International Journal of Geographical Information Science
Vol. 20, No. 9, October 2006, 1067–1085

Research Article

Agent-based modelling of shifting cultivation field patterns, Vietnam

M. R. JEPSEN*, S. LEISZ, K. RASMUSSEN, J. JAKOBSEN, L. MØLLER-JENSEN and L. CHRISTIANSEN

Institute of Geography, University of Copenhagen, Øster Voldgade 10, DK-1350 Copenhagen K, Denmark

(*Received 30 October 2005; in final form 26 April 2006*)

Shifting cultivation in the Nghe An Province of Vietnam's Northern Mountain Region produces a characteristic land-cover pattern of small and larger fields. The pattern is the result of farmers cultivating either individually or in spatially clustered groups. Using spatially explicit agent-based modelling, and relying on empirical data from fieldwork and observations for parameterization of variables, the level of clustering in agricultural fields observed around a study village is reproduced. Agents in the model act to maximize labour productivity, which is based on potential yield and labour costs associated with fencing of fields, and are faced with physical constraints. The simulation results are compared with land-cover data obtained from remote sensing. Comparisons are made on patterns as detected visually and using the mean nearest-neighbour ratio. Baseline simulation outputs show high degrees of spatial clustering and similarity to the land-cover data, but also a need for further calibration of model variables and controls.

Keywords: Land use; Spatial patterns; Shifting cultivation; Northern Mountain region; Vietnam

1. Introduction

Human transformation of the surface of the Earth receives much attention. As a process in itself, land-cover change has huge impacts on ecosystems, biodiversity, and the global climate. It is also a spatial footprint of societal processes. Research on land use and land-cover change (LUCC) has been a programme element of the International Geosphere–Biosphere Programme and the International Human Dimensions Programme on Global Environmental Change (Gutman *et al.* 2004), and is now continued within the Global Land Project (GLP) (Rindfuss *et al.* 2004, Verburg and Veldkamp 2005), with the aim to understand the coupled human–environmental system, and focusing on human decision-making and actions.

Shifting cultivation is a farming system commonly found in the Tropics (see, for instance, Halenda 1989, Angelsen 1995, Read and Lawrence 2003, Lawrence 2005). The basic rationale of shifting cultivation is what may be termed 'temporal concentration' of nutrients (Christiansen 1992, Rasmussen and Møller-Jensen 1999), achieved by building up a nutrient pool in the fallow vegetation and releasing them

*Corresponding author. Email: mrj@geogr.ku.dk

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810600830848

at the start of the cultivation period. Alternatively, or as a supplement, the use of fallow may be explained as a means of suppressing weeds (see Mertz 2002, for an overview of work related to these subjects).

The typical shifting cultivation landscape is a mosaic of cultivated plots and secondary regrowth, i.e. the fallow plots, and understanding the human processes configuring the landscape is important when addressing global environmental change.

A range of quantitative techniques exists for describing the spatial patterns of landscapes. See, for example, Turner *et al.* (1989), Turner (1990), and Antrop and Van Eetvelde (2000), for examples on global landscape indices (diversity, dominance, and contagion) and patch metrics (patch size, perimeter, fractal dimension, edges, and adjacency probability), or Leung *et al.* (2003) and Amarasinghe *et al.* (2005) for examples of cluster detection (local Moran's I, local Geary's C, and Getis G). These methods have also been applied to LUCC studies by Pan *et al.* (2004) and Parker and Meretsky (2004), for example.

Agent-based spatially explicit modelling offers a potential insight to the processes configuring the land cover. Referred to as multi-agent systems of land use and cover change (MASLUCC), the purpose of these models is to simulate the behaviour of actors (e.g. farmers) by setting up rules describing their decision-making (Parker *et al.* 2003), and thereby contribute to an understanding of the processes creating the landscape. This understanding is not only interesting from a pure academic point of view but also an insight in how people act, and the resulting impact on the landscape can also be useful in a policy context.

This paper presents a case study from a shifting cultivation system in the uplands of the Nghe An province, North-Central Vietnam. Analysis of satellite images covering the study area shows a characteristic patchiness of regrowth/fallow areas and cultivated fields ranging in size from very small to large fields. The large fields result from a preference by local farmers to cultivate adjacent fields. Based on classical shifting cultivation theory (Boserup 1965), fieldwork data, and remote sensing data, we constructed a spatially explicit agent-based model of shifting cultivation land use in Ban Que (Que village) and compared the simulated landscape against remotely sensed images using basic spatial statistics to answer the following research questions:

1. Are local farmers' behaviours in accordance with classical shifting cultivation theory?
2. Will a spatially explicit agent-based model of farmer's decisions regarding field location result in a spatial pattern of smaller and larger fields?
3. Will the modelled land-cover pattern match the land-cover pattern observed by satellite?

2. Theory and methods

In constructing the model presented in this paper, we rely on two basic scientific principles: Inductive modelling and deductive modelling. Inductive models are based on real-world observations, while general theories or knowledge drive deductive modelling (Brown *et al.* 2004). In this section, we will focus on the deductive part of the model and return to the empirically fitted model aspects afterwards.

2.1 *Shifting cultivation*

What distinguishes shifting cultivation from permanent agriculture is the use of fallow which can be seen as a labour-minimizing cultivation strategy in two ways:

1. By allowing nutrients to build up in the fallow biomass for subsequent release to the soil, shifting cultivators let time do the hard work of restoring soil fertility after cultivation.
2. To the extent that fallow vegetation also has an out-shading impact on weeds, time also substitutes for labour in this respect.

If space is available, and production is not market-driven, shifting cultivation can be regarded as a rational production strategy with the main rationale of farmers being to optimize returns to the sparsest production input, namely labour (Boserup 1965), while ensuring that enough food is produced to ensure food sufficiency (Kates *et al.* 1993). This rationale is expressed through the behaviour of farmers, deciding when and where to cultivate, and thus configuring the landscape. Some decisions are based on perfect information, and farmers are capable of acting in order to make perfectly rational choices à la 'economic man' (Rasmussen and Møller-Jensen 1999), while other decisions are made within a bounded rationality (Manson 2000, Benenson and Torrens 2004). In the highly simplified model presented here, decisions behind farmer behaviour are understood in light of empirical observations in Que village and the theory of labour productivity maximization, while ensuring that the production meets the demand.

2.2 *Agent-based modelling*

Agent-based modelling (ABM), multi-agent systems, and multi-agent simulation have become the terms for a range of applications involving hard- or software agents (Bousquet and Le Page 2004). In a LUCC context, ABM is being used to simulate farming or natural-resource management decisions, often by letting a set of agents, representing local decision-makers, act on a simulated environment represented as a georeferenced grid map of cells (often derived from remote sensing or GIS), in which natural processes occur (Deadman *et al.* 2004, Klügl *et al.* 2005). Previous applications of ABM include: a multi-agent simulation of land-tenure arrangements and land-use allocation in rural Vietnam (Castella *et al.* 2005), a simulation of agricultural innovation diffusion and hydrological and economic processes in Chile in a spatially explicit model (Berger 2001), and use of spatial pattern metrics to measure the outcome of a spatially explicit model working on an abstract rural–urban landscape (Parker and Meretsky 2004). Manson (2005) combined a spatial ABM and genetic programming model of actors, institutions, and environment in Mexico, and Deadman *et al.* (2004) constructed a spatially explicit ABM of smallholder farming in the Amazon rainforest. Excellent reviews of ABM applications to land-use systems can be found in Agarwal *et al.* (2002), Parker *et al.* (2003), and Brown *et al.* (2004).

The model presented in this paper is a discrete-choice model (Parker and Meretsky 2004), where agents choose among a discrete set of land-use choices. The model outcome is a map of land use that forms the basis for comparison of model outputs with observed land-use patterns.

2.3 *Validation methodology and spatial metrics*

Comparison of the agent-induced land-cover configurations with those from satellite images of the study area provides a limited means of validating the model outputs. A suite of spatial similarity techniques exists for comparing model outputs with observed patterns (Brown *et al.* 2005), but as the shifting cultivation landscape is changing every cropping season, a pixel-by-pixel or section-by-section comparison of the simulated output to the validation data is not fruitful.

As the overall goal of the model is to reproduce the observed landscape as constituted by the field patterns, we use simple spatial metrics to describe the clustering tendency. The mean nearest-neighbour ratio (MNNR) is chosen for two reasons: (1) it is calculated using simple mathematics and (2) it is well suited for description of clustering. The spatially explicit mean nearest-neighbour tool, available in ESRI's ArcGIS 9.0 spatial statistics toolbox, calculates the ratio of the observed mean Euclidean distance from the centre of a cultivated cell to the centre of the nearest other cultivated cell against a hypothetical random distribution, calculated as

$$\mathrm{NNR} = \frac{\sum_{i=1}^{n} C_i}{n} \bigg/ \frac{0.5}{\sqrt{n/A}} \tag{1}$$

where $\sum_i C_i$ is the sum of the distances to each feature's nearest-neighbour, n is the number of features, and A is the areal extent of the study area. The resulting value indicates a clustered pattern if the ratio is less than 1; ratios greater than 1 indicate a dispersed pattern. Along with the ratio, a Z-score is also reported. Statistically significant clustering is indicated by low negative Z-scores, while dispersed patterns have high positive Z-scores. Z-scores around 0 indicate a random distribution of data.

In this context, landscapes dominated by large fields will appear as clustered because many small cultivated fields are adjacent. The opposite is the case for a very fragmented landscape composed of cultivated cells spread throughout the landscape. As the MNNR is highly sensitive to spatial extent, it is important to keep this constant when comparing simulation results and satellite-derived data. In this analysis, the spatial statistics will be calculated for the area representing the village territory. To compare land-cover data and simulation results to a truly random distribution of plots, an image was produced in which 214 plots (the average number of plots observed in the remote sensing validation data) were distributed randomly within the village territory, and the MNNR for this image was computed.

3. Data

3.1 *Field data*

Within the 'University Support to Environmental Planning and Management in Lao PDR, Vietnam and Cambodia (USEPAM)' project (www.usepam.org), research has been conducted on shifting cultivation production systems in villages across the Nghe An province of Vietnam. A combination of rapid rural appraisal (RRA) and participatory rural appraisal (PRA) techniques were used to obtain information on cultivation strategies, natural-resource use, and labour inputs by methods such as semi-structured interviews, targeted single and group interviews, and cropping calendars. Interviews were done with groups gathered by the village leader in 2003 and 2005, and a structured questionnaire was distributed to the heads of 30 randomly sampled households in 2003. To complement remote-sensing data on land use and land-cover distribution, transects were walked with local farmers. This paper presents results from Que village.

The population of Ban Que is 409, distributed in 69 households, and we assumed that it was constant across the entire period of the study. Of the 69 households, 65 practise shifting cultivation. The land cover within the village territory of 1574 ha consists of secondary regrowth (fallow) of grass, bamboo and deciduous broad-leafed trees, cultivated shifting cultivation fields, and irrigated paddy fields. Farmers did not reveal any preferences for slashing, burning, and cultivating bamboo areas as opposed to forested areas. In this simulation, the two land-cover classes will therefore not be treated separately. Households are located centrally in the territory of Que village along a large stream surrounded by steep slopes and ridges with elevations ranging from 175 to 500 m (Jakobsen 2005). The flat areas along the river are used for paddy fields, and areas on steep slopes above 35° are not considered for cultivation, mainly due to a lack of soil moisture (figure 1).

The fields cultivated by households averaged 5080 m^2 in size, and the mean yields after 4 and 5 years of fallow are 950 kg ha^{-1} and 1200 kg ha^{-1}, respectively. The maximum yield observed is 1875 kg ha^{-1}, and one crop is raised annually. We fit a logistic curve to the relationship between yield and fallow period as reported by villagers during the interviews (figure 2, equation (2)), and assume that cultivating without prior fallowing will yield only 300 kg ha^{-1} and that the maximum yield of 1875 kg ha^{-1} is obtained after minimum 12 years of fallow. The relationship is as follows

$$y = \frac{a}{1 + b \cdot \exp(-cx)} \qquad (2)$$

where y=yield; x=fallow age; a=1,866; b=8.07; and c=0.52. R^2 is 0.99.

Figure 1. Ban Que territory. The Village territory is shown in light grey, and cells with a slope less than 6° and higher than 35° are shown in black and white, respectively.

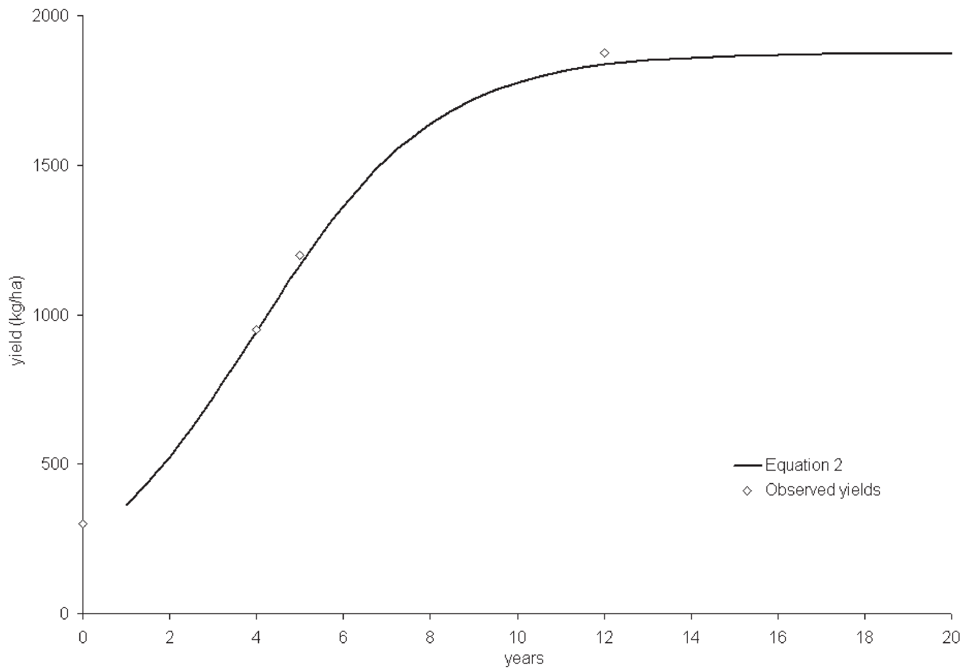

Figure 2. Exponential fallow–yield relationship. The line shows a logistic regression fitted to empirical data (Equation 2).

As a protection against free-ranging livestock and wild animals, fences are commonly constructed along the perimeter of cultivated fields. Fences are made of bamboo, either cut from the regrowth or taken from old fences. Due to the hot and humid climate, most fence materials decompose during a year, and new fences have to be built at the beginning of each crop season. The effort involved in fencing fields is thus considerable, and the construction of cropping calendars with the villagers showed an average of 134 days spend per field, of which 22 days were used for fencing. Based on the rough assumption that the 22 days spent on fencing represents fencing two sides of a field, we estimate the average labour requirements for cultivating a field without fences as $134-22=112$, and the labour requirements for cultivating and fencing a field on all four sides as $134+22=156$ days. Values for cultivating and fencing fields on one, two, or three sides are linearly interpolated between these values. A formalization of this is that labour requirement $=156-(11\times$ the number of uncultivated neighbouring fields); see table 1.

The effort involved in fencing fields may be reduced by cultivating adjacent plots and thereby sharing the workload associated with fencing. This implies that contiguous field areas of up to 20 ha are sometimes found. When deciding where to locate a plot, the farmer thus may have to trade-off between (1) choosing a plot with high productivity and (2) choosing a plot requiring low labour inputs.

3.2 *Satellite data and pattern metrics*

We derived the slope from a digital elevation model (DEM) calculated from a 15×15 m resolution ASTER along-track optical stereo data from November 2001 (Tottrup in press). Field patterns from Que village territory were obtained from

Table 1. Labour productivity as a function of fallow age and labour required for fencing.[a]

Fenced perimeter	Labour days	Fallow age 1	2	3	4	5	6	7	8	9	10
0	156	2.38	3.44	4.75	6.20	7.65	8.95	9.99	10.77	11.32	11.69
1	145	2.55	3.68	5.08	6.64	8.19	9.58	10.70	11.53	**12.12**	12.51
2	134	2.74	3.96	5.46	7.14	8.81	10.30	11.51	**12.40**	**13.03**	13.46
3	123	2.97	4.29	5.91	7.72	9.53	11.15	**12.45**	**13.42**	**14.10**	14.56
4	112	3.23	4.67	6.44	8.41	10.38	**12.14**	**13.56**	**14.62**	**15.36**	15.86

[a]Labour productivity is calculated as the ratio of yield (calculated using equation (2)) to assumed labour costs. Values shown in bold are the combinations of fallow age/fencing that make a cell more attractive to the agent than a 10-year-old fallow without fencing.

analysis of Landsat TM satellite images from 20 November 1991, 27 December 1993 and 7 November 1998, with a resolution of 30×30 m. These three images were geometrically co-rectified to each other and subsequently transformed to the UTM coordinate system and resampled to 15×15 m resolution to match the DEM. The fields were identified by visual interpretation and digitized on screen. The resulting field polygons were post-processed by overlaying the digital elevation model described above, in order to discriminate between shifting cultivation fields and other field types, prevalent in valley bottoms. The output from the analysis was three vector files describing field boundaries at each date. To ensure consistency in the modelled data and the validation data, the geographical scale (measured by spatial resolution and extent) of the two data sets has been matched. The validation polygons have been converted to raster format with the same extent as the model input and output, and resampled to a spatial resolution of 70 m (4900 m^2). The conversion and resampling procedures introduce minor errors in the spatial accuracy of field locations and the spatial extent of each field but also ensure uniformity of the structure of model data and validation data. Finally, the validation data were converted to vector domain to allow for the MNNR to be calculated.

Comparison of the three images (figure 3) clearly reveals that shifting sites for cultivation is an important property of the agricultural system, as is a decision about

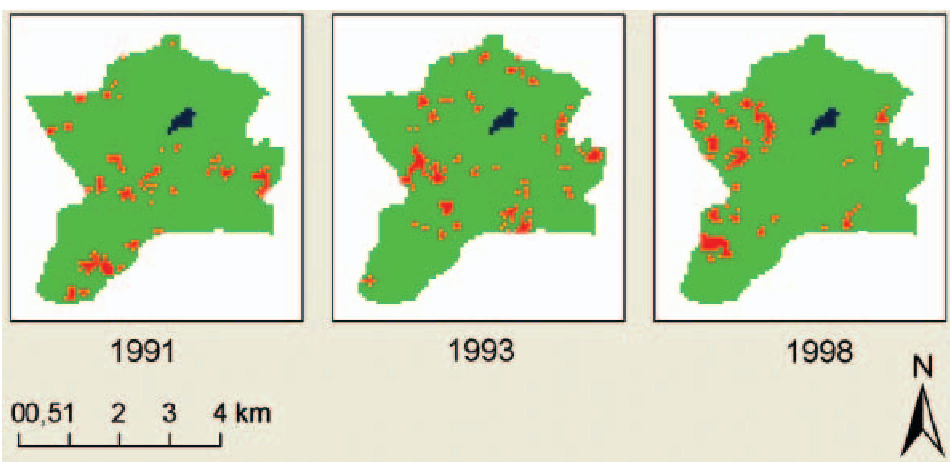

Figure 3. Satellite-derived land-cover data. Legend as in figure 5.

adjacent farming. What appears on the figure (and on the remote sensing data) as clearcuts, a land-cover phenomenon, is caused by land use, namely the practice of farming adjacently as described in section 3.1. Each clear cut is composed of at least one unit of 4900 m^2, the smallest unit detectable given the geographical resolution of the project.

The MNNR for all three validation years indicates a high degree of clustering, which is supported by very significant Z-score values (table 2). For the years 1993 and 1998, the nearest-neighbour observed mean distance (NNOMD) is close to 70. As this is the lag of the input map (the distance between centres of two adjacent cells; Duncan *et al.* 2002), this figure alone reveals strong clustering. As the area of analysis is always kept constant, the Expected Mean Distance reflects the number of cultivated fields in the data.

4. Modelling shifting cultivation field patterns

The model presented here is driven by a combination of empirical data and theoretical assumptions regarding shifting cultivation strategies. The goal was to empirically test the ability of agent behaviour based on cultivation theory to reproduce observed field patterns. The main decision the agents in the model make is the choice of field location. This decision is guided by expected return to labour investment which, in turn, is determined by the state of the potential field in terms of the age of the fallow vegetation, its slope, and its location relative to cultivated fields. The dependence on the location relative to other cultivated fields stems from the importance of protecting the fields from free-roaming livestock and wild animals by building fences. Sharing the outer perimeter of fields with neighbouring fields implies that the work input is reduced substantially (table 2).

The model's time step is a year, and agents find one or more cells to cultivate each time step (depending on their household size; see description in section 4.2). At the end of the time step, the agents return to Ban Que, and the cultivated areas return to regrowth.

4.1 *Environment*

The model is spatially distributed covering the area belonging to Ban Que. The total area of 1574 ha is represented in a discrete two-dimensional grid map with cell sizes of 70×70 m (0.49 ha), and the spatial resolution approximates the observed mean plot size (0.51 ha). A basic location map, imported from GIS, discriminates between the village, the village territory, and areas not belonging to Ban Que.

Table 2. Number of fields (4900 m^2), mean nearest-neighbour ratio, and Z-score for the satellite data and a random distribution of 214 fields within the village territory.

	1991	1993	1998	Random
Number of cultivated cells	190	231	222	214
Nearest-neighbour observed mean distance (NNOMD)	80.87	71.537	71.89	137.20
Expected mean distance (EMD)	134.43	121.92	124.37	126.67
Mean nearest-neighbour ratio (MNNR)	0.60	0.59	0.58	1.08
Z-score	−10.51	−12.06	−12.03	2.33

Each cell holds a discrete state representing the simulated land cover and variables describing the age of the land cover and the slope of the cell. To avoid biasing the results by using an already clustered real-world land-cover configuration as input, all the village territory is set to be fallow (regrowth) at the onset of the simulation, and *age* of regrowth cells is distributed randomly.

A very central variable is the labour productivity variable. The labour productivity variable equals the potential productivity of a cell calculated using equation (2), divided by the labour required to cultivate the cell which is a function of the number of adjacent cultivated cells, counted in the variable *perimeter*. Table 3 shows the values and transition rules for states, variables, and parameters.

4.2 *Agents*

The 65 households of Ban Que engaged in shifting cultivation are each represented as an agent in the simulation, comprising a total of 385 individuals. Each agent is described by two variables related to the population; a variable, *hh*, with the number of people in the household, and a variable holding the plot requirement. The *hh* variable values are distributed to all agents based on a Poisson distribution around the average household population of 6. The plot requirement variable is informing the agent about the demand for area based on the household population size. In the simulation, the relation between household population size and the household plot requirement is not continuous, as the potential farming area is represented in discrete 4900 m^2 cells. Therefore, the relation between the two variables is defined by threshold values, representing the assumption that an average person consumes approximately 500 g of rice per day (FAO 2003). This implies that the average yield of 1200 kg ha^{-1} from a 4900 m^2 plot provides rice for

$$\frac{\left(1200\,\mathrm{kg\,ha^{-1}\,year^{-1}}\right)\times 0.49\,\mathrm{ha}}{\left(0.5\,\mathrm{kg\,day^{-1}\,person^{-1}}\right)\times 365\,\mathrm{days}} = 3.22\,\mathrm{persons}$$

on an annual basis.

Table 3. Transition rules and values for the cell states, variables, and parameters *a*, *b*, and *c* of equation (2).

State	Transition rule/value
Cultivated	This parameter is controlled by agent decisions
Regrowth	If state_{t-1}=cultivated then set state_t=*regrowth*
	If state_{t-1}=regrowth then set $\mathrm{state}_t=\mathrm{state}_{t-1}$
Village	Village is a static state
Variable	
Slope	Slope is a static variable
Age	If state_{t-1}=*regrowth* and $\mathrm{state}_t=\mathrm{state}_{t-1}$ then set *age*=*age*+1
	If state_{t-1} ![a]=state_t then set *age*=0
Perimeter	If state_t=*cultivated* then set *perimeter* of cells in the von Neumann neighbourhood=*perimeter*+1
Labour productivity	$(\mathrm{Yield}_{max}/(1+(b\times\exp(-c\times\mathrm{age}))))/(-11\times\mathrm{perimeter}+156)$
Global parameters	
a (Yield_{max})	1875 kg ha^{-1}
b	6.75
c	0.48

[a]!: different from.

Agents with *hh* values lower than 3.5 are hence set to a plot requirement value of 1 by default. Agents with *hh* ranging between 3.5 and 7.5 have a plot requirement of 2, and agents representing households with more than 7.5 members are set to 3.

Rules specifying agent behaviour are kept very simple. Here, the agents do not have complete information on total production costs and potential yields. Rather, they act on very few inputs to find a cell to cultivate:

- the productivity based on the time in fallow (calculated using equation (2));
- the labour required to cultivate the cell (calculated based on the cell *perimeter* variable);
- the slope of the cell.

Information on the first two inputs, the fallow/regrowth age and the labour requirements for cultivation, are available to the agents from the cell variable *labour productivity*. As no cells are fenced at the beginning of each time step, the first agent finds a cell to fence and cultivate based solely on the slope of the cell and the productivity of the cell. In accordance with the findings from fieldwork interviews (section 3.1), agents consider only cells with slopes between 6 and 35° suitable for shifting cultivation. When choosing a cell, the labour productivity of the four neighbouring (von Neumann) cells is updated to reflect the fencing done by the agent and thus the decreased labour costs associated with fencing/cultivation of these particular cells, and the particular cell chosen by the agent changes its state from *regrowth* to *cultivated*. The next agent will therefore have an incentive to choose their cell next to the cell occupied by the first agent. This procedure continues until all agents have found a place to cultivate, and the procedure is repeated every time step. If an agent faces a set of cells with similar labour productivity, a random cell will be chosen from the set for cultivation. Figure 4 shows the model structure.

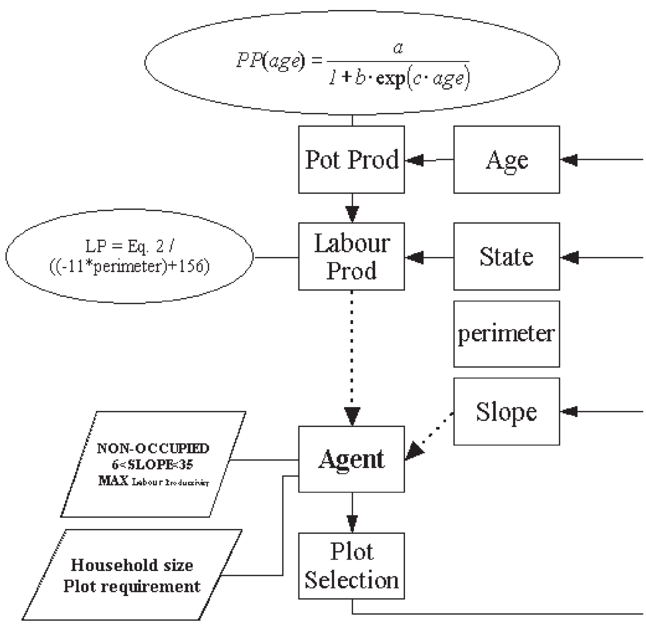

Figure 4. Model structure. PP(age): the potential production for a cell with a given age. LP: labour productivity. Labour productivity, age, state, perimeter, and slope are cell variables.

Table 1 shows the labour productivity matrix for given combinations of yield (based on fallow age) and degrees of fencing. Highlighted in grey are the combinations of fallow age/fencing making a cell more attractive to the agent than a 10-year-old fallow without fencing. For 9-year-old fallow cells, only one cultivated neighbour is necessary to raise the labour productivity over that of the 10-year unfenced cell, while 6-year-old cells must be cultivated on at least three sides to attract an agent compared with the 10-year-old unfenced cell.

4.3 *Model runs and scenarios*

The model is dynamic in space and time, resulting in distinct land-cover patterns. As a closed system, however, no information, energy, or mass is introduced to or exported from the model world once the simulation is set up. This implies that the simulated shifting cultivation system has a constant population and a constant area to cultivate, and as the model does not keep track of nutrient pools or declining yields, the simulated environment will reach a steady state with respect to the average age of the regrowth cells—the fallow period. At this point, the model stops, and the graphics are exported to ArcGIS for conversions and analyses.

To construct a baseline, 30 simulations were made with parameters and rules as specified in section 3.1 and table 3. Each simulation varies slightly due to randomness in the initial distribution of fallow ages, the household size distribution, and the specific cell chosen by an agent for cultivation if cells with similar labour productivities exist.

To evaluate how the system might change under different parameter values, three scenarios were constructed:

1. The '100-agent scenario,' where the number of agents is changed from 65 to 100, thereby changing the total population from 385 to 600. It is not unlikely that this scenario will happen in reality, given an annual population growth rate in Vietnam of more than 1%.
2. The 'plot-requirement scenario,' where the number of cells cultivated by each agent is changed by varying the *hh* and plot requirement variables. The default variable values are based on the average Vietnamese rice consumption which could be different in the study village due to different dietary composition. In this scenario, agents with households smaller than two persons cultivate two cells, agents with households between two and four persons cultivate three cells, and agents with households larger than (or equal to) four persons cultivate four cells.
3. The 'slope scenario,' building on the plot-requirement scenario, where the minimum and maximum thresholds for slope are gradually changed, while the *hh* and plot requirement variables are similar to the plot requirement scenario. The slope scenario tests the sensitivity of the model outcome to loosening the physical barriers for agent movement. Each of the model scenarios was simulated 30 times for comparison of results to the base scenario.

5. Simulation results

The cultivation patterns for the 30 simulations comprising the baseline, of which two are shown in figure 5, exhibit some similarity to the patterns observed from satellite imagery (figure 3). The average statistics for the baseline show that the

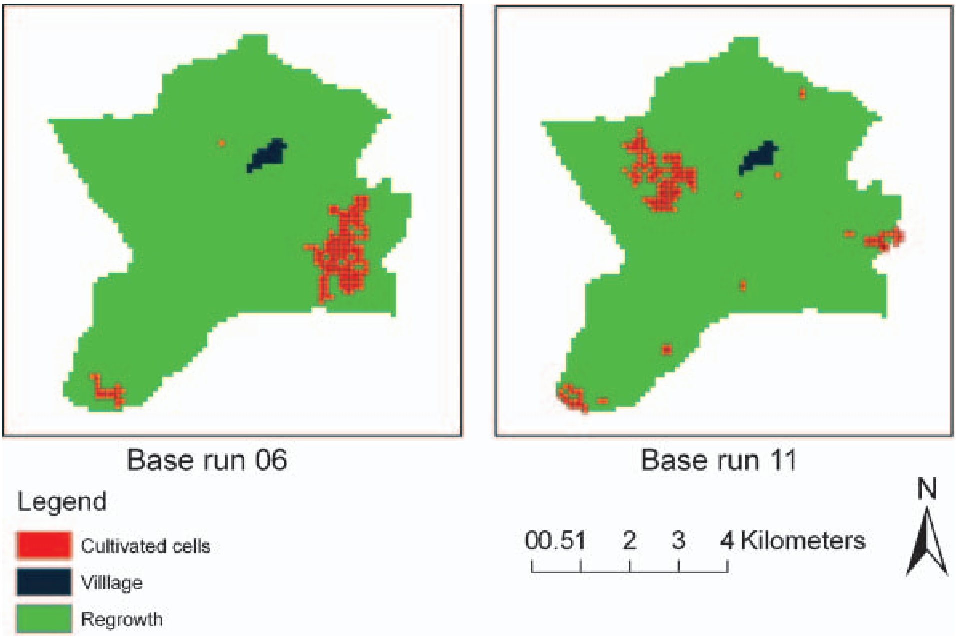

Figure 5. Two examples of the 30 runs comprising the baseline.

NNOMD, i.e. $\left(\sum_i C_i\right) \Big/ n$ of equation (1), is around half of the expected mean distance (EMD), based on the number of cultivated fields and the village territory area. As a consequence, the MNNR is 0.48, and the clustering index is supported by a strong negative Z-score of -11.72.

The MNNR statistics from the baseline runs compared well to the satellite-derived data. The values of MNNR for the observed data were 0.60, 0.59, and 0.58 (table 2). The baseline Z-score was similar to the satellite-based Z-scores. Examining the NNOMD and EMD of the baseline and the satellite data reveals important differences: the observed mean distance of the baseline was 75.36, while the satellite data had values in the range 71.54–80.87. The expected mean distance was 158.72 for the baseline compared with 122–134 for the satellite data. Most important, however, is the difference in the number of cultivated cells: the baseline value is 136, while the satellite validation data on average contain 214 cultivated cells (table 4).

The 100-agent scenario shows the same visual trends as the baseline, with a mixture of smaller and larger fields, but naturally with a greater number of plots than the baseline and with a less clustered field pattern. The MNNR is 0.83, and the Z-score is -4.65.

The plot requirement scenario forces the agents to cultivate a larger number of cells than the baseline scenario, the average number of 210 cells cultivated is close to the average of 214 for the validation data, and the EMD of this scenario (127.87) is therefore close to the EMD of the validation data (126.67). The NNOMD of 116.93 reveals a long distance between neighbouring cultivated cells, and the MNNR is 0.92 with a Z-score of -2.32, expressing rather weak clustering.

The slope scenario consists of a series of runs where the low slope threshold is varied from 1 to 6, and the high slope threshold is varied between 20 and 50 in 10° intervals (figure 6, table 5). The MNNR ranges between 1.03 with a z-score of 0.90

Table 4. Mean nearest-neighbour ratio, Z-score, and number of cultivated cells for the baseline simulation and three scenarios[a].

	Baseline (30 runs)	100 agents (30 runs)	Plot requirements (30 runs)	Slope (4 × 30 runs with low slope constant at 6°)			
				20	30	40	50
No. of cultivated cells	136	211	210	135	138	136	137
NNOMD	75.36 (4.06)	106.2 (7.74)	116.93 (7.49)	156.48 (10.41)	98.13 (8.32)	70.96 (2.37)	70.17 (0.92)
EMD	158.72 (2.82)	127.53 (1.96)	127.87 (3.28)	159.32 (2.38)	158.1 (2.92)	158.9 (2.99)	158.69 (3.02)
MNNR	0.48 (0.03)	0.83 (0.06)	0.92 (0.07)	0.92 (0.07)	0.62 (0.05)	0.45 (0.02)	0.44 (0.01)
Z–score	−11.72 (0.57)	−4.65 (1.61)	−2.31 (1.87)	−1.76 (1.61)	−8.5 (1.2)	−12.34 (0.33)	−12.46 (0.13)

[a]Numbers in parentheses are standard deviations.

for a low slope=6 and a high slope=20 and 0.57 with a z-score of −11.80 for a low slope=1 and a high slope=50.

6. Discussion

The baseline scenario produces a pattern of cultivated cells with a stronger clustering than the validation data, measured as the MNNR. But inspecting the NNOMD, the cultivated cells in the validation data are placed more adjacent than in the baseline scenario. The higher MNNR for the baseline arises because the scenario contains a lower number of cultivated cells than the validation data and

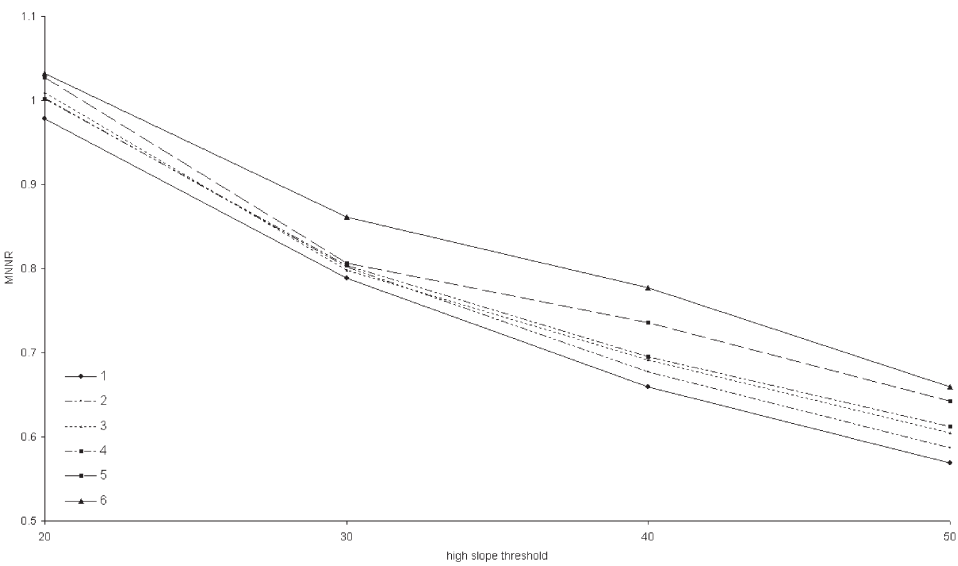

Figure 6. The spatial output from four simulations: 1) Baseline with default settings; 2) Baseline with the number of agents increased from 65 to 100; 3) Baseline with the c-constant of Equation 2 changed from 0.2 to 0.4 and 4) Baseline with maximum slope threshold changed from 35 to 50.

Table 5. Mean nearest-neighbour ratio, Z-score and number of cultivated cells for the slope scenario[a].

high_slope	low_slope	NNOMD	EMD	Z-score	Cultivated cells	MNNR
20.00	1.00	124.75 (5.96)	127.54 (3.33)	−0.58 (1.45)	211.50 (11.10)	0.98
20.00	2.00	126.95 (6.06)	126.56 (3.54)	0.10 (1.24)	214.83 (11.77)	1.00
20.00	3.00	128.70 (6.72)	127.55 (2.42)	0.25 (1.35)	211.27 (8.07)	1.01
20.00	4.00	127.81 (7.04)	127.53 (2.73)	0.06 (1.38)	211.40 (8.83)	1.00
20.00	5.00	132.15 (7.61)	128.69 (3.79)	0.74 (1.38)	207.87 (12.14)	1.03
20.00	6.00	131.38 (7.83)	127.24 (3.31)	0.90 (1.53)	212.50 (11.05)	1.03
30.00	1.00	100.57 (7.25)	127.57 (3.63)	−5.90 (1.38)	211.50 (11.99)	0.79
30.00	2.00	102.40 (6.23)	127.76 (3.58)	−5.51 (1.23)	210.83 (11.86)	0.80
30.00	3.00	101.82 (7.71)	127.59 (3.84)	−5.65 (1.37)	211.47 (12.68)	0.80
30.00	4.00	102.47 (6.71)	127.49 (2.73)	−5.47 (1.30)	211.53 (9.24)	0.80
30.00	5.00	103.80 (6.65)	128.62 (3.42)	−5.30 (1.45)	207.97 (10.84)	0.81
30.00	6.00	110.10 (6.21)	127.89 (3.70)	−3.83 (1.43)	210.43 (11.63)	0.86
40.00	1.00	83.99 (4.88)	127.39 (3.19)	−9.44 (1.17)	211.97 (10.48)	0.66
40.00	2.00	85.74 (7.10)	126.52 (2.70)	−8.98 (1.70)	214.80 (9.26)	0.68
40.00	3.00	87.42 (5.84)	126.46 (4.39)	−8.56 (1.52)	215.43 (14.29)	0.69
40.00	4.00	88.92 (7.86)	127.94 (3.46)	−8.39 (1.82)	210.20 (11.14)	0.70
40.00	5.00	93.79 (6.57)	127.42 (3.29)	−7.28 (1.69)	211.90 (10.95)	0.74
40.00	6.00	99.05 (7.39)	127.38 (3.64)	−6.09 (1.98)	212.13 (12.31)	0.78
50.00	1.00	73.39 (4.33)	128.99 (3.26)	−11.81 (1.01)	206.77 (10.58)	0.57
50.00	2.00	74.79 (4.71)	127.48 (3.70)	−11.44 (1.15)	211.80 (12.03)	0.59
50.00	3.00	76.90 (4.63)	127.07 (3.22)	−10.97 (1.11)	213.03 (10.56)	0.61
50.00	4.00	77.52 (3.77)	126.64 (2.96)	−10.83 (0.91)	214.43 (10.00)	0.61
50.00	5.00	81.37 (5.62)	126.74 (3.47)	−9.95 (1.38)	214.23 (11.62)	0.64
50.00	6.00	84.05 (4.08)	127.53 (2.88)	−9.45 (0.98)	211.43 (9.60)	0.66

[a]Each row presents mean values of 30 runs with standard deviations in parentheses.

hence has a higher EMD value. In general, the algorithm used for calculation of the labour productivity value of the cells (equation (2) divided by the cell perimeter variable) and the rules describing how the agents select cells to cultivate succeed in producing a clustered cultivation pattern, but the number of cultivated cells in the baseline scenario falls below the observed number.

The results from the 100-agent scenario were surprising. As indicated by the EMD of 127.53, the number of cultivated fields was similar to the satellite data, but the NNOMD was 106.20, higher than the baseline value. This results from cultivated cells not being situated as adjacent as in the baseline scenario and the validation data. The reason for the high NNOMD could be an artefact of the method used to stop the model; when the landscape change, measured as the cultivated-to-regrowth ratio, is constant between two time steps, the model stops. To test if the model is stopped prematurely by a 'quasi-stationary' level occurring before the landscape reaches a steady state, a test was carried out where the simulation runs 10 times longer than the average number of time steps for the 100-agent scenario, but this only decreased the NNOMD from 106 to 98.68. Another reason for the high NNOMD is that the physical barrier for cultivation, the slope threshold of 35°, is set too low, thus forcing the agents to avoid steeply sloping cells despite adjacency to other cultivated fields and related high labour productivity values, and thereby fragmenting the pattern of cultivated cells.

The 100-agent scenario partly compensates for the fact that the baseline underestimates the number of cultivated fields by letting more agents cultivate.

But this approach relaxes the relation between model and reality, as the empirical data on the number of households in Que village (65) are regarded as robust. Instead of changing this variable to increase the number of cultivated cells, we change the number of cells each of the 65 agents should cultivate by varying the *hh* and plot requirement variables. In this scenario, the plot-requirement scenario, the model reproduces the number of cultivated cells observed on the satellite images but fails to reproduce the clustering. Like the 100-agent scenario, this can be attributed to the physical limitations posed on agent choice by the slope thresholds.

In the slope scenario, the threshold values were varied, thereby testing the impact of the slope constraint on the pattern produced. With *hh* and plot-requirement variables like the plot-requirement scenario, the EMD matches the validation data and as the area available to the agents is increased by losing the slope barriers, the OMNND of the simulations approaches the average OMNND of the validation data. In the slope scenario, the simulations with a low slope=2 and high slope=50 yield results closest to the validation data, in terms of both MNNR and *z*-score (0.59 and -11.44 for the simulation, 0.59 and -11.53 for the validation data).

The mean nearest-neighbour ratio is a simple measure of spatial clustering. Inspecting the ratio's two components, the observed and expected nearest-neighbour distances, allows a more precise description of the patterns analysed. In the present context, validation of the model by the ratio alone would reveal similarities in clustering tendencies between modelled and observed data but would not measure if the model produces the correct number of cultivated cells or whether the mean nearest-neighbour distance in the model results fits the validation data. For instance, the baseline scenario matches the mean nearest-neighbour distance and the *z*-score of the validation data but only produces around two-thirds of the cultivated cells observed in the satellite data. As a consequence, the baseline EMD is 25% larger than the validation data EMD, and the MNNRs of the baseline and the validation data do not match.

Both the 100-agent scenario and the plot-requirement scenario arrive at the number of cultivated cells observed in the satellite data and thus match the EMD. But the OMNNDs and MNNRs for the two scenarios show that cultivated cells are spread across the landscape with weak clustering. Two reasons could exist for this: (1) the benefit of cultivating adjacently, expressed as the labour productivity variable of the cells, is too low; and (2) the area available to the agents is too restricted by the slope thresholds. While the first reason is not addressed in this study, the second reason is examined in the slope scenario which yields a result very similar to the validation data. This indicates that the slope thresholds reported by the interviewed villagers could be misestimates and that the farmers are willing to cultivate slopes as steep as 50°. Indeed, overlaying the cultivated cells observed in the satellite images with the slope map shows some cultivated cells situated on slopes between 40 and 53°. But even though the model contains a modest amount of variables, the number of combinations is huge, as is the formulation of rules governing agent's behaviour. It is therefore not unlikely that other rules and variable combinations would produce results similar to the slope scenario, making it difficult to reach a firm conclusion about the model's performance.

6.1 *Model shortcomings*

The model is a rough simplification of reality. In a real-world situation, yield is partly dependent on the labour invested in production. When setting the labour productivity of a cell in the model, yield is calculated from the cell age, and labour costs are calculated based on the number of adjacent cells that are cultivated. No feedback exists between labour costs (or investments) and yield. In addition, while real-world subsistence farmers probably do consider labour productivity as an important factor, the size of the actual yield, the productivity, might be crucial for decision-making. This is simulated here through the plot requirement variable, which does not respond linearly to the household size but responds stepwise following the plot-requirement thresholds. An interesting improvement in the model would be to keep track of which cells an agent is cultivating and to obtain empirical data on the 'ownership' of fields. This would enable a micro-scale validation of the simulations by testing if the modelled tendencies of each agent to locate their own cells adjacently fits with the empirical observations.

In contrast with one of the often claimed principles of cellular automata and multi-agent systems, the agents here do not process information in parallel but 'line up' and execute the decision serially for each time step. This approach is highly intentional: before selecting a cell, agents check the productivity value of the cells and check if the cell is occupied by another agent. Then, the agent chooses an unoccupied cell to cultivate. If this sequence of commands were to be executed in parallel, all agents could in principle end up at the same cell. The serial processing is done in the same order for each time step. This means that the same agent can always choose first. As agent performance in terms of productivity or labour costs is not part of the model outcome, this approach does not impact the resulting land cover.

Finally, the procedure for stopping simulations could be improved. A method yet to be tested is to calculate the MNNR for each model time step and keep a list of MNNRs for a range of time steps. The model could then be stopped with the MNNR, for a given time step is within a given threshold of the average of the MNNRs in the list.

The mean nearest-neighbour distance and mean nearest-neighbour ratio provide a reliable first-order neighbourhood measure. But as a validation tool, it has a shortcoming: if all cultivated cells have a neighbour, the OMNND will be 70, the smallest distance between any two cells, thus indicating perfect clustering. This will also be the case if cultivated cells are located in dispersed pairs. To measure the size of clusters, and thus provide a more precise measure of patterns, the nearest-neighbour distance should be expanded to include distances to the second nearest neighbour, third nearest neighbour, etc. Application of a *k*-order nearest-neighbour distance index (Mitchell 2005) would be an interesting feature in the validation of spatially explicit agent-based models.

7. Conclusion

Agent-based models operating in simulated environments have a strong logical and pedagogical appeal. Decision-makers, be they farmers, farm households, or actors at other organizational levels or geographical scales, are represented as agents acting on their surroundings, represented as cells, which in turn provide opportunities and constraints to the agents.

The baseline simulation presented here is almost entirely based on empirical data, and fieldwork has served to identify and parameterize variables. The rules in this simulation are simple, revolving around labour productivity and the impact of physical barriers on adjacent farming.

The ability of the model simulations to generate spatial clustering and approach the validation data indicates that the 'core' rule in the model, i.e. the search for cells with high labour productivity, could be a realistic representation of farmer decision-making. The model succeeds in producing a land-use pattern consisting of smaller and larger fields, and the pattern produced in the slope scenario is similar to the validation data measured by the spatial statistics applied. This supports traditional shifting cultivation theory: farmers act to maximize their labour productivity, sometimes at the expense of maximizing the yield. But it should be kept in mind that validating rules of locational behaviour by their spatial expression is still a 'proxy validation', and that the observed patterns could be reproduced using many different rules and model types.

References

AGARWAL, C., GREEN, G.M., GROVE, M.J., EVANS, T.P. and SCHWEIK, C.M., 2002, A review and assessment of land-use change models: dynamics of space, time and human choice. USDA Forest Service Northeastern Forest Research Station, Center for the Study of Institutions, Population, and Environmental Change at Indiana University–Bloomington and the USDA Forest Service. CIPEC collaborative report series.

AMARASINGHE, U., SAMAD, M. and ANPUTHAS, M., 2005, Spatial clustering of rural poverty and food insecurity in Sri Lanka. *Food Policy*, **30**, pp. 493–509.

ANGELSEN, A., 1995, Shifting cultivation and 'deforestation': A study from Indonesia. *World Development*, **23**, pp. 1713–1729.

ANTROP, M. and VAN EETVELDE, V., 2000, Holistic aspects of suburban landscapes: visual image interpretation and landscape metrics. *Landscape and Urban Planning*, **50**, pp. 43–58.

BENENSON, I. and TORRENS, P.M., 2004, Geosimulation: Automata-based modeling of urban phenomena. Wiley.

BERGER, T., 2001, Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. *Agricultural Economics*, **25**, pp. 245–260.

BOSERUP, E., 1965, *The Conditions of Agricultural Change* (London: Earthscan).

BOUSQUET, F. and LE PAGE, C., 2004, Multi-agent simulations and ecosystem management: a review. *Ecological Modelling*, **176**(3–4), pp. 313–332.

BROWN, D.G., PAGE, S., RIOLO, R., ZELLNER, M. and RAND, W., 2005, Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science*, **19**(2), pp. 153–174.

BROWN, D.G., WALKER, R., MANSON, S. and SETO, K., 2004, Modeling land use and land cover change. In *Land Change Science. Observing, Monitoring and Understanding Trajectories of Change on the Earth's Surface*, G. Gutman, A.C. Janetos, C.O. Justice, E.F. Moran, J.F. Mustard, R.R. Rindfuss, D.L. Skole, B.L. Turner II and M.A. Cochrane (Eds), pp. 395–409 (Dordrecht: Kluwer Academic).

CASTELLA, J.C., BOISSAU, S., TRUNG, T.N. and QUANG, D.D., 2005, Agrarian transition and lowland-upland interactions in mountain areas in northern Vietnam: application of a multi-agent simulation model. *Agricultural Systems*, **8**, pp. 312–332.

CHRISTIANSEN, S., 1992, A new attempt at an ecological classification of land utilization systems. *Danish Journal of Geography*, **92**, pp. 54–60.

DEADMAN, P., ROBINSON, D., MORAN, E. and BRONDIZIO, E., 2004, Colonist household decisionmaking and land-use change in the Amazon Rainforest: an agent-based simulation. *Environment and Planning B*, **31**, pp. 693–709.

DUNCAN, J.L., PERRY, J.N., DALE, M.R.T., LEGENDRE, P., CITRON-POUSTY, S., FORTIN, M.J., JAKOMULSKA, A., MIRITI, M. and ROSENBERG, M.S., 2002, A balanced view of scale in spatial statistical analysis. *Ecography*, **25**, pp. 626–640.

FAO 2003, *Proceedings of the 20th Session of the International Rice Commission*, Bangkok, 23–26 July 2002 (Rome: FAO).

GUTMAN, G., JANETOS, A.C., JUSTICE, C.O., MORAN, E.F., MUSTARD, J.F., RINDFUSS, R.R., SKOLE, D., TURNER II, B.L. and COCHRANE, M.A., 2004, *Land Change Science. Observing, monitoring and understanding trajectories of change on the earth's surface.* (Dordrecht: Klumer Academic Publishers).

HALENDA, C.J., 1989, *The Ecology of a Fallow Forest after Shifting Cultivation in Niah Forest Reserve* (Sarawak, Malaysia: Forest Department).

JAKOBSEN, J., 2005, A shifting cultivation system in transition—a village case study from the uplands of North Central Vietnam. Master thesis, Institute of Geography, University of Copenhagen.

KATES, R.W., HYDEN, G. and TURNER II, B.L., 1993, Theroy, evidence, study design. In *Population Growth and Agricultural Change in Africa*, B.L. Turner II, G. Hyden and R.W. Kates (Eds), pp. 1–41 (Gainesville: University Press of Florida).

KLÜGL, F., FEHLER, M. and HERRLER, R., 2005, About the role of the environment in multi-agent simulations. *Environments for Multi-Agent Systems*, **3374**, pp. 127–149.

LAWRENCE, D., 2005, Biomass accumulation after 10–200 years of shifting cultivation in bornean rain forest. Ecology, **86**, pp. 26–33.

LEUNG, Y., MEI, C.L. and ZHANG, W.X., 2003, Statistical test for local patterns of spatial association. *Environment and Planning A*, **35**, pp. 725–744.

MANSON, S.M., 2000, Agent-based dynamic spatial simulation of land-use/cover change in the Yucatán peninsula, Mexico. In *4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4)*, Banff, Canada.

MANSON, S.M., 2005, Agent-based modeling and genetic programming for modeling land change in the Southern Yucatan Peninsular Region of Mexico. *Agriculture Ecosystems & Environment*, **111**, pp. 47–62.

MERTZ, O., 2002, The relationship between length of fallow and crop yields in shifting cultivation: a rethinking. *Agroforestry Systems*, **55**(2), pp. 149–159.

PAN, W.K.Y., WALSH, S.J., BILSBORROW, R.E., FRIZZELLE, B.G., ERLIEN, C.M. and BAQUERO, F., 2004, Farm-level models of spatial patterns of land use and land cover dynamics in the Ecuadorian Amazon. *Agriculture, Ecosystems & Environment*, **101**, pp. 117–134.

MITCHELL, A., 2005, *The ESRI Guide to GIS Analysis: Volume 2: Spatial Measurements and Statistics* (Redlands, CA: ESRI Press).

PARKER, D.C., MANSON, S.M., JANSSEN, M.A., HOFFMANN, M.J. and DEADMAN, P., 2003, Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers*, **93**, pp. 314–337.

PARKER, D.C. and MERETSKY, V., 2004, Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics. *Agriculture, Ecosystems & Environment*, **101**, pp. 233–250.

RASMUSSEN, K. and MøLLER-JENSEN, L., 1999, A generic model of shifting cultivation. *Danish Journal of Geography*, Special Issue, **1**, pp. 157–164.

READ, L. and LAWRENCE, D., 2003, Recovery of biomass following shifting cultivation in dry tropical forests of the Yucatan. *Ecological Applications*, **13**, pp. 85–97.

RINDFUSS, R.R., WALSH, S.J., TURNER II, B.L., FOX, J. and MISHRA, V., 2004, Developing a science of land change: challenges and methodological issues. *PNAS*, **101**, pp. 13976–13981.

TOTTRUP, C., in press, Forest and land cover mapping in a tropical highland region using linear mixture modeling and decision tree classification of high-spatial resolution image data. *Photogrammetric Engineering and Remote Sensing*.

TURNER, M.G., 1990, Spatial and temporal analysis of landscape patterns. *Landscape Ecology*, **4**, pp. 21–30.

TURNER, M.G., O'NEILL, R.V., GARDNER, R.H. and MILNE, B.T., 1989, Effects of changing spatial scale on the analysis of landscape pattern. *Landscape Ecology*, **3**, pp. 153–162.

VERBURG, P.H. and VELDKAMP, A., 2005, Introduction to the special issue on spatial modeling to explore land use dynamics. *International Journal of Geographical Information Science*, **19**, pp. 99–102.