This paper introduces a methodological framework for performance assessment of spatial dynamic models by means of map comparison. The objective is to discern to what extent model performance, expressed by a variety of metrics, can be attributed to processes that are endogenous to the model or to exogenous model inputs. For this purpose, neutral models of landscape change are introduced that are subject to the same boundary conditions and constraints as the probed model, but otherwise represent no specific processes except reluctance to change. The neutral models serve as benchmark and the difference in performance with the model under investigation can be attributed to the endogenous qualities of the model. Furthermore, the framework makes performance measures over multiple criteria and scales mutually comparable and helps identifying strengths and weaknesses of the model.

The framework is applied for the performance assessment of a Constrained Cellular Automata land use model for La Réunion (Fr.). Map comparison metrics of land use presence and structure are evaluated at multiple scales. For criteria of land use presence the land use model outperforms the neutral models only at coarse scales, but for criteria of land use structure it performs better on all scales.

Keywords: map comparison, neutral model, spatial dynamics

1 Introduction

Assessing model performance is a continuous challenge for modelers of landscape dynamics. A common approach is historical validation (Rykiel 1996) where a predicted map is compared to an actual map. Recently, numerous map comparison methods have been proposed that take into account the spatial relation between cells, as opposed to simple cell-by-cell overlap. These new methods consider for example proximity (Fewster and Buckland 2001; Hagen-Zanker et al. 2005; Kuhnert et al. 2005), the presence of recognizable structures, i.e. features (Ebert and McBride 2000; Power et al. 2000) and information contained at different scales, whereby coarser scales are found by aggregation (Costanza 1987; Pontius et al. 2004; Remmel and Csillag 2006); moving windows (Hagen-Zanker 2006; Pijanowski et al. 2002) or wavelet decomposition (Briggs and Levine 1997; Zepeda-Arce et al. 2000). Others have evaluated model performance on the basis of metrics summarizing the whole landscape (Barredo et al. 2003; Turner et al. 1989; White et al. 1997)

The diversity of these methods and their possibly conflicting outcomes calls for a methodological framework to apply them in a coherent fashion. This paper seeks to answer the following questions:

1. How can diverse metrics of model goodness-of-fit be mutually compared to indicate strengths and weaknesses of the evaluated model?

2. To what extent must goodness-of-fit be attributed to modeled processes or to boundary conditions and constraints?

The second question may seem more abstract, but in fact refers to a real and persistent problem; It is not uncommon for land use models to attain a percentage correct of about 95%, but this 'very good performance' is then due to the fact that the landscape at the end of the simulation period is largely identical to the given initial situation. Another common constraint on land use models is the total area per land use class. This is applied when in fact the model is not intended to represent the change in area (quantity) of land use classes, but only the spatial distribution. The exogenously imposed areas per land use class will influence model accuracy, regardless of the adequacy of the modeled processes. These issues have been only partially addressed before in studies that account for the impact of the initial situation, but not for other constraints (Hagen 2003; Pontius et al. 2004).

Neutral landscape models that are common in landscape ecological studies, may be useful to this problem. Neutral models are algorithms that create landscape patterns in absence of specific processes (With and King 1997). Comparison of the performance of the investigated model to that of neutral models can help determining to what extent the model performance must be attributed to processes that are absent in the neutral model.

'Textbook' neutral landscape models (Turner et al. 2001, chapter 6) are not appropriate as reference for dynamic models however. These models create a landscape starting from a blank or randomized initial situation and therefore can not account for the effect of the initial situation. A recent extension to neutral landscape models (Gardner and Urban 2007) applies a mask to separate the landscape in static (not changing) and dynamic (changing) region. In that extension, however, the landscape in the dynamic region is still created from a blank initial situation and the model is therefore not adequate for the purpose of this paper either.

To resolve this shortcoming, this paper introduces a new class of neutral landscape models: *neutral models of landscape change*. These models do not create a landscape from a blank initial situation, but instead modify an existing initial landscape. These models are subject to the same boundary condition and constraints as the probed model and therefore pose an adequate reference level. Turner (1987) and Pontius et al. (2007) provide earlier applications of simple models of landscape change as benchmarks for more complex ones.

The evaluated model in the case study is a Constrained Cellular Automata (CCA) land use model (Engelen et al. 2003; White et al. 1997; White and Engelen 1993) calibrated for the island La Réunion. The comparison methods used within the framework combine the wavelet decomposition approach of Briggs and Levine (1997) with the structure indicator maps of Hagen-Zanker (2006). This is the first application where these two approaches are combined. This combination is particularly suited to test and demonstrate the methodological framework since it presents diverse indicators of model performance at multiple scales.

2 Methods

2.1 General procedure

The performance of spatial models is assessed by running them for a period in the past and comparing the output of the model to reality. Model and reality are compared on a number of criteria that are evaluated at multiple scales. These diverse indicators of model performance are normalized to the real change that occurred over the simulation period, measured according to the same criteria. This means of normalization reflects the need to account for persistence already

observed by Tobler (1970) and is inspired by recent findings of Pontius et al. (in press) that model performance correlates strongly with the amount of change over the simulation period.

The comparison is not just made for the evaluated model, but also for one or more neutral models of landscape change. These models are subject to the same boundary conditions and constraints as the evaluated model. Otherwise the neutral models are conservative, which means that they minimize change. What constitutes minimal change remains ambiguous, therefore multiple neutral models of landscape change may be applied. The model performance is expressed relative to that of the neutral models of landscape change.

Figure 1 presents a flowchart of the general procedure. The specific neutral models, performance criteria and means of normalization that are applied in this paper are given in the following sections.

Figure 1. Flowchart of model performance evaluation. The dashed lines indicate that Model constraints are not necessarily based on the initial and final map of reality, but in the current case they are.
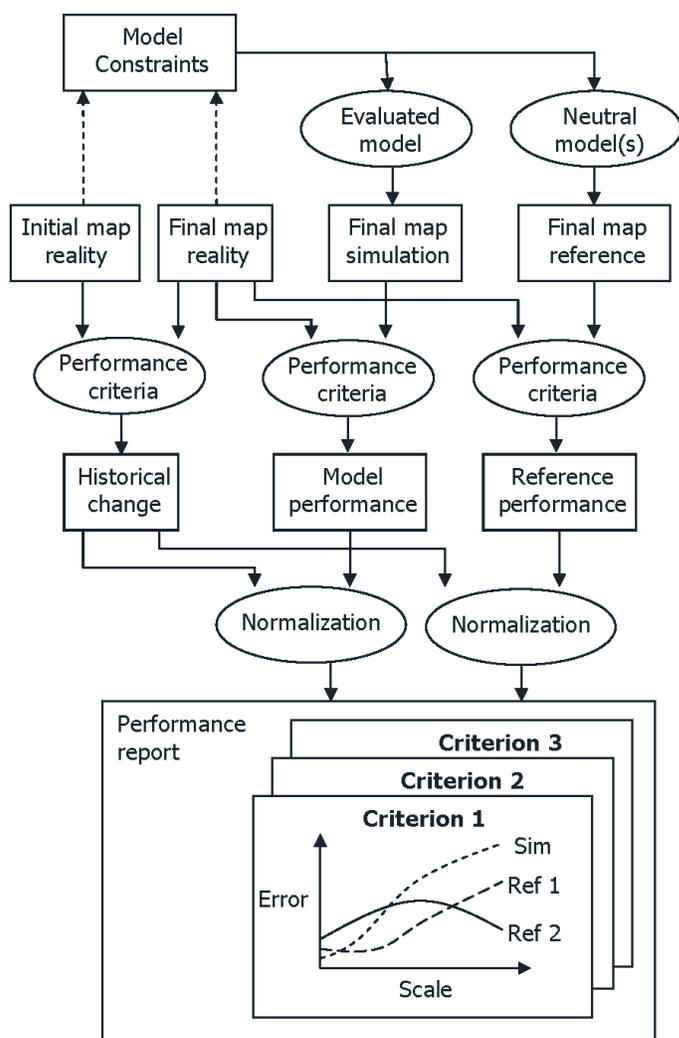
2.2 Neutral models of landscape change

This section introduces two neutral models of landscape change. Both neutral models modify the initial map to the effect that it has the same composition (total area per class) as the simulated map. This is the same constraint that is also applied in the evaluated model. Both neutral models

5 follow the notion of changing as little as possible, the difference is that the random constraint match model only minimizes the area of change, whereas the growing clusters model also minimizes the change in the spatial distribution of classes, by placing new cells of a class adjacent to existing cells of that class.

Pseudo code of both models is given in the appendix.

10 ### 2.2.1 Neutral model: Random constraint match

The random constraint match model finds locations of change randomly and evolves towards a 'speckled' map of small clusters.

The model first assesses for each class by how many cells it is under- or overrepresented in the initial map relative to the constraints. For each overrepresented class it then randomly selects the

15 surplus cells on the initial map. Then, the underrepresented classes are randomly distributed over the selected cells on the initial map.

For example, consider a dataset of four land use classes (agriculture, urban, nature, water). Agriculture and nature become less prevalent over time, while water and urban are gaining. Initially there are 450 cells of agriculture whereas the constraint for the end situation is 410,

20 which means that the class agriculture is overrepresented by 40 cells. Out of the 450 cells of agriculture, 40 are selected to change class. Likewise, the class nature is overrepresented by 10 cells and out of the initial nature cells 10 are selected. In total 50 cells have been selected to change class. The class urban is underrepresented by 45 cells and the class water by 5 cells. Thus out of the 50 cells that will change class, 45 cells are randomly selected. These change to the class

25 urban. The remaining 5 cells change to the class water. Now all constraints are satisfied.

2.2.2 Neutral model: Growing clusters

The growing clusters model find locations for changes along the edges of existing patches (clusters). This model evolves towards 'blobby' map of large patches.

The model first counts cells for all classes both in the initial map and in the target map. The

30 model then compiles a list of all edges in the initial map where an underrepresented category borders an overrepresented category. Then, it repeatedly selects edges from this list and changes the cell with the overrepresented category to take the value of the underrepresented category. The transition will only take place if that cell has not changed yet and the categories along the edge are still overrepresented, respectively underrepresented.

35 It is possible that the list of edges is exhausted before all required changes are made. In that case the procedure is repeated with the adjusted map as the new initial map. Another possibility is that no more edges between over- and underrepresented classes are found. In that case, for each underrepresented class randomly a cell of an overrepresented class is found and the transition is made. The new random changes can be the seeds of growing clusters and the procedure is

40 repeated until all constraints have been met.

2.3 Performance criteria

Real and model maps are compared on several characteristics that are selected on the basis of the intended and perceived strengths of the evaluated model. The spatial interaction of the Constrained Cellular Automata land use model is in particular expected to represent the dynamic processes underlying urban morphology, therefore the performance is measured in terms of both structure and presence at multiple scales. With 'presence' is meant the extent to which occurrence of a class coincides in both maps. Likewise, with 'structure' is meant the extent to which the spatial configuration is similar values at the same location in both maps.

The structure metrics that are applied are based on the patch size and patch fractal dimension metrics for individual land use classes. Urban patch size is included because we perceive it to be the main criterion in the visual assessment of map similarity for this kind of model. Urban fractal dimension is included because earlier model assessments put emphasis on the area to perimeter relation of urban clusters.

The class agriculture is only evaluated in terms of presence, because this class is modeled primarily on the basis of site characteristics and not spatial interaction. Multi-scale analysis is appropriate since the site characteristics have multi-scale features.

2.3.1 Criterion maps

The first step of the comparison derives criterion maps from the two compared land use maps for each of the criteria introduced above. The criterion maps for urban and agricultural presence have value 1 where the respective class is present and 0 elsewhere. Urban patch size and urban patch fractal dimension criterion maps give every cell the value of the patch that it belongs to, i.e. an urban cell belonging to a patch with size s, obtains value s.

The identification of patches is based on Queen Contiguity, i.e. diagonally adjacent cells are considered contiguous. Patch size is simply the number of cells in a patch. Patch fractal dimension (**Equation 1**) is based on the FRAGSTATS definition (McGarical et al, 2002), which means that 'fractal dimension' is a shape characteristic and not a fractal dimension in mathematical terms (Frankhauser 1994).

$$ F = \frac{\log(a)}{\log\left(\frac{p}{4}\right)} \tag{1} $$

where:

F fractal dimension of a patch

a area of the patch expressed in cell units

p perimeter of the patch in cell units.

2.3.2 Multi-scale analysis by wavelet transformation

A wavelet transformation is used to decompose the criterion maps into criterion maps at multiple scales. The idea of the transformation is that large clusters of similar values are registered at coarse scales (large wavelengths) and small clusters of similar values at fine scales (small wavelengths). The transformation is quite different from block and moving window aggregation techniques that are more commonly used. In aggregation approaches coarse scale information remains present at finer scales implying that some errors are registered at multiple scales. This is not the case with wavelet transformation since it finds different scales by separation, not

aggregation. Another attractive property of the wavelet approach is that the squared difference between two transformed maps over all scales amounts to the difference of the original untransformed maps. Therefore it can be said that the wavelet approach decomposes the difference between maps into multiple scales.

5   Wavelet transformation is a technique that transforms an original signal consisting of *n* values to an accumulation of the same number of *n* coordinates, each corresponding to one from a family of wavelets. These wavelets are template patterns that all have the same basic shape (resembling a localized wave), but different amplitudes, wavelengths and phases. Each coordinate relates to a unique combination of a wavelength and a phase and the coordinate value itself expresses the
10   amplitude. Wavelets are commonly used in image processing, in particular image compression. For such two-dimensional applications the wavelet transformation is first applied horizontally and then vertically.

In this paper wavelet transformation separate the data into different scales. These separate scales are found by selecting from all coordinates of the wavelet transformation those that correspond to
15   the particular wavelength and perform an inverse wavelet transformation on those only. Colloquially said: The wavelet approach transforms the original dataset into a large collection of local waves. The comparison at a scale is based on only those waves whose wavelength corresponds to that particular scale.

There are different types of wavelets and one aspect differentiating them is whether they are
20   continuous or discrete. We follow Briggs and Levine (1997) and use discrete wavelets. This means that wavelengths come in powers of 2 and wavelets are arranged head to tail. For instance, a signal consisting of 16 values will be decomposed into 8 wavelets of length 2, 4 of length 4, 2 of length 8, 1 of length 16 and finally one value for the overall mean.

The second differentiating factor of wavelet types is their template pattern (i.e. the shape of the
25   wavelet), called mother wavelet. Restrictions to possible forms of mother wavelets arise from the condition that wavelets derived from a single mother wavelet need to be orthogonal and orthonormal. The simplest of mother wavelets is the Haar wavelet, most common are Daubechies wavelets.

This paper applies the Haar wavelet. It is not selected on the basis of simplicity, but following the
30   information based criterion recommended by Briggs and Levine (1997). A second motivation for using the Haar wavelet is that the earlier mentioned restrictions make it the only symmetrical wavelet and all others would introduce anisotropy. Although these are valid arguments, there are no overwhelming arguments against other mother wavelets. The choice of mother wavelet is, thus, more or less arbitrary and will be considered in the sensitivity analysis (section 4.2).

35   For an in depth discussion of wavelet based map comparison we refer to Briggs and Levine (1997) and Zepeda-Arce et al. (2000).

Equation 2 expresses that the original criterion map is decomposed into a series of maps by wavelet transformation.

$$C_{real} \xrightarrow{WT} \left\{ C_{real}^1, C_{real}^2 ..., C_{real}^{nmax} \right\}$$
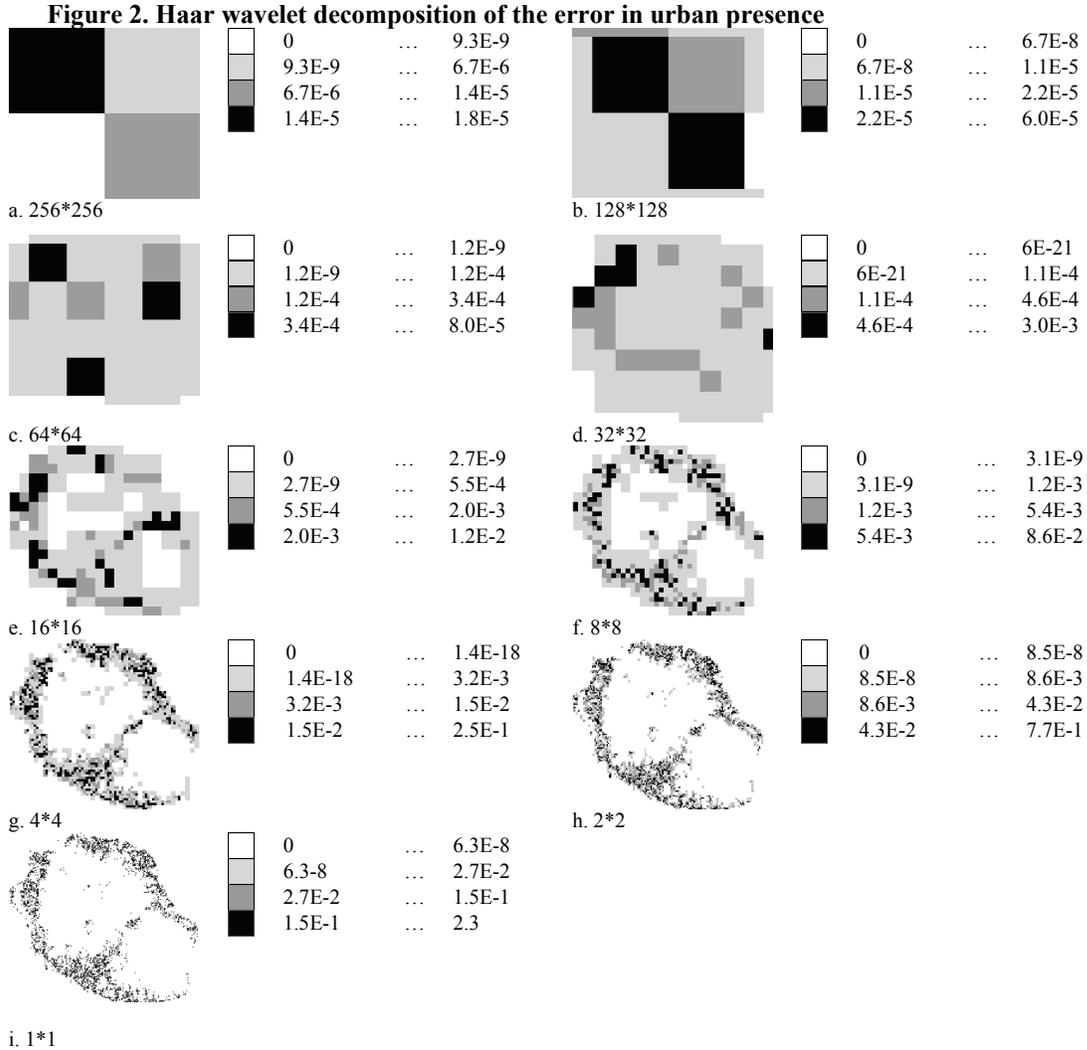$$C_{real}^1 + C_{real}^2 ... + C_{real}^{nmax} = C_{real}$$

(2)

40   where the numbers 1, 2,…,*nmax* indicate different scales. *C* is any of the 4 criterion maps. The wavelet transform is performed for the *real* as well as the *model* map.

The cell-to-cell errors at every scale are aggregated to a mean squared error (MSE).

$$MSE_{C,s}\left(real,model\right) = \frac{\sum\limits_{cell}\left(C^s_{real,cell} - C^s_{model,cell}\right)^2}{nCells} \tag{3}$$

where *C* is the evaluated criterion and *s* the scale at which it is evaluated. The index *cell* iterates over all cells of which there are *nCells*. As an illustration, **Figure 2** shows the squared error per cell for the criterion urban presence at multiple scales. The underlying data is presented in section 3.

Figure 2. Haar wavelet decomposition of the error in urban presence

	0	...	9.3E-9
	9.3E-9	...	6.7E-6
	6.7E-6	...	1.4E-5
	1.4E-5	...	1.8E-5

a. 256*256

	0	...	6.7E-8
	6.7E-8	...	1.1E-5
	1.1E-5	...	2.2E-5
	2.2E-5	...	6.0E-5

b. 128*128

	0	...	1.2E-9
	1.2E-9	...	1.2E-4
	1.2E-4	...	3.4E-4
	3.4E-4	...	8.0E-5

c. 64*64

	0	...	6E-21
	6E-21	...	1.1E-4
	1.1E-4	...	4.6E-4
	4.6E-4	...	3.0E-3

d. 32*32

	0	...	2.7E-9
	2.7E-9	...	5.5E-4
	5.5E-4	...	2.0E-3
	2.0E-3	...	1.2E-2

e. 16*16

	0	...	3.1E-9
	3.1E-9	...	1.2E-3
	1.2E-3	...	5.4E-3
	5.4E-3	...	8.6E-2

f. 8*8

	0	...	1.4E-18
	1.4E-18	...	3.2E-3
	3.2E-3	...	1.5E-2
	1.5E-2	...	2.5E-1

g. 4*4

	0	...	8.5E-8
	8.5E-8	...	8.6E-3
	8.6E-3	...	4.3E-2
	4.3E-2	...	7.7E-1

h. 2*2

	0	...	6.3E-8
	6.3-8	...	2.7E-2
	2.7E-2	...	1.5E-1
	1.5E-1	...	2.3

i. 1*1

2.4 Normalization

The change over the simulation period is used as a means of normalization. The change over the simulation period is measured according to the same metric that expresses the model error. The two neutral models described in the previous section are used as a reference level. In summary, the following five maps are input to the analysis.

1. INIT: Initial ground truth map (initial year)
2. FINAL: Final ground truth map (final year)

3. SIM: Evaluated land use model (final year)
4. RCM: Random Constraint Match (final year)
5. CLUST: Growing Clusters (final year)

Applying these abbreviations the normalized performance of the three models is calculated according to the following equations.

$$P_{C,s}^{SIM} = \frac{MSE_{C,s}(FINAL,SIM)}{MSE_{C,s}(FINAL,INIT)}$$

$$P_{C,s}^{RCM} = \frac{MSE_{C,s}(FINAL,RCM)}{MSE_{C,s}(FINAL,INIT)} \tag{4}$$

$$P_{C,s}^{CLUST} = \frac{MSE_{C,s}(FINAL,CLUST)}{MSE_{C,s}(FINAL,INIT)}$$

where P is the performance metric applied in this paper. C stands for any of the four criteria (agricultural presence, urban presence, urban patch size and urban fractal dimension) and s is the scale at which the model is evaluated. P is the final measure of goodness-of-fit and can be summarized in plots and tables to concisely express the performance at multiple scales and relative to the different reference models. It is an error metric, so lower values indicate better performance.

3 The case of La Réunion: study area, data and model

3.1 Land use model

The CCA (White et al. 1997) is a spatial explicit and dynamic land use model. The main variable of the model is the land use (raster) map. The state of every cell is its land use class (e.g. residential, industrial, forest, agriculture). Over the course of a simulation the land use map is iteratively updated, typically in time steps of one year.

The land use map is updated to satisfy the exogenous land area demands (the constraints). In other words, the model does not determine the total area of different land use classes, but only their spatial distribution.

The land use transitions are driven by the endogenous variable of land use potential, which is calculated for every location (cell) and each land use class. Land use potential is a function of several aspects, in particular neighbourhood effect, physical suitability, accessibility and zoning according to the following equation:

$$P_{k,a} = f\left(r_{k,a} N_{k,a} S_{k,a} A_{k,a} Z_{k,a}\right) \tag{5}$$

where:

$N_{k,a}$ potential for land use class k at location a

$r_{k,a}$ Random perturbation factor for land use class k at location a

$N_{k,a}$ neighbourhood effect for land use class k at location a

$S_{k,a}$ Physical suitability for land use class k at location a

$A_{k,a}$ Accessibility of land use class k at location a

$Z_{k,a}$ Zoning status for land use class k at location a

f(...)   Total potential function. The precise form of the total potential function differs between applications, but generally it is a product of the different factors.

5   Of these, physical suitability, accessibility and zoning are static data layers; they are constant over the course of a simulation. Physical suitability reflects aspects of soil, topography and climate that impact the aptness of a location for sustaining different land use classes. Accessibility reflects the degree to which every location is serviced by the infrastructure network and is modelled on the basis of the distance to links and nodes in the network. Zoning maps reflect the zoning status of

10   locations (allowed, not-allowed) at different moments in time.

The neighbourhood effect is the dynamic component of the land use potential. It expresses how the presence of land use classes in the surrounding of a location impact the aptness of that location to sustain different land use classes.

$$N_{k,a} = \sum_{b} w_{d(a,b),k,L(b)}$$ 6

15   where:

$N_{k,a}$ the neighborhood effect of cell a for land use class k.

$w_{d,k,l}$ the impact of land use class l has on the neighbourhood effect for land use k at distance d.

b this index iterates over all cells in the neighbourhood of cell a The neighborhood consists of all cells within a given distance of cell a.

20    $d(a,b)$ the distance between cells a and b, since the model operates on a regular raster, the number of within a given radius possible distances is limited.

$L(b)$ the land use class found at cell b

The neighborhood effect in the potential function qualifies the model as a cellular automata model and underlies the self-organizing behaviour of the model. The complexity of the model

25   arises from the fact that the land use classes at locations depends on the land use potential of these location which in turn depend on the land use.

The neighbourhood interactions are in first place intended to represent urban land use dynamics. The potential for agricultural and natural land use classes is typically dominated by the layers representing physical suitability and zoning.

30   The interaction between total area demands and land use potential is via the allocation mechanism. This mechanism iteratively finds the (yet unassigned) cell with the highest potential for any land use class (of which area demands are yet unmet), and assigns the cell to the particular land use class.

3.2 Study area

35   La Réunion belongs to the Mascareignes archipelago in the Indian Ocean with Mauritius and the Gambrani islands situated east of Madagascar (**Figure 3**). Politically, La Réunion belongs to the overseas French departments. It is one of the seven Ultraperipheral European Regions with the same status as regions on the European mainland.

The island covers 2512 km^2 but only 1000 km^2 are available for human activities, because of the

40   steep slopes that are found on the island. The highest peak is 3070 m (Piton des Neiges) and is

one of the world's most active volcanoes. Another restriction is formed by tropical forest which covers 1000 km^2 and has a highly protected status.

Figure 3. Geographical setting of La Réunion

Author : Gilles LAJOIE, 2007
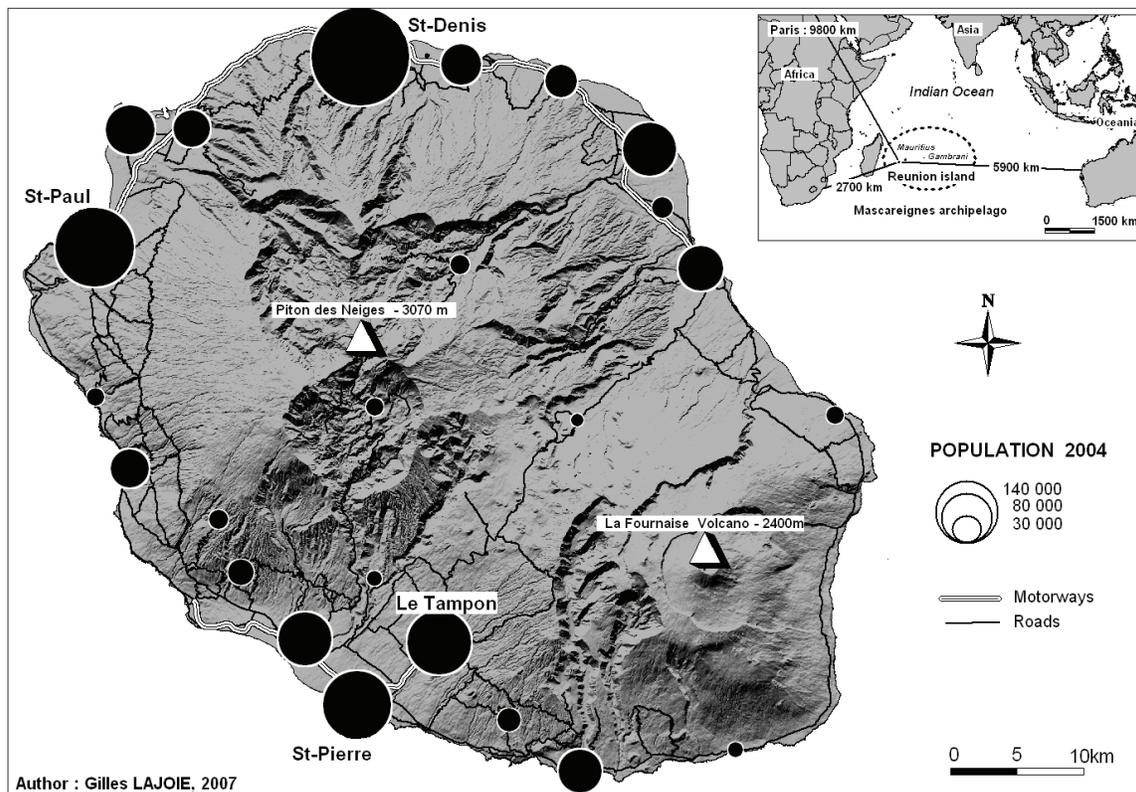
5

Over the last twenty years the island has undergone some profound changes. The population increased from 590 000 inhabitants in 1989 to 740 000 in 2002 and will reach 1 000 000 in 2028 according to demographic projections of the National Institute of Statistics report (Lardoux 2002). The economy has grown too and land has become scarce.

10 The land use model is calibrated over the period 1989 – 2002. In this period, urban development and the associated recession of forest and agriculture were the main processes of land use change report (Lagabrielle et al. 2005). The recession of agriculture is expressed by a decrease of 3000 ha of land of which 60% transformed into forest and 20% transformed into urban. This classic case of rapid recession of agriculture came unexpectedly because agriculture was the primary

15 economic activity on the island for several decades. Transitions took place both from agriculture to forest and vice versa.

While agriculture declined in the Northeast, it expanded in the Southwest. The explanation is simple; urban growth found ideal conditions in the Northeast (absence of slope, good accessibility to the capital where economic activity is concentrated, low land prices, etc.). At the same time,

20 irrigation made land in the Southwest suitable for agriculture thanks to a large construction project financed by the European Union that is planned to transform 7000 ha of savannah into sugar cane plantations over a period of thirty years (1983-2013).

3.3 Data

The calibration of the model requires two land use datasets. The best available data for this purpose are the remote sensing classifications from the TEMOS program and in particular the Data Base Isle project (CNES 2005). This project supplies scientists with pre-processed, multi-temporal satellite and multi-sensor data, as well as multi-thematic measures of land characteristics.

These source data are two SPOT images with a resolution of 20 meters for 1989 and 2002. The classification of this data is detailed in the TEMOS project report (Lagabrielle et al. 2005). Fifty land cover categories are distinguished, of which the land use model uses five particularly stable aggregate classes (mixed vegetation, forest, agriculture, urban, bare land). Additionally, to complete the cover of the island, steep slopes (>30%) and the littoral fringe have become separate classes.

Both images of 1989 and 2002 have been aggregated to grids of regular cells of 200 meters aside, every cell taking the dominant class among seven classes. Eventually, La Réunion is represented on a matrix of 286 rows and 323 columns. The choice of spatial resolution and categorical detail is not trivial. Categorical detail determines to a large extent the parameter space in the model, since there are a number of parameters for every pair of land use classes. As a consequence this choice impacts the degrees of freedom of the model and the risk of over calibration. For this reason the number of land use classes is kept as low as reasonably possible. The choice of model resolution is not just a matter of model precision; in fact the resolution of the model determines the dynamic behavior of the model (Benenson 2007). It does so in a technical sense because the software allows neighborhoods with a radius of maximally 8 cells, but also more fundamentally since the resolution determines the spatial unit at which the land use changes.

Suitability maps are based on a number of characteristics that are considered constant in time, in particular soil type, elevation and slope. Although some of this data has been collected after 1989, there is no reason to assume that the data is not a good representation of the conditions in 1989. The same holds for the road network that underlies the accessibility layer of the model. This data was collected for 2002, but the road network has not changed between 1989 and 2002 (new projects are planned for 2009-2019).
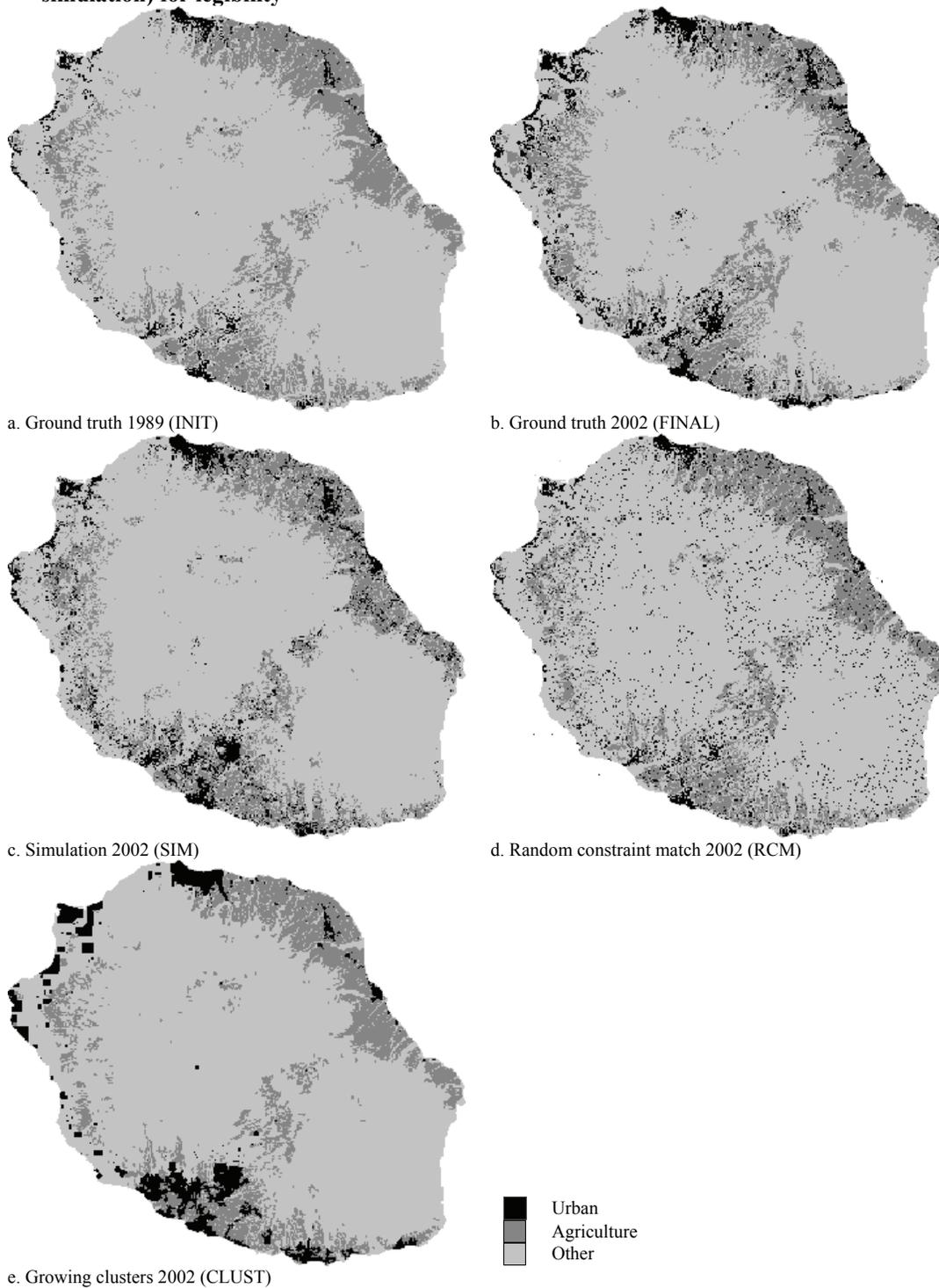
The land use maps of 1989 and 2002, as well as the simulated maps are presented in **Figure 4**.

4 Results

4.1 Model performance at multiple scales and criteria

The results for the performance indicator presence of agriculture are presented in **Table 1**. The results indicate that the changes that occurred over time (the column INIT compares the initial map of 1989 to the final map of 2002) vary with scale. At finer scales the changes are larger. The errors of the three models (SIM, RCM and CLUST) follow this same trend. In fact there is a near perfect linear fit between error and historical change for all three models (the correlation is 0.997 or higher for all). When the errors (**Table 1a**) are normalized to the historical change the results (**Table 1b**) fall within a range that is more naturally interpreted. The value 0 still stands for a perfect match, but the value 1 stands for the error that corresponds to the level of historical change, in other words the performance of the model of persistence.

Figure 4. Five land use maps, the original legend has been reduced to three main classes (after simulation) for legibility

a. Ground truth 1989 (INIT)

b. Ground truth 2002 (FINAL)

c. Simulation 2002 (SIM)

d. Random constraint match 2002 (RCM)

e. Growing clusters 2002 (CLUST)

Urban
Agriculture
Other

At coarser scales the model all models perform better than a model of persistence. The fact that the random constraint match and the growing clusters model display these 'good results' at coarse

scales (and even perfect results at the coarsest scale), indicates that this achievement must be attributed to the constraints imposed on the model.

Table 1. Goodness-of-fit results for the criterion presence of agriculture.

a. Mean squared error (MSE)

Scale	INIT	SIM	RCM	CLUST
512X512	1.1E-05	0.0E+00	0.0E+00	0.0E+00
256X256	8.4E-06	3.8E-06	3.5E-06	6.3E-06
128X128	3.8E-04	3.4E-05	2.7E-04	3.7E-04
64X64	2.0E-04	1.2E-04	1.9E-04	2.3E-04
32X32	4.6E-04	3.4E-04	3.8E-04	4.6E-04
16X16	7.2E-04	6.3E-04	5.9E-04	7.8E-04
8X8	1.2E-03	1.3E-03	1.1E-03	1.3E-03
4X4	2.3E-03	3.1E-03	2.2E-03	2.4E-03
2X2	5.5E-03	8.1E-03	5.7E-03	5.6E-03
1X1	1.7E-02	2.5E-02	1.8E-02	1.6E-02

b. MSE normalized to INIT (P)

Scale	SIM	RCM	CLUST
512X512	0	0	0
256X256	0.46	**0.41**	0.74
128X128	**0.09**	0.69	0.95
64X64	**0.58**	0.92	1.15
32X32	**0.74**	0.82	1
16X16	0.87	**0.82**	1.08
8X8	1.09	**0.95**	1.08
4X4	1.37	**0.98**	1.06
2X2	1.48	1.04	**1.03**
1X1	1.49	1.08	**0.95**

c. MSE cumulative over scales

Scale	INIT	SIM	RCM	CLUST
512X512	1.1E-05	0	0	0
256X256	2.0E-05	3.8E-06	3.5E-06	6.3E-06
128X128	4.0E-04	3.8E-05	2.7E-04	3.7E-04
64X64	6.1E-04	1.6E-04	4.6E-04	6.1E-04
32X32	1.1E-03	5.0E-04	8.3E-04	1.1E-03
16X16	1.8E-03	1.1E-03	1.4E-03	1.8E-03
8X8	3.0E-03	2.4E-03	2.6E-03	3.1E-03
4X4	5.3E-03	5.6E-03	4.8E-03	5.6E-03
2X2	1.1E-02	1.4E-02	1.0E-02	1.1E-02
1X1	2.8E-02	3.9E-02	2.9E-02	2.7E-02

d. MSE cumulative over scales normalized to INIT

Scale	SIM	RCM	CLUST
512X512	0	0	0
256X256	0.20	**0.18**	0.32
128X128	**0.09**	0.67	0.92
64X64	**0.26**	0.75	1.00
32X32	**0.47**	0.78	1.00
16X16	**0.63**	0.8	1.03
8X8	**0.81**	0.86	1.05
4X4	1.05	**0.91**	1.05
2X2	1.27	**0.97**	1.04
1X1	1.41	1.04	**0.99**

5   **INIT = Initial map (1989)**
        SIM = Constrained Cellular Automata simulation (2002
        RCM = Neutral Model: Random Constraint(2002)
        CLUST = Neutral Model: Growing Clusters (2002)
        The best fit per scale is highlighted

10   Of most interest is the performance of the CCA model (SIM) relative to the neutral models. **Table 1b** indicates that only at the mid-range scales of 32*32 to 128*128 the model outperforms both. **Table 1d** shows that in terms of cumulative results at each scale and all coarser scales the model outperforms the neutral models of landscape change only at scales of 8*8 to 128*128. **Figure 5** shows the graphs corresponding to **Table 1b** and **1d**. All other results are presented in
15   this visual form.

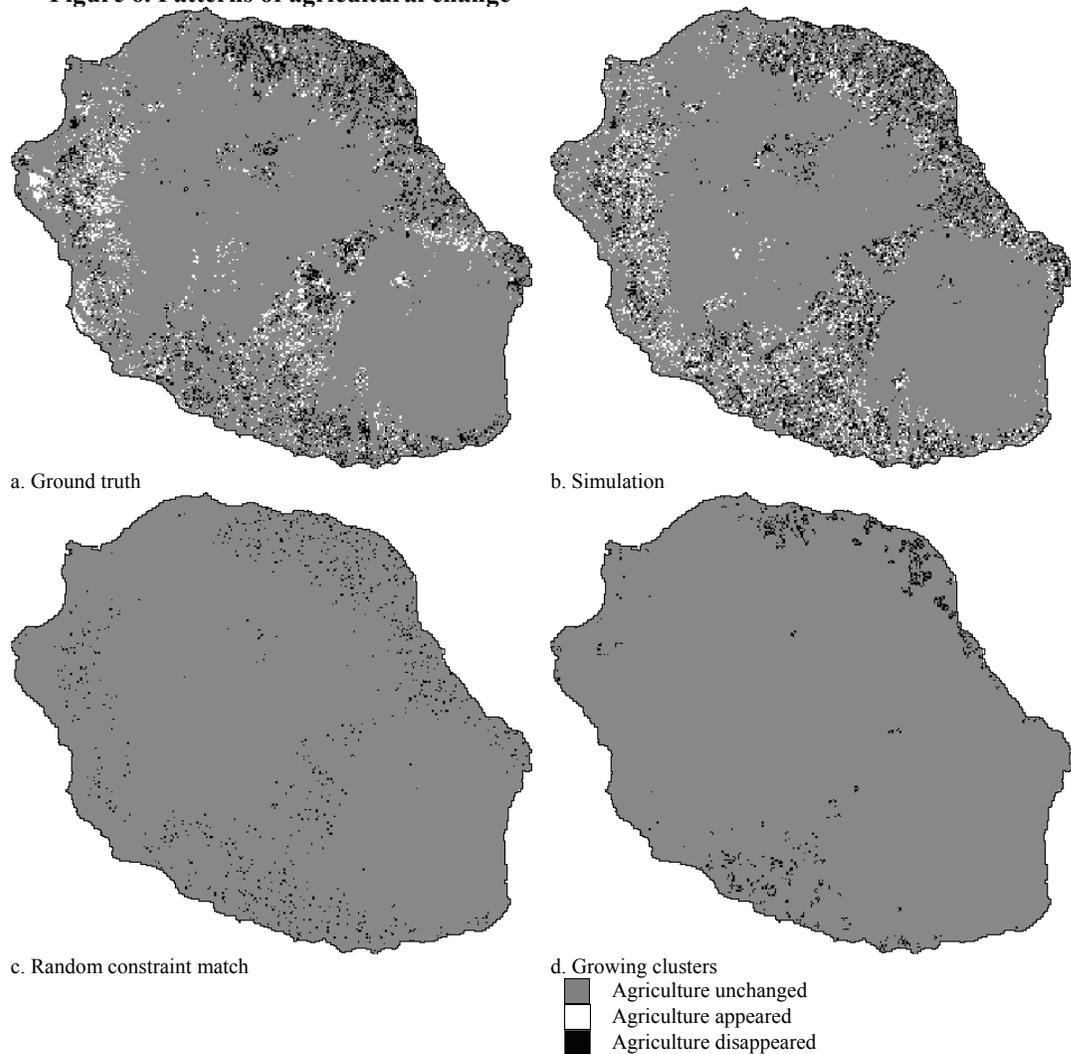**Figure 5. Multi-scale goodness-of-fit of presence of agriculture**
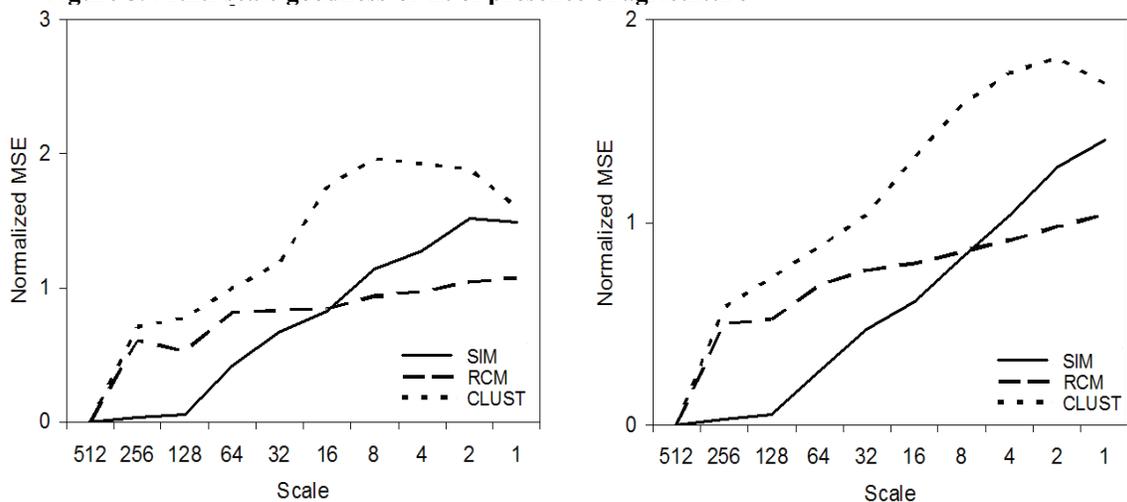
a. Each scale represents a wavelength

b. Cumulative: Each scale represents a wavelength and all larger wavelengths combined

A closer look at the change patterns of the agricultural land use class may illuminate these results. **Figure 6** illustrates that the changes patterns that occurred in reality consisted of appearing agriculture at some cells and disappearing at other. The same holds for the changes that occurred in the model, but not in the random constraint match model, since this model minimizes overall change.

Figure 6. Patterns of agricultural change

a. Ground truth

b. Simulation

c. Random constraint match

d. Growing clusters

Agriculture unchanged
Agriculture appeared
Agriculture disappeared

Results for presence criterion of urban are similar to those of agriculture (**Figures 7a** and **7b**); the CCA only outperforms the neutral models of landscape change at the mid range of scales (8*8 to 32*32). When spatial structure is considered the results are quite different. In terms of urban patch size (**Figure 7c**) the neutral models are outperformed at all scales and on the criterion of fractal dimension at all scales except 32*32 (**Figure 7d**).

Figure 7. Multiscale goodness-of-fit for the class urban

a. Presence

b. Presence cumulative

c. Patch size

d. Fractal dimension
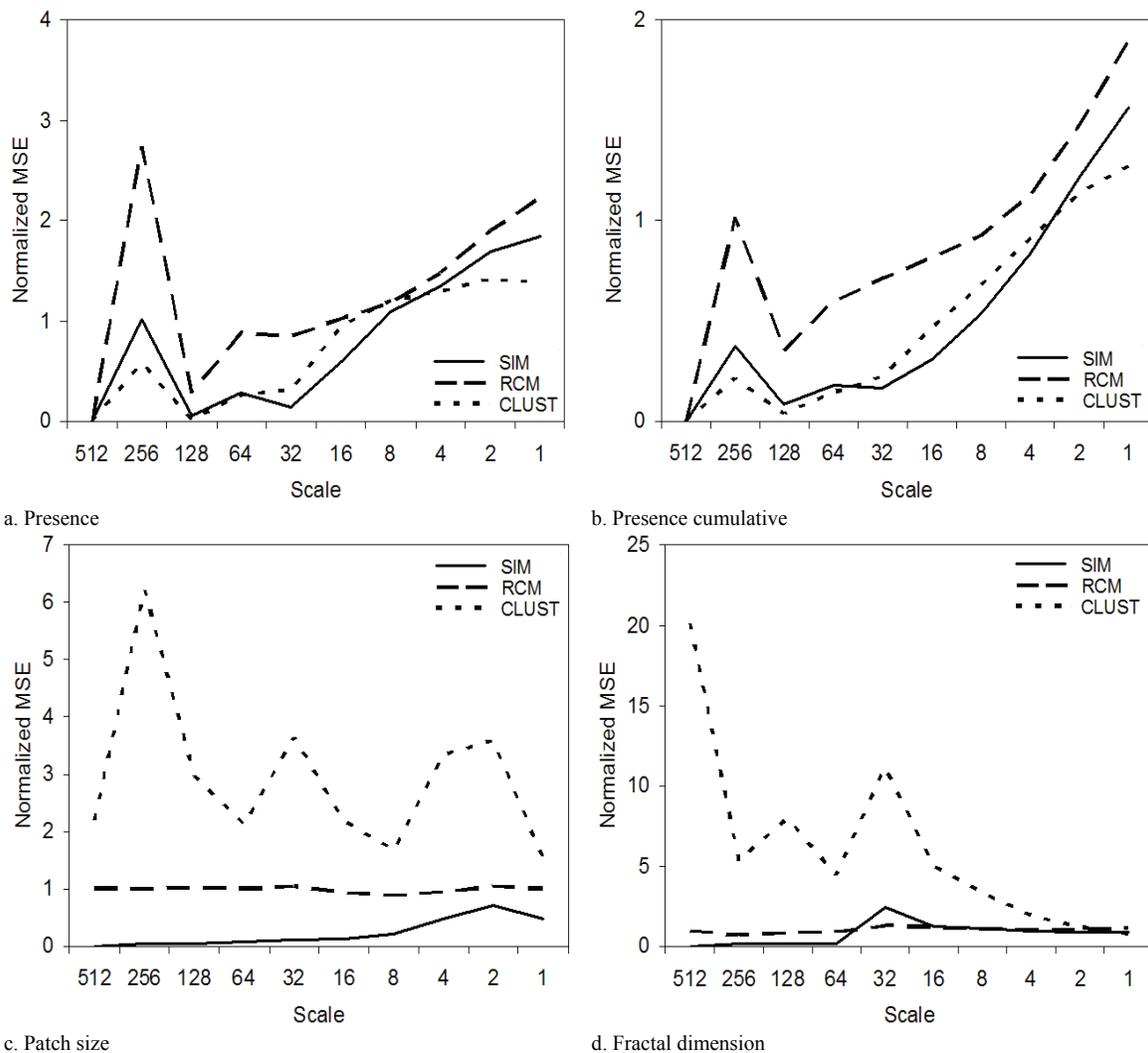
4.2 Sensitivity

In the course of the analysis, several choices have been made that are essentially arbitrary. In this section the impact of those decisions is investigated. The following three arbitrary decisions are considered.

5

1. The choice of the mother wavelet
2. The strategy for padding the image to 512 by 512 cells
3. The choice of evaluating only a single model run

For the analysis we used the Haar wavelet. Although we provided motivation for this choice of mother wavelet, the selection is more or less arbitrary.

10

The evaluation of similarity of agriculture presence is recalculated twice, on the basis of a Daubechies 4 and Daubechies 16 wavelets (**Figure 8**). Comparison to the earlier results (**Figure 5a**) makes clear that the choice of mother wavelet substantially impacts the analysis. For instance, at scale 16*16 the CCA model outperforms the reference models according to the Haar wavelet, but not according to the Daubechies wavelets. At the coarsest scales the CCA model outperforms

15

the neutral models of landscape change according to the Daubechies 4 and 16 wavelets, but not according to the Haar wavelet.

Figure 8. Goodness-of-fit of agriculture according to Daubechies wavelets
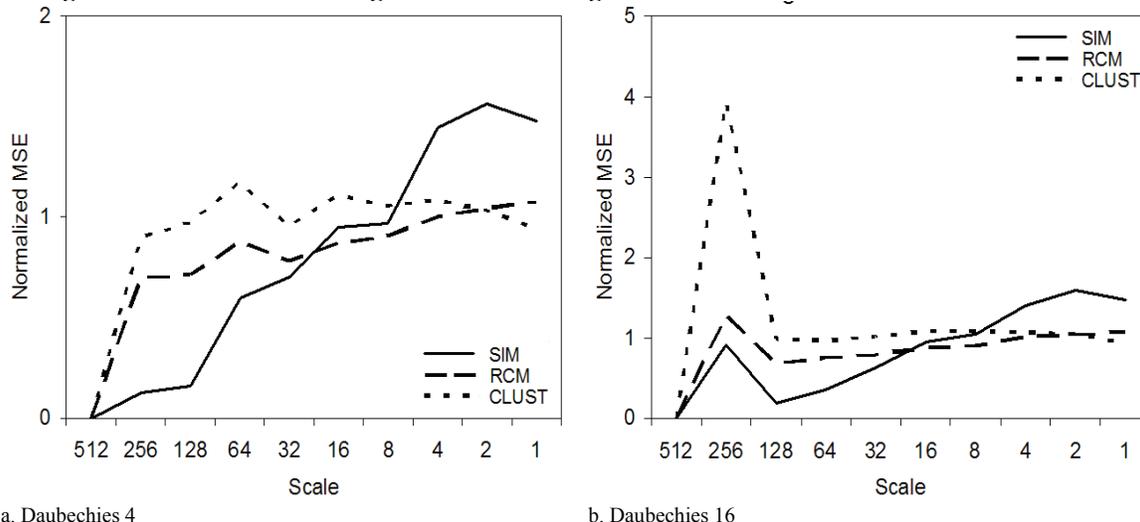
a. Daubechies 4

b. Daubechies 16

5

The discrete wavelet transform requires data in strings of 2^n elements. The original map has the dimension of 286 by 323 cells. In order to be processed in the wavelet analysis, all maps had to be padded to 512 by 512 cells. With padding is meant that at the sides of the map rows and columns are added to obtain the right dimension. Since La Réunion is an island, in effect padding includes a larger part of the ocean in the analysis. There are different options available considering how many rows are added in the North or the South and columns in the East or the West. The results presented so far have been based on the strategy of centering the original map on the padded map (i.e. the island lays in the middle of the ocean). One alternative is to place the original map on the upper-left corner of the padded map. This is the approach used to attain the results of (**Figure 9a**). In comparison to earlier results (**Figure 7a**) the errors have shifted between scales. Most strikingly, the peak in the error that was originally registered at the scale of 128*128 is now found at 256*256. At the finer scales the errors are approximately the same.

All three models have a stochastic component. Monte Carlo analysis of the distribution of goodness-of-fit is an option for further statistical rigor. The error bands in **Figure 9b** indicate the highest and lowest value in five runs of the models. It shows that the relative ordering of the models is most unclear at scale 8*8 and the coarse scales 8*8 to 256*256.

Of the three arbitrary decisions, the padding strategy influences the results most strongly.

Figure 9. Goodness of fit of urban presence under alternative options
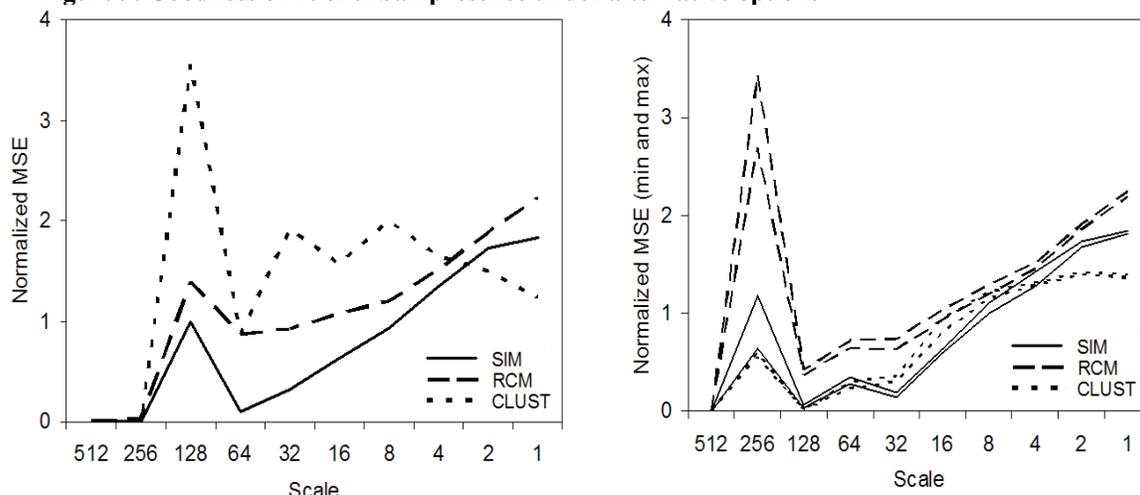
a. Alternative padding strategy   b. Monte Carlo

5. Discussion

The model evaluation period is the same period that is also used for the model calibration, namely 1989-2002. This is not the ideal case as preferably model performance is measured over a
5   separate validation period or otherwise independent data set. Good results for the calibration period do therefore not necessarily indicate good predictive quality. Temporal stationarity is not investigated and there is a risk of over-calibration. The consequence is that our conclusions only reflect how well the model can reproduce historical land use patterns and not how well it can be expected to predict future developments.

10   Results at the coarsest scale form a special case. At this scale, maps are represented by the mean over the map. Since the total presence of agriculture and urban area is exogenous to both the CCA model and the reference models, all perform identically (and perfectly) at this scale for the presence criteria. Other coarse scales must be considered with caution; the discrete nature of wavelets and relatively small number of coordinates at this level (refer back to **Figure 1**) make
15   results at this scale highly sensitive to the arbitrary positioning of the wavelet patterns. This is confirmed by the sensitivity analysis. The multi-scale assessment seems stable at scales 1*1 to 64*64 at coarser scales there is too little information and the results are overshadowed by the uncertainties posed by arbitrary decisions. Note that at 64*64 the whole island is captured by less than 30 cells.

20   The comparison of normalized results of the four criteria indicates that the model performs best in terms of urban patch size (**Figure 10** plots all criteria on the same scale). The strong contrast between performance of the model and the neutral models of landscape change indicates that the similarity in structure can be attributed to the processes of the CCA model. Barring the unreliable results at the coarsest scales, the results indicate that the Constrained Cellular Automata
25   outperforms neutral models at coarse scales for all criteria. In terms of structure, and particularly patch size, the CCA outperforms the neutral models at all scales.

The performance of the random constraint match model for fractal dimension is remarkable. How can this blunt neutral model form competition to the CCA when it comes to a criterion of patch shape? A visual assessment of the maps shows that the shape of the patches is not at all realistic;
30   in fact the model only adds speckles of size 1. The fractal dimension metric may have been ill-chosen for the current case. At the resolution of 200m many urban patches consist of only a few

cells and the metric value is then determined largely by the small number of possible configurations rather than the actual shape of the patch.

Figure 10. Multi criteria goodness-of-fit. Agriculture presence (AP), Urban presence (UP), Urban Urban patch size (US) and Urban fractal dimension (UF)
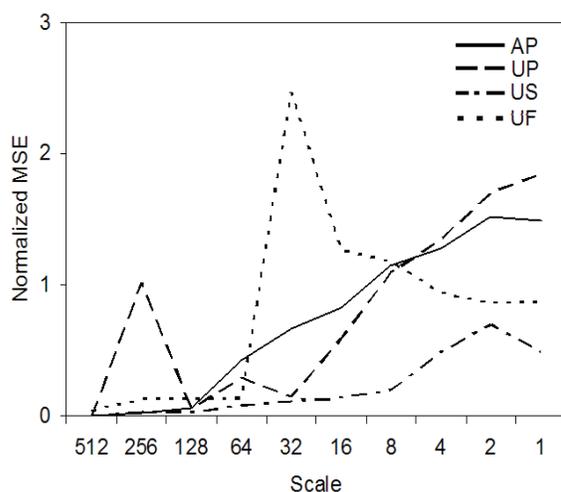
5

A remarkable error of urban presence is found at the coarse scale of 256*256 (128*128 in the analysis based on alternative padding). This is an indication that one or more large scale processes are insufficiently captured by the model. Several strong gradients over the island may be the explanation. The first gradient is climatic; the East receives 4000-5000 mm rain per year whereas the West receives less than 1000 mm per year. Another gradient is socio-economic; the model considers only a single urban class, but the characteristics of this class differ regionally in terms of population density and economic profile.

6. Conclusion

The performance of spatial models can be measured according to many different criteria. The problem arises that scores for these criteria are not mutually comparable and moreover that scores on identical criteria may not be comparable when applied at different scales or for different applications, since the context of the model determines the appropriate interpretation. The framework for model performance assessment proposed in this paper addresses both problems by first normalizing to the level of historical change and secondly expressing the model performance relative to that of neutral models of landscape change.

The development of rigorous frameworks for the evaluation of model performance is crucial to advance the field of land use modeling. As the field is developing, advances increasingly depend on systematic empirical analysis. An example of a frontier of new developments is the investigation of scale-dependency in CA land use models (Benenson 2007). Scale dependency is of course known for a long time, but only recently studies are appearing that systematically investigate the sensitivity to resolution and scale (e.g. Jantz & Goetz 2005; Kocabas & Dragicevic 2006; Menard & Marceau 2005, Samat 2006). All these studies present multi-scale analyses in similar fashion to this paper. The difference is that these studies analyze the results of models that are applied on multiple scales, whereas the model in this paper is only applied at a single resolution. These studies are confronted by the same problem that as the scale of analysis changes, the appropriate reference level for different performance criteria changes as well. Application of the methodological framework of this paper would enhance the interpretation given to the results in these studies.

Neutral models of landscape change are a new class of neutral landscape models. Their innovation is that instead of creating a landscape from scratch, they modify an initial landscape until all constraints are satisfied. This dynamic character is crucial if neutral landscape models are to be used in relation to the process of land use change. The potential for application of neutral

5 models of landscape change is not restricted to land use modelling purposes. Any landscape analysis with temporal aspects can benefit from neutral models of landscape change. The models can be employed to test hypotheses about patterns of change in the landscape rather than the landscape patterns themselves.

References

10 Barredo, J.I., Demicheli, L., 2003. Urban sustainability in developing countries' megacities: modelling and predicting future urban growth in Lagos. Cities 20, 297-310.

Benenson, I., 2007. Warning! The scale of land-use CA is changing! Computers Environment and Urban Systems 31, 107-113.

Briggs, W.M., Levine, R.A., 1997. Wavelets and field forecast verification. Monthly Weather
15   Review 125, 1329-1341.

CNES 2006. Portail KALIDEOS. Data portal available online: http://kalideos.cnes.fr (17 May 2007)

Costanza, R., 1989. Model goodness of fit: A multiple resolution procedure. Ecological Modelling 47, 199-215.

20 Ebert, E.E., McBride, J.L., 2000. Verification of precipitation in weather systems: determination of systematic errors. Journal of Hydrology 239, 179-202.

Engelen, G., White, R., de Nijs, T., 2003. Environment Explorer: Spatial support system for the integrated assessment of socio-economic and environmental policies in the Netherlands. Integrated assessment 4, 97-105.

25 Fewster, R.M., Buckland, S.T., 2001. Similarity indices for spatial ecological data. Biometrics 57, 495-501.

Frankhauser, P., 1994. La fractalité des structures urbaines. Anthropos : Diffusion, Economica, Paris.

Gardner, R.H., Urban, D.L., 2007. Neutral models for testing landscape hypotheses. Landscape
30   Ecology 22, 15-29.

Hagen-Zanker, A., 2006. Map comparison methods that simultaneously address overlap and structure. Journal of Geographical Systems 8, 165-185.

Hagen-Zanker, A., Straatman, B., Uljee, I., 2005. Further developments of a fuzzy set map comparison approach. International Journal of Geographical Information Science 19, 769-
35   785.

Jantz, C.A., Goetz, S.J., 2005. Analysis of scale dependencies in an urban land-use-change model. International Journal of Geographical Information Science 19, 217-241.

Kocabas, V., Dragicevic, S., 2006. Assessing cellular automata model behaviour using a sensitivity analysis approach. Computers, Environment and Urban Systems 30, 921-953.

40 Kuhnert, M., Voinov, A., Seppelt, R., 2005. Comparing raster map comparison algorithms for spatial modeling and analysis. Photogrammetric Engineering and Remote Sensing 71, 975-984.

Lagabrielle E., Metzger P., Martignac C., Durieux L., Grellet G. 2005. Guide critique d'utilisation des informations produites dans le cadre du projet TEMOS à La Réunion.
45   CIRAD. Report available online: http://www.cirad.fr/reunion/actualites/archives_2006/-rapport_temos (17 May 2007)

Lajoie, G., 2007. La simulation de l'étalement urbain à La Réunion: apport de l'automate cellulaire Metronamica® pour la prospective territoriale. In: Cybergeo, vol. Article 405.

Lardoux, J.-M., 2002. La population en 2030 : dans le cadre national. Revue économie de La Réunion 112.

McGarigal, K., Cushman, S.A., Neel, M.C., Ene, R., 2002. FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: www.umass.edu/landeco/research/fragstats/fragstats.html. In:

Menard, A., Marceau, D.J., 2005. Exploration of spatial scale sensitivity in geographic cellular automata. Environment and Planning B-Planning & Design 32, 693-714.

Pijanowski, B.C., Brown, D.G., Shellito, B.A., Manik, G.A., 2002. Using neural networks and GIS to forecast land use changes: a Land Transformation Model. Computers, Environment and Urban Systems 26, 553-575.

Pontius Jr., R.G., Boersma, W., Castella, J.-C., Clarke, K., de Nijs, T., Dietzel, C., Zengqiang, D., Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C.D., McConnell, W., Pijanowski, B., Pithadia, S., Sood, A.M., Sweeney, S., Trung, T.N., Veldkamp, A.T., Verburg, P.H., in press. Comparing the input, output, and validation maps for several models of land change. Annals of Regional Science.

Pontius Jr., R.G., Huffaker, D., Denman, K., 2004. Useful techniques of validation for spatially explicit land-change models. Ecological Modelling 179, 445-461.

Pontius Jr., R.G., Walker, R., Yao-Kumah, R., Arima, E., Aldrich, S., Caldas, M., Vergara, D., 2007. Accuracy assessment for a simulation model of amazonian deforestation. Annals of the Association of American Geographers 97, 677-695.

Power, C., Simms, A., White, R., 2001. Hierarchical fuzzy pattern matching for the regional comparison of land use maps. International Journal of Geographical Information Science 15, 77-100.

Remmel, T.K., Csillag, F., 2006. Mutual information spectra for comparing categorical maps. International Journal of Remote Sensing 27, 1425-1452.

Rykiel, E.J., 1996. Testing ecological models: The meaning of validation. Ecological Modelling 90, 229-244.

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. Computers, Environment and Urban Systems 30, 905-920.

Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46, 234-240.

Turner, M.G., 1987. Spatial simulation of landscape changes in Georgia: A comparison of 3 transition models. Landscape Ecology 1, 29.

Turner, M.G., Costanza, R., Sklar, F.H., 1989. Methods to evaluate the performance of spatial simulation-models. Ecological Modelling 48, 1-18.

Turner, M.G., Gardner, R.H., O'Neill, R.V., 2001. Landscape ecology in theory and practice : pattern and process. Springer, New York.

White, R., Engelen, G., 1993. Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land use patterns. Environment and Planning A 25, 1175-1199.

White, R., Engelen, G., Uljee, I., 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. Environment and Planning B: Planning and Design 24, 323-343.

With, K.A., King, A.W., 1997. The use and misuse of neutral landscape models in ecology. Oikos 79, 219-229.

Zepeda-Arce, J., Foufoula-Georgiou, E., Droegemeier, K.K., 2000. Space-time rainfall organization and its role in validating quantitative precipitation forecasts. Journal of Geophysical Research-Atmospheres 105, 10129-10146.

Appendix: Pseudocode

```
function RandomConstraintMatch( initialMap, targetMap, resultMap )

        resultMap = initialMap

        for cat = 1 to lastCat 'cat is short for category
                adjustment( cat ) = Count( initialmap, cat ) - Count( targetmap, cat )
                catlist( cat ) = MakeListOfAllCells( initialmap, cat )

                if( adjustment(cat) > 0 ) then
                        for sample = 1 to adjustment( cat )
                                cell = PickRandom( catlist( cat ) )
                                AddToList( cellList, cell )
                                RemoveFromList( catList( cat ), cell )
                        next sample
                endif
        next cat

        for cat = 1 to lastCat
                if( adjustment( cat ) < 0 ) then
                        for sample = 1 to -adjustment( cat )
                                cell = PickRandom( cellist )
                                resultMap( cell ) = cat
                                RemoveFromList( cellList, cell )
                        next sample
                endif
        next cat

end function

function growingclusters( initialmap, targetmap, resultmap )

        adjustment( cat ) = Count( initialmap, cat ) - Count( targetmap, cat )
        resultmap = initialmap

        while( not all adjustment(cat) = 0 ) do
                for cell = 1 to lastCell
                        for nb = 1 to 4 'nb is short for neighbour
                                cat = resultMap(cell)
                                nbCat = resultMap(cell, nb)
                                catRight = ( adjustment( cat ) > 0 )
                                nbRight = ( adjustment( nbCat ) < 0 )
                                if ( catRight and nbRight ) then
                                        edge = { cell, cat, nbCat }
                                        AddToList( edgeList, edge )
                                endif
                        next nb
                next cell

                if edgeList = empty
                        for cat = 1 to lastCat
                                if( adjustment( cat ) < 0) then
                                        repeat
                                                cell = PickRandom( initialMap )
                                                randomCat = resultMap(cell)
                                        until( adjustment( randomcat ) > 0 )

                                        resultmap( cell ) = cat
                                        adjustment( randomCat )--` decrement
                                        adjustment( cat )++ ` increment
                                endif
                        next cat
                endif

                while ( not edgeList = empty ) do
                        edge = PickRandom( edgeList )
                        RemoveFromList( edgeList, edge)

                        stillOver = ( adjustment( edge.cat ) > 0 )
                        stillUnder = ( adjustment(edge.nbhcat) < 0 )
```

```
                        stillNeed = ( stillOver and stillUnder )
                        stillSame = ( resultmap(edge.cell) == edge.cat )

                        if stillNeed and stillSame then
                                adjustment( edge.cat )--
                                adjustment( edge.nbhCat )++
                                resultmap( edge.cell ) = edge.nbhCat
                        endif
                loop
        loop
    end function
```