Agent-based modeling and genetic programming for modeling land change in the Southern Yucatán Peninsular Region of Mexico

Steven M. Manson

Department of Geography, 414 Social Sciences, 267 19th Avenue South, University of
Minnesota, Minneapolis, MN 55455

Abstract: Land-use and land-cover change research increasingly takes the form of integrated land-change science, the explicit joining of ecological, social, and information sciences. Traditional interdisciplinary methods are buttressed by new ones stemming from computational intelligence research and the complexity sciences. Several of these—genetic programming, cellular modeling, and agent-based modeling—are applied to land-change in the southern Yucatán peninsular region (SYPR) of Mexico through the SYPR Integrated Assessment (SYPRIA). This work illustrates how computational intelligence techniques, such as genetic programming, can be used to model decision making in the context of human-environment relationships. This application also contributes to methodological innovations in multicriteria evaluation and modeling of coupled human-environment systems. This effort also demonstrates the importance of considering both social and environmental drivers of land change, particularly with respect to the decision making of change agents within the context of key socioeconomic and political drivers, particularly as channeled through market institutions and land tenure, and ecological factors, especially characteristics of land use and land cover such as state, history, and fragmentation. SYPRIA demonstrates the utility of modeling methods based in computational intelligence and the complexity sciences in helping understand the decision making of land-change agents as a function of both social and environment drivers.

Keywords: agent-based model, genetic program, land-use and land-cover change, multicriteria evaluation, symbolic regression

Citation: Manson, S.M. (2005). Agent-based modeling and genetic programming for modeling land change in the Southern Yucatan Peninsular Region of Mexico. *Agriculture, Ecosystems and Environment* 111(1): 47-62.

PREPRESS VERSION

1. Introduction

Global environmental change is occurring at an unprecedented rate and magnitude, giving rise to a situation unparalleled in human experience (Steffen et al., 2003). Important to this change is human activity that impinges on land-use systems and surficial features, known as ***land-use and land-cover change*** (LUCC). In addition to contributing roughly a quarter of anthropogenic carbon dioxide, LUCC affects myriad ecosystem services and governs the vulnerability of hundreds of millions of people. LUCC is also a phenomenon that defines a good deal of the dynamics occurring at the interface between human agricultural systems and ecological systems. Helping us understand LUCC is the emerging field of integrated ***land-change science*** (LCS), the joining of social, ecological, and information sciences designed to aid in understanding LUCC (Gutman et al., 2004). A core endeavor is the development of spatially explicit, dynamic LUCC models that examine human decision making in social and ecological contexts. LCS is a powerful means of addressing land change but it contends with two general issues.

First, there is a continuing need to develop appropriate methodologies that run the gamut from simple mathematic formulae to intricate simulations. There is mounting interest in methods based in the so-called complexity sciences, which deal with deterministic complexity, particularly as understood by chaos and catastrophe theories, and aggregate complexity, or how individual elements working in concert create complex systems that exhibit learning and emergence, with approaches including genetic algorithms, classifier systems, cellular automata, and neural networks (Manson, 2001; Agarwal et al., 2002; Verburg et al., 2005). Holding particular promise are ***agent-based models*** (ABM, also known as multiagent systems), systems of software agents that represent adaptive, autonomous entities that extract information from their surroundings and apply it to functions such as perception, planning, and learning. Agents can be used to represent the complex behavior (particularly in terms of aggregate complexity) of interacting entities in the context of coupled human-environment systems (Gimblett, 2002; Janssen, 2003). ABM represent complexity resulting from interactions among entities; provide an analytically tractable means of modeling spatial and temporal interdependencies and heterogeneity among agents and with their environment; examine how emergent system outcomes stem from actor behavior; and explicitly represent individual rationality and learning (Parker et al., 2003).

Second, there is a need for explicit consideration of decision making as key to understanding the behavior of change agents (Agarwal et al., 2002). Another complexity method that is promising in this respect is ***evolutionary computing***, which draws on biological metaphors to create computer programming systems. One variant of evolutionary programming uses genetic programs, software programs formatted as decision trees, with branchings of mathematical functions, logical operators, or user-defined functions and terminal leaves of function arguments (Koza, 1992). Genetic programming is a directed search method that uses the computational equivalent of natural selection to evolve software programs to solve problems. Simultaneous evolution of program structure and parameters enables this method to find solutions in highly dimensional, noisy, stochastic environments. It is applied over domains including industrial design, pattern recognition, imagery classification, neural net construction, time series modeling, and signal processing (Banzhaf et al., 1998; Krzanowski and Raper, 2001).

There is a growing body of research on how genetic programming and ABM methods may be used together to model decision making. An agent invested with a population of genetic programs has multiple strategies with which to deal with structural changes in its environment

(Edmonds, 1999). The evolution of these strategies can represent memory, learning and innovation and act as a means by which agents can share strategies in order to engage in imitation and communication (Brenner, 1999). Exploring the potential for using computational intelligence approaches such as genetic programming to represent human decision making speaks directly to the long-term goal of LUCC researchers to understand the complex dynamics of coupled human-environment systems. At the same time, these methods are the subject of continuing research because they can be difficult to interpret and there are dangers in conflating human decision making with biologically-inspired models of computer programming.

This article describes the SYPR Integrated Assessment (SYPRIA), a LUCC model applied to the southern Yucatán peninsular region of Mexico. Section 2 describes the study area, data used, and the SYPRIA actor-institution-environment conceptual framework; all of which support a computational model comprised of a geographic information system (GIS) tied to agent-based modeling, cellular modeling, and genetic programming. Section 3 presents model results with respect to LUCC theory as embodied in the conceptual framework. The final sections discuss the model results and give concluding thoughts.

2. Methodology

Study area and data

The southern Yucatán peninsular region lies on Mexico’s border with Guatemala and Belize (Figure 1). The 22 000 km^2 region is home to deforestation that affects arable land availability, regional climate, secondary succession, and soil quality. The region was sparsely populated until Highway 186 was built through its width in 1967, opening the way to immigration. Into the 1980s, smaller, more densely populated agricultural settlements began to absorb immigrants from other parts of the country. In 1992, the national government established the 7232 km^2 Calakmul Biosphere Reserve (CBR) for conservation and ecoarcheological tourism initiatives (Primack et al., 1998). Overall, the population of the region increased from approximately 2500 in 1960 to over 30 000 today (Klepeis and Turner, 2001).

< Figure 1 here >

The Land-Cover and Land-Use Change in the Southern Yucatán Peninsular Region Project (SYPR) examines LUCC dynamics by integrating historical, social-science, remote sensing, and ecological analysis (Turner et al., 2004). The project provides the bulk of the data for SYPRIA. This includes a 1997/1998 survey of eleven *ejidos* (local administrative regions) and 188 randomly selected *ejidatarios* (residents of an ejido) for household characteristics, production activities, labor, institutional ties, and information on distinct plots. SYPR also generated spatial data, including land-use and land-cover maps from Thematic Mapper (TM) imagery for 1987, 1992, and 1995; elevation, slope, and aspect from a 1:50 000 digital elevation model; soil categories from a 1:250 000 map created by the Mexican National Institute for Statistics, Geography, and Information (INEGI, *Instituto Nacional de Estadistica, Geografía e Informatica*); a road network from INEGI 1:50 000 topographic maps; precipitation interpolated from twenty-one climate stations; surface hydrology; and socioeconomic and demographic data from the federal census.

3.1 Conceptual framework

Much global environmental change research employs the so-called ***driving forces and proximate causes framework*** (Stern et al., 1992). It sees five drivers of change: population growth; economic growth; technological change; political and economic institutions; and cultural attitudes and beliefs. Proximate causes are activities that directly impinge on the environment or indirectly when the products of these activities affect the environment. The LUCC research community rearticulates this schema as the foci of social systems, ecological systems, and land managers (IGBP-IHDP, 1995), which SYPRIA recasts as an actor-institution-environment conceptual model in which institutions and the environment act as conduits through which driving forces are channeled into proximate causes via decision making.

3.1.2 Actors

Key actors in the region are smallholder households that diversify their production activities to include subsistence-oriented *milpa* cultivation, agroforestry, modest logging, and market-oriented cultivation (Klepeis, 2003; Abizaid and Coomes, 2004). Extensive agriculture is the most common production activity and has the greatest LUCC impact, but commercial cultivation is becoming increasingly important as well (Keys, 2004). Two bodies of theory help us understand how households decide where to locate these production activities with respect to institutional and environmental factors. First, theories of relative space posit returns to land as a function of distance to phenomena that influence costs of inputs or prices of output, such as distance to market or specific land cover classes (e.g., bid-rent Alonso or Von Thünen circles). Related theories of absolute space consider *in situ* landscape characteristics (e.g., the Ricardian view), such as soil quality or land tenure (Bockstael, 1996). Second, agrarian theories account for the interplay of population, resource availability, technology, and market forces that affect production strategies through intensification, extensification, technological changes, fertility and labor adjustments, and migration (Bilsborrow and Carr, 2001).

3.1.3 Environment

Actor production has a biogeophysical context defined by climate, soil characteristics, secondary growth regimes, and disturbances such as fire and pest invasions. Production strategies feedback on the environment through pathways such as cover conversion, soil erosion, or nutrient depletion (Lawrence and Foster, 2003). The southern Yucatán peninsular region has a subhumid tropical climate and annual precipitation varies north to south from 900 mm to 1400 mm. Given the general tendency towards dry conditions, areas of higher precipitation are expected to be more likely to have agriculture. The region sits on a limestone plateau with elevation ranging from 200 m to 300 m amsl while the surrounding lowlands descend to less than 100 m. The thin rocky soils (lithosols) found at higher elevations and clays (gleysols and vertisols) found in depressions are generally less productive than the clay loams (rendzinas) found elsewhere, and agriculture is therefore more likely on the latter. Elevation is expected to be negatively correlated with agriculture because higher areas tend to be more rugged, less accessible, and have thinner, rockier soils with less moisture. Expected effects of slope and aspect are mixed given their varying relationships to soil quality and insolation and the role of agricultural technology (e.g., gentle slopes are better suited to mechanized agriculture).

< Table 1 here >

The region is dominated by seasonal tropical forests, including lowland (*bajo*) forest, upland (*mediana*) forest, savanna, agricultural land, secondary succession, and bracken fern (*Pteridium aquilinum*) (Table 1). Fallow-cycle dynamics dictate the nature of weed competition, pests, disease, and soil fertility. Time in agriculture of 2 to 5 years and a fallow of 20–30 years is optimal for restoring soil fertility and forest biomass sufficient to act as fuel for fires to kill pest plants and insects, while ten years is the shortest fallow period seen as adequate as long as herbicides and fertilizer are applied (Lawrence and Foster, 2003). Accordingly, agriculture is expected to replace secondary succession, agriculture, and upland forests. Land-use fragmentation is also posited to affect agricultural location because it gives access to the forest interior and greater accessibility for hunting (Sierra, 2000).

3.1.4 Institutions

Institutions guide smallholder decision making, particularly ejidos, the market, and conservation programs. ***Ejidos*** are administrative units that serve as communal land management councils, each defined by an area with a collective tenure system and its community (Figure 2). The Institutional Analysis and Development (IAD) formulation conceives institutions as "shared concepts used by humans in repetitive situations organized by rules, norms, and strategies" (Ostrom, 1999). Rules are predictably enforced shared prescriptions while costs and inducements enforce norms. Strategies are the regularized plans that individuals make given their understanding of rules, norms, and the likely behavior of others. IAD sees institutions as conduits for driving forces such as population or culture that by themselves are difficult to link to proximate causes. Driving forces are difficult to characterize at the regional scale but they almost necessarily act through institutions such as markets or land tenure that can be modeled.

< Figure 2 here >

IAD provides concrete measures of institutions by delineating seven kinds of rules: 1) boundary rules establish actor roles according to characteristics such as age or gender; 2) position rules define the place of actors in institutions; 3) authority rules define allowable participant behavior; 4) scope rules define which outcomes are allowed, mandated, or forbidden, and as such, are the context in which authority rules play out; 5) aggregation rules account for the net impact of individual behavior; 6) information rules determine what is known by whom; and 7) payoff rules establish the outcomes of actor behavior.

IAD rules developed from the household survey and the research literature characterize ejidos, markets, and conservation institutions. Boundary and position rules define key actors in ejidos as individual households. Authority, boundary, and scope rules establish permissible land uses by expressing how government agencies limit land use both inside and outside of ejidos (national lands and the CBR) and how ejido councils determine land tenure within the ejidos (usufruct and forest reserves). Aggregation rules map onto ejidal land use resulting from net actor land use. Information rules make institutional information and imperatives available to pertinent actors. Payoff rules are determined by households as a function of their production activities and household needs (e.g., market versus subsistence orientation) as tempered by the market and government programs. The Direct Rural Support Program (PROCAMPO, *Programa de Apoyo Directo al Campo*) provides direct payments to cultivators of basic food crops, like corn, beans, and rice, on a per-hectare basis (Klepeis, 2003).

Decision-making theories lend additional context to IAD rules. Deforestation is expected to be less likely the further a location is from a road, market, or village. While the relationship between population and land use can be difficult to articulate under agrarian theories, population is expected to be positively related to extensive land use as it generally requires more land under cultivation. Under theories of absolute space, land use should occur on secondary succession, as ejidos assign plots to households on a long-term basis with the understanding that actors engage in commonly held crop-fallow practices. Finally, government payments are expected to encourage agricultural land use.

3.2 Computational framework

The actor-institution-environment framework provides the conceptual underpinnings for the SYPRIA model methodology. The environment, represented by a cellular model, represents effects of actor behavior and endogenous transitions (i.e., ecological transitions separate from actor influence). Actors are represented by software agents in the ABM and their decision making is influenced by the simulated environment, other actors, and institutions represented by a separate collection of agents. SYPRIA is written in C++ and works with the Idrisi Geographic Information System to provide control structures, variables, and functions. It has a cellular model with which to represent the environment and an ABM with which to model actors and institutions. During a simulation run, SYPRIA iterates across single model 'years' in which three processes occur. First, institutions change variables related to actor decision making. Second, the environment changes according to endogenous ecological rules and the effects of actor decision making during the previous time step. Third, each actor in the region makes land-use decisions, essentially a multicriteria evaluation, that account for institutional and environmental factors. Each simulation run is part of a Monte Carlo series where actor heterogeneity and probabilistic environmental transitions introduce stochasticity.

Like other simulation models, SYPRIA can be treated in a variety of ways (e.g., scenario generation or exploratory 'what-if' analysis). The focus here, however, is on examining links between land use and environmental and institutional factors. A response variable, agricultural land use in 1992, varies as a function of 1987 environmental and institutional predictor variables chosen for their effects as hypothesized by land-use theory. A total of 3200 agent decision-making strategies were sampled over 100 Monte Carlo runs in order to understand how environmental and institutional factors influence actor decision making. Actors use these rules to project 1995 land-use that in turn provides the basis for comparison to actual 1995 land use. The effects of the environment cellular model are minimized by the relatively short time horizon of the application described here; secondary succession and changes in soil quality are explicitly modeled but other unchanging environmental factors are also important to understanding actor decision making.

3.2.1 Environment

A cellular model represents the environment. It is based on a generalized cellular automata, a two-dimensional grid where cell values change over time according to rules based on the value of adjacent cells, previous states, and external inputs (after Takeyama and Couclelis, 1997). Grids are identical to the Idrisi GIS raster layers used to store spatial data and SYPRIA therefore uses these layers as a base for the cellular model, defined as

$$[\mathbf{S},\mathbf{N},\mathbf{T}] \tag{1}$$

where $\mathbf{S}$ is a set of finite cell states, $\mathbf{N}$ is the set of states of neighboring cells, and $\mathbf{T}$ is a set of transition rules that pair input-state tuples to an output state for the cell. At regular time steps, each cell is tested to ascertain its state in the next time step as a function of $\mathbf{T}$. While $\mathbf{T}$ is a fully enumerated set, the cellular model allows generalized rules with wildcards that saves the user from having to code every permutation, conditional statements, references to cells in other layers, and probabilistic rules. All data were converted to the Idrisi GIS raster data structure with a cell resolution of 28.5 m^2 native to the satellite imagery.

Soil fertility and vegetation succession are represented by transition rules based on the conceptual framework and the spatial data detailed above. A fertility index (Q) acts as a simple proxy to actual soil fertility and is the subject of transition rules detailed in Table 2. Q is a function of: 1) land cover in the past time step; 2) soils of the types described in the conceptual framework and their respective value for agriculture; 3) soil fertility in the past time step; 4) duration of the cell's present land use, which can draw down soil quality without inputs; and 5) application by actors of inputs such as fertilizer. Possible transitions between land-use and land-cover types are constrained to those where a cell moves to agriculture, from agriculture, and between secondary succession classes. These transitions are governed by transition rules for secondary succession including bracken (Table 3). These relate to: 1) time since last transition; 2) total number of neighboring cells acting as seed sources; and 3) fallow-cycle dynamics reflecting soil fertility (Q) given in Table 2 as a function of past soil quality and inputs used by actors.

< Table 2 here >

< Table 3 here >

The cellular model also represents less dynamic (with respect to the timescale in question) environmental factors of importance to decision making, including elevation, slope, aspect, precipitation, and a soils dummy variable corresponding to soils considered poor for agriculture. The model recalculates distance to land use and land cover classes that change due to endogenous transitions or actor behavior. Fragmentation (F) is also updated

$$F = \frac{(n-1)}{(c-1)} \tag{2}$$

where c is the number of cells considered in the kernel (25 in this application) and n is the number of different classes present in the kernel.

3.2.2 Institutions

SYPRIA represents actors and institutions with an ABM. Agents are capable of movement across cellular model layers, communication, and decision making about production activities. Most of this functionality is invested in actors, considered below, although SYPRIA also uses agents to represent institutions. IAD rules define institution-agents as having spatial characteristics (expressed as GIS layers) and rules that influence actors. Employing institutions as a proxy for social systems is useful because changes in social systems that are important to actor decision making are channeled through institutions. Changes in national scale biodiversity policy, for example, can be seen in changes to land tenure and land access institutions like the CBR.

Population information is used as a proxy for local population, with the key caveat that relationships between population and land use are one-to-one under agrarian theories (Bilsborrow and Carr, 2001). A population institution-agent spawns actor-agents probabilistically according to population densities calibrated to match actual population densities given by census data for 1987 (interpolated between 1985 and 1990 with a second-order polynomial), 1990 (directly from 1990 data), and 1995 (directly from 1995 data). Figure 2 shows the 1995 distribution of the entire population in the study site with respect to ejidos. The mean household size in the southern Yucatán peninsular region is 6.4 (SD 1.4).

IAD rules developed from the household survey and the conceptual framework determine how SYPRIA represents institutions computationally. Boundary and position rules define key actors in ejidos as individual households. Ejidos probabilistically assign an average of 5 ha under milpa (4.6 – 5.4 ha/household) and 2 ha under commercial cultivation (1.3 - 2.5 ha/household). A three-year rotation for milpa and use of inputs to maintain constant acreage under market agriculture give an average of 17 ha for any given three year period (5 ha milpa × 3 years + 2 ha chili). Within this acreage, households determine where to locate production. Ejidos and conservation agents use authority and scope rules to limit land uses in the CBR and ejidal forest reserves. The market institution makes available to actors information on government direct payments (varying by ejido) and travel-cost surfaces for markets, population centers, and roads as a function of primary and secondary roads.

3.2.3 Actors

Actors exist in a landscape that corresponds to GIS spatial layers subject to the cellular model environment and ABM institutions. Each actor can be seen as solving its own multicriteria evaluation problem (after Eastman et al., 1995) where an actor-agent determines the suitability, *S*, of a set of grid cells for a given production activity:

$$S = \sum_{i=1}^{m} w_i v_i \prod_{j=1}^{n} b_j \tag{3}$$

as a function of continuously varying spatial factors $V = \{v_1, \ldots, v_m\}$, a set of factor weights $W = \{w_1, \ldots, w_m\}$, and a set of Boolean constraints $B = \{b_1, \ldots, b_n\}$. Environmental and institutional conceptual framework components determine the spatial factors V and constraints B while agents determine W. Eqn (3) gives each agent a continuous surface within their ejido that indicates propensity for agriculture as a function of environmental and institutional factors. Each agent denotes a subset of cells by choosing a threshold for S corresponding to its allotted acreage and then allocating cells to maximize aggregate suitability on a first-come, first served basis.

Actor-agents treat Eqn (3) as an instance of ***symbolic regression***, where response-variable observations are used to estimate parameters of independent predictor variables. Function $f(x)$ is known solely through observations $X = \{x_1, \ldots, x_n\}$. The observed value of the function at data point x_i, or $f(x_i)$, is denoted $\bar{f}_i$ and is related to the true value f_i through error defined as $\varepsilon_i = f_i - \bar{f}_i$. Symbolic regression gives an approximation of $f(x)$:

$$\hat{f}(x) \approx \sum_{j=1}^{n} a_j \phi_j(x) \tag{4}$$

where $\hat{f}(x)$ is a combination of functions $\phi_j(x)$ with coefficients a_j estimated by some means that minimizes ε_i over the observations (Ralston and Rabinowitz, 2001). $X = \{x_1, \ldots, x_n\}$ are cells in spatial layers corresponding to the response variable of agricultural land use in 1992 that vary as a function of 1987 environmental and institutional predictor variables. Function $\hat{f}(x)$ is an approximation to Eqn (3) determined by individual actors with genetic programming, which offers the methodological and theoretical advantages noted above.

Genetic programs serve as symbolic regression solutions with a function set ($\boldsymbol{F}$) composed of arithmetic operators (+ – ÷ ×) and a terminal set ($\boldsymbol{T}$) defined by spatial factors V and B in Eqn (3). Terminals also include ephemeral random constants, random numbers used by genetic programs to solve symbolic regression; if a genetic program tree were to be composed solely of terminals that correspond to spatial and aspatial variables, it is unlikely that a suitable tree could evolve to solve the symbolic regression equation, since the genetic program could only work with values given by the spatial variables. The power of genetic programming lies in computational analogs to natural selection (and therein lies the name). Each member of the initial population of programs has a tree structure randomly constructed from functions and terminals (Figure 3). Population K_0 is the first of G generations where individual programs are parents to offspring that constitute the following generation K_{i+1}. The initial generation are poor solutions because they are random agglomerations of functions and terminals, but each succeeding generation becomes better because individuals create offspring via three operators—mutation, crossover, and reproduction—that are applied to the programs' tree structures over many generations in order to evolve better programs.

< Figure 3 here >

Crossover is equivalent to breeding because it involves trading portions of two parent programs to create two offspring programs (agent strategies) that contain random portions of parental tree structures (analogous to genes). Reproduction is like cloning because it involves placing a duplicate of a parent into the next generation as offspring. It is useful for maintaining coherence of strategies across generations. Mutation is akin to genetic mutation because it randomly changes constituent nodes of a parent to create a new offspring program. In addition to ensuring that programs can jump into new parts of the search space, mutation protects terminals and functions from disappearing from the population.

While mutation is applied randomly, the two parents for crossover and the one parent for reproduction are selected probabilistically in proportion to their fitness, f_k, in order to make it more likely that fitter parents have offspring in an analog to survival of the fittest. A fitness function $f\left(k_j\right)$, squared difference from the mean, determines fitness by minimizing error over observations $X = \{x_1, \ldots, x_n\}$ for $\hat{f}(x)$. Other genetic program system parameters, such as population size and number of generations (Table 4), follow generally accepted practices (see Koza, 1992; Banzhaf et al., 1998).

< Table 4 here >

4 Results

How does land-use relate to driving forces posited by the conceptual model? The strength and direction of relationships between predictor variables and the agriculture response variable mediated by agent land-use strategies give an indication, as do the spatial residuals in accounting for spatial trends. Analysis of results takes place against the larger backdrop of evaluating LUCC models, a research area that on the whole remains underexamined (Verburg et al., 2005). This is especially so for complexity-based models, such as ABM and genetic programming, given their relatively recent use in LUCC modeling.

4.1 Symbolic regression

SYPRIA measures the relative importance of spatial factors in household decision making via a joint frequency and directionality analysis of terminals in the genetic programs that represent actor strategies. In every program, a given terminal t in terminal set $\boldsymbol{T}$ is labeled with a raw utility score $\hat{u}(t)$ when activated by the fitness test $f\left(k_j\right)$:

$$\hat{u}(t) = 1/d \tag{5}$$

where the utility $\hat{u}(t)$ of terminal t is inversely proportional to the depth d of the node, given the hierarchical, branching nature of the decision tree. Labels are aggregated across all nodes and programs to give a single value $\hat{U}(t)$ that is in turn normalized across $\boldsymbol{T}$ to give the value $U(t)$ for each terminal. This measure is not directly testable for significance with respect to variance, noisiness, or non-linearity in data in the way that Z values are assessed for statistical regression, although this is a topic of ongoing research.

Table 5 gives $U(t)$ for each terminal aggregated across the 3200 programs sampled, where each program represents the final multicriteria evaluation strategy for an agent. For institutional factors, likelihood of deforestation decreases the further a cell is from a road (–4.24) and increases with proximity to market (4.66) or a village (5.64). This is in keeping with expectations raised by theories of relative space, whereby distance to market and infrastructure is important to determining suitability of land for cultivation. The model finds a positive relationship between agriculture and both population density (3.16) and change in population density (4.48). This is expected given the predominance of extensive agriculture, the subsistence orientation of actors, and the relative abundance of land in ejidos. Open-tenure rules exert a strong influence (4.77) but the magnitude of forest protections (– 3.94) is tempered by the fact that, in reality, actors cultivate in areas to which they are not entitled (discussed below). Government support is positively related to agricultural land use but only weakly (1.67), which is likely due to 88% of the ejidatarios receiving PROCAMPO support, which makes this factor less important in the sense that it does not distinguish well between ejidatarios within individual ejidos, only between ejidos.

< Table 5 here >

In terms of environmental factors, the model agrees with the corollaries of spatial land-use theory with few exceptions. Agents uniformly ignored slope and aspect; further investigation is required to understand this apparent relationship, although likely explanations are that the scale of the DEM may be too coarse or that these characteristics affect extensive agriculture less than other styles of cultivation. As expected, agriculture is negatively related with elevation (– 1.88) given the effects on soil quality and landscape ruggedness. Agriculture is positively related to precipitation (1.63) and the soils dummy variable (3.67) that reflects soils posited as suitable for agriculture (the clay loams noted above). Per theories of relative space, distance to nearest agricultural land use (– 5.65) and secondary succession (– 6.90) is negatively related to deforestation because households pursue a shifting crop-fallow cycle within their ejidal plot assignments. Distance to lowland forest (5.44) and upland forest (– 3.46) also meets expectations because the latter is generally seen as being better for clearing for agriculture. In absolute spatial terms, extant agriculture (5.00), secondary succession (4.62), and upland forest (3.85) are positively related to agriculture occurring in the same location given the suitability of these land cover classes for agriculture, the effects of multi-year cropping within the crop-fallow cycle, and effects of long-term ejidal plot assignment. The model finds that fragmentation of surrounding land use and land cover is important (1.94 to 12.27), but further research is necessary to determine the relative importance of different kinds of fragmentation.

4.2 Spatial residuals

Examining spatial residuals lends further context to the use of genetic programs for representing LUCC decision making. Actor decisions are manifest in spatial patterns of land use in the simulated year 1995 that are validated against actual 1995 land-use data. Each of the 100 Monte Carlo runs produces a binary map of projected agriculture for all agents. Additively overlaying these Boolean layers and rescaling the resultant layer gives a composite map in which a zero value for a given cell indicates that none of the Boolean masks contributing to the composite map has agriculture in that cell while a value of one indicates the highest level of agreement that agriculture exists there.

General agreement between projected and actual land use is indicated by two measures (Gardner and Urban, 2003). The Kappa Index of Agreement (KIA) calculates how closely each binary categorical map compares to the 1995 reference image while accounting for chance. The Relative Operating Characteristic (ROC) compares the composite suitability layer to the reference layer. SYPRIA averaged KIA scores of 0.482 over the Monte Carlo runs and the composite scored a ROC of 0.905, both of which are good for a LUCC modeling exercise. The amount of actual land use in 1995 is used as a threshold value to denote which cells are considered agriculture in the composite layer in order to focus analysis on locational differences. Figure 4 is a crosstabulation image of projected agriculture given by the composite 1995 image vs. actual 1995 agriculture with three categories: 1) correctly projected non-agriculture in white; 2) incorrectly projected land state in dark gray; and 3) correctly projected agriculture in black. Figure 5 portrays a subregion of the study site at higher resolution with four categories: 1) correctly projected agriculture denoted in white; 2) correctly projected non-agriculture in light gray; 3) incorrectly projected non-agriculture in dark gray; and 4) incorrectly projected agriculture in black.

< Figure 4 here >

< Figure 5 here >

Several kinds of error are essentially inherent to a regional-scale model like SYPRIA yet still in keeping with the expectations of the conceptual framework. The most common error is randomly distributed patches that result from micro-scale processes that typically elude LUCC models (Point A in Figure 5). Errors also occur due to proximity to roads (Point B) or current agriculture (Point C). These small-scale errors are acceptable in the sense that they occur where actors choose the correct location in general but not specific cells in particular, especially when choosing plots spanning several grid cells.

The model could be better specified with respect to two kinds of error. One, errors are present where agriculture occurs outside of institutional bounds, such as in protected areas within the biosphere reserve (Point D in Figure 4). The model could conceivably account for these errors but only at the expense of potentially over-fitting the model with area-specific institutional rules. Two, large patches of under-projection and over-projection occur in several locations in the study area (e.g., Point E in Figure 5). This is due largely to SYPRIA placing restrictions on actor choice through a reduced suite of production activities. SYPRIA model configurations that incorporate greater institutional controls and persistence of commercial agriculture, for example, ameliorate these large-scale errors but that discussion lies beyond the scope of this article.

5 Discussion

SYPRIA is an example of integrated land-change science that supports the basic utility of the driving-forces conceptualization of LUCC and global environmental change. It also addresses theoretical and methodological needs in human-environment modeling, LUCC, and geographic information science. It does so by combining genetic programming, cellular models, and agent-based models in a GIS-based simulation framework. These techniques combined offer a way to invest individual actors with decision-making analogs that incorporate learning over time. They also afford the modeler the analytical tractability necessary to incorporate spatial and temporal interdependencies between agents and their environment through movement and decision making at a spatially local scale while simultaneously allowing examination of their net effects at larger scales. SYPRIA is one of a number of efforts that link ABM and cellular models to leverage these advantages in examining LUCC, but remains the only one to our knowledge that uses genetic programs as a decision making analog. Outstanding questions about complexity-based methods such as ABM, cellular models, and genetic programming remain, however, particularly with respect to: standardizing implementations; building better theories of their underlying mechanics; crafting calibration, verification, and validation regimes specific to these techniques; and closely examining the role of these methods in constructing theories of natural, human, and human-environment systems.

In addition to advancing modeling methodology, the model upholds basic precepts of relative and absolute theories of space with respect to both environmental and institutional factors posited to be of importance to household decision making through measures of the relative importance and directionality of spatial factors of terminals in the genetic programs that represent actor-agent decision-making strategies. The importance of both social and environmental factors within the context of actor decision making also supports use of the actor-institution-environment conceptual framework, itself a mirror of the foci identified by the larger LUCC and global environmental change research communities.

The SYPRIA modeling effort also highlights areas of further exploration, such as the need to distinguish among production technologies and to have more nuanced relationships between decision making and household characteristics. There is also a need for a greater focus on the environmental and institutional dynamics that are minimized by the relatively short period examined here. Longer time scales will also support expanded validation of the model given a degree of temporal dependence between calibration and validation data. With longer periods comes a greater need to examine error propagation, minimized at present by the relatively short periods examined. The SYPR project continues to develop econometric models that will support better specification of regional economic effects on household decision making. SYPRIA will also continue to be better integrated with ongoing work on actor-environment interaction over longer periods. Supporting these endeavors is a second round of actor interviews and remotely sensed imagery interpretation designed to give land-use and land-cover data for 1999/2000 and 2003/2004.

6 Conclusion

Land-use and land-cover change research is increasingly cast in the form of integrated land-change science—the explicit joining of ecological, social, and information sciences—through the means of land-change models as one of the few ways of combining various theories, methods, and data. Traditional modeling methods are joined by those from the complexity sciences and computational intelligence, such as genetic programming, cellular modeling, and agent-based modeling. Despite the need for ongoing research in the use of these methods, this application demonstrates the use of a computational intelligence technique, genetic programming, to model decision making in the context of human-environment relationships that are captured in turn by cellular modeling and agent-based modeling. It also reiterates the importance of considering the spatial and temporal dynamics of both social and environmental drivers of LUCC, in contrast to conceptual frameworks that favor sole drivers of land change—especially biophysical, demographic, or economic—in stylized equilibrium settings. Of particular importance to the application considered here is the decision making of change agents with respect to social drivers, particularly market institutions and land tenure, and ecological factors, especially characteristics of land use and land cover such as state, history, and fragmentation. While land change continues to offer a host of escalating challenges for local to global scale human-environment systems, the land-change research community is drawing on a long history of research to meet these challenges with continually evolving data, methods, and theories.

Acknowledgements

This work is supported by the National Aeronautics and Space Administration (NASA) Earth System Science Fellowship program (ESS 99-0000-0008) and a National Science Foundation Doctoral Dissertation Improvement grant (NSF 99-07952). It is also supported by NASA's Land-Cover and Land-Use Change program through the Southern Yucatán Peninsular Region project (NAG 56406 and NAG 511134) and the Center for Integrated Studies of Global Environmental Change, Carnegie Mellon University (NSF-SBR 95-21914). The author gratefully acknowledges the assistance of SYPR project personnel. Responsibility for the opinions expressed herein is solely that of the author.

References

Abizaid, C., Coomes, O.T., 2004. Land use and forest fallowing dynamics in seasonally dry tropical forests of the southern Yucatan Peninsula, Mexico. Land Use Policy 21, 71-84.

Agarwal, C., Green, G.L., Grove, J.M., Evans, T., Schweik, C., 2002. A Review and Assessment of Land-Use Change Models: Dynamics of Space, Time, and Human Choice. Center for the Study of Institutions Population and Environmental Change (Indiana University Bloomington Indiana) and the USDA Forest Service Northeastern Forest Research Station (Burlington Vermont).

Banzhaf, W., Nordin, P., Keller, R.E., Francone, F.D., 1998. Genetic Programming: An Introduction. Morgan Kauffman Publishers, San Francisco, California.

Bilsborrow, R.E., Carr, D.L., 2001. Population, agricultural land use and the environment in developing countries. In: Lee, D.R., Barrett, C.B. (Eds.), Tradeoffs and Synergies: Agricultural Intensification, Economic Development and the Environment. Commonwealth Agricultural Bureau International, London, pp. 35-55.

Bockstael, N.E., 1996. Modeling economics and space: the importance of a spatial perspective. American Journal of Agricultural Economics 78, 1168-1180.

Brenner, T., 1999. Computational Techniques for Modelling Learning in Economics. Kluwer Academic Publishers, Boston, Massachusetts.

Eastman, R., Jin, W., Kyem, P.A.K., Toledano, J., 1995. Raster procedures for multi-criteria/multi-objective decisions. Photogrammetric Engineering and Remote Sensing 61, 539-547.

Edmonds, B., 1999. Modelling bounded rationality in agent-based simulations using the evolution of mental models. In: Brenner, T. (Ed.), Computational Techniques for Modelling Learning in Economics. Kluwer Academic Publishers, Boston, Massachusetts, pp. 305-332.

Gardner, R.H., Urban, D.L., 2003. Model validation and testing: past lessons, present concerns, future prospects. In: Canham, C.D. (Ed.), Models in Ecosystem Science. Princeton University Press, Princeton, New Jersey, pp. 184-203.

Gimblett, H.R., 2002. Integrating Geographic Information Systems and Agent-Based Modeling Techniques for Simulation of Social and Ecological Processes. Oxford University Press, New York.

Gutman, G., Janetos, A., Justice, C., Moran, E., Mustard, J., Rindfuss, R., Skole, D., Turner, B.L., II., Janetos, A., Justice, C., Moran, E., Mustard, J., Rindfuss, R., Skole, D., 2004. Land Change Science: Observing, Monitoring, and Understanding Trajectories of Change on the Earth's Surface. Kluwer Academic Publishers, Dordrecht, Netherlands.

IGBP-IHDP, 1995. Land-Use and Land-Cover Change: Science/Research Plan. International Geosphere-Biosphere Programme and International Human Dimensions Programme, Stockholm.

Janssen, M., 2003. Complexity and Ecosystem Management: The Theory and Practice of Multi-Agent Approaches. Edward Elgar Publishers, Northampton, Massachusetts.

Keys, E., 2004. Chili cultivation in the southern Yucatan region: plant-pest disease as land degradation. Land Degradation and Development 15, 397-409.

Klepeis, P., 2003. Development policies and tropical deforestation in the southern Yucatán peninsula: centralized and decentralized approaches. Land Degradation & Development 14, 541-561.

Klepeis, P., Turner, B.L., II, 2001. Integrated land history and global change science: the example of the Southern Yucatan Peninsular Region project. Land Use Policy 18, 27-39.

Koza, J., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, Massachusetts.

Krzanowski, R., Raper, J., 2001. Spatial Evolutionary Modeling. Oxford University Press, Oxford, United Kingdom.

Lawrence, D., Foster, D.R., 2003. Changes in forest biomass, litter dynamics and soils following shifting cultivation in southern Mexico: an overview. Interciencia 27, 400-408.

Manson, S.M., 2001. Simplifying complexity: a review of complexity theory. Geoforum 32, 405-414.

Ostrom, E., 1999. Institutional rational choice: an assessment of the institutional analysis and development framework. In: Sabatier, P.A. (Ed.), Theories of the Policy Process. Westview Press, Boulder, Colorado, pp. 35-71.

Parker, D.C., Manson, S.M., Janssen, M., Hoffmann, M.J., Deadman, P.J., 2003. Multi-agent systems for the simulation of land use and land cover change: a review. Annals of the Association of American Geographers 93, 316-340.

Primack, R.D., Bray, D.B., Galletti, H.A., Ponciano, I., 1998. Timber, Tourists, and Temples: Conservation and Development in the Maya forest of Belize, Guatemala, and Mexico. Island Press, Washington, D.C.

Ralston, A., Rabinowitz, P., 2001. A First Course in Numerical Analysis. Dover Publications, New York.

Sierra, R., 2000. Dynamics and patterns of deforestation in the western Amazon: the Napo deforestation front, 1986-1996. Applied Geography 20, 1-16.

Steffen, W., Sanderson, A., Tyson, P., Jäger, J., Matson, P., Moore, B., III, Oldfield, F., Richardson, K., Schellnhuber, H.-J., Turner, B.L., II, Wasson, R., 2003. Global Change and the Earth System: A Planet Under Pressure. Springer-Verlag, Berlin.

Stern, P.C., Young, O.R., Drukman, D., Young, O.R., Drukman, D., 1992. Global Environmental Change: Understanding the Human Dimensions. National Academy Press, Washington, D.C.

Takeyama, M., Couclelis, H., 1997. Map dynamics: integrating cellular automata and GIS through Geo-Algebra. International Journal of Geographical Information Science 11, 73-91.

Turner, B.L., II, Geoghegan, J., Foster, D., 2004. Integrated Land-Change Science and Tropical Deforestation in the Southern Yucatán: Final Frontiers. Clarendon Press, Oxford, United Kingdom.

Verburg, P.H., Schota, P., Dijsta, M., Veldkamp, A., 2005. Land use change modelling: current practice and research priorities. GeoJournal (In press).

Tables

Table 1.
Regional land-use and land-cover change (1987-1995)

Land State	**Year**						**Change**	
	1987		***1992***		***1995***		***1987-1995***	
	ha	*%*	*ha*	*%*	*ha*	*%*	*ha*	*%*
Lowland Forest	306,685	17.8	306,530	17.8	299,403	17.4	-7282	-2.4
Upland Forest	1,201,979	69.8	1,170,013	68.0	1,095,391	63.6	-106,588	-8.9
Secondary Succession	121,247	7.0	112,969	6.6	212,176	12.3	90,929	75.0
Agriculture	85,864	5.0	124,992	7.3	106,544	6.2	20,679	24.1
Bracken Fern	458	0.0	1,837	0.1	2,919	0.2	2,462	537.7
Savanna	3,590	0.2	3,407	0.2	3,476	0.2	-114	-3.2
Water	2,049	0.1	2,052	0.1	2,047	0.1	-2	-0.1

Source: SYPR project

Table 2.
Soil fertility modifiers for cellular model for agricultural model component

Situation in previous time period (Q^{t-1})	Fertility in current time period (Q^{t})
Agriculture – Milpa (s^{Am})	If S^{t-1} == s^{Am} then Q^{t} – – 2
Agriculture – Commercial (s^{Ac})	If S^{t-1} == s^{Ac} then Q^{t} – – 3 unless fertilized
Pasture (s^{P})	If S^{t-1} == s^{P} then Q^{t} – – 1
Fallow (s^{F3})	If S^{t-1} == s^{F3} then Q^{t} + + 1

Notes:

==	Equal to
++x	Add (x) to current value of Q
– –x	Subtract (x) from current value of Q
s^{F3}	Cell state = secondary succession
s^{Am}	Cell state = milpa
s^{Ac}	Cell state = commercial agriculture
s^{P}	Cell state = pasture

Base values: Lowland soil (Q^{t} = 10) and Upland soil (Q^{t} = 12)

Table 3.
Land-use and land-cover transition rules for environmental model component

State in last time period (S^{t-1})	State in current time period (S^t)	
Water (s^W)	If	$S^{t-1} == s^W$
	then	$S^t = s^W$ and $t^* = 0$
Savanna (s^S)	If	$S^{t-1} == s^S \bullet S^t \sim(s^A ˇ s^P)$
	then	$S^t = s^S$ and t^*++ 1
Lowland Forest (s^{FL})	If	$S^{t-1} == s^S \bullet S^t \sim(s^A ˇ s^P)$
	then	$S^t = s^{FL}$ and t^*++ 1
Upland Forest (s^{FU})	If	$S^{t-1} == s^S \bullet S^t \sim(s^A ˇ s^P)$
	then	$S^t = s^{FU}$ and t^*++ 1
Bracken Fern (s^{F0})	If	$S^{t-1} == s^{F0} \bullet S^t \sim(s^A ˇ s^P)$
	then	$S^t = s^{F0}$ and t^*++ 1
Secondary Succession of 5, 10, or 25 years (s^{F1}, s^{F2}, or s^{F3})	If	$S^{t-1} == s^F \bullet S^t \sim(s^A ˇ s^P) \bullet t^*$++ 1 $< \beta$
	then	$S^t = s^{F\text{ (Current class)}}$
	else	$S^t = s^{F\text{ (Next class)}}$ and $t^* = 0$
	where	β = 5, 10, or 25 threshold for time in s^{F1}, s^{F2}, or s^{F3}
Agriculture (s^A)	If	$S^{t-1} == s^A \bullet S^t \sim(s^A ˇ s^P)$
		$\bullet$ (If ($N^{Count}[s^{F0}] > \beta_1$) ˇ $Q > \beta_2$)
	then	$S^t = s^{F0}$ and $t^* = 0$
	else	$S^t = s^{F1}$ and $t^* = t^* \times \beta_3(N^{Count}[s^{F?}])$
	where	$\beta_1 = 3$ (threshold for s^{F0} neighbors)
		$\beta_2 = 2$ (soil fertility threshold)
		$\beta_3 = 0.3$ (constant for seed availability from $s^{F?}$)
Pasture (s^P)	If	$S^{t-1} == s^P \bullet S^t \sim(s^A ˇ s^P)$
	then	$S^t = s^{F1}$ and $t^* = 0$

Notes:
Logical operators: Not (~); And (•); Or (ˇ); Equal to (==); Make equal to (=)
Cellular operators: Majority (N^{Major}); Counting (N^{Count})
Other: Add X to current value (++X); Time since last transition (t^*); Wildcard (?)

Table 4.
SYPRIA genetic program parameter settings

Parameter	**Value**
Population size (M): programs in the population	300
Generations (G): generations over which programs evolve	50
Crossover probability (P_c): probability that a program will be the offspring of two other programs	0.9
Mutation probability (P_m): probability that a program mutates	0.001
Reproduction probability (P_r): probability that a program will be the offspring of one parent program	$1 – (P_c – P_m)$
Selection type: process by which parents are selected	Tournament
Tournament size	5
Crossover: probability that a program will be crossed or mutated at a function	0.7
Creation depth: maximum depth of programs when created	6
Crossover depth: maximum depth at which programs cross	17
Randomness: probability of choosing an ephemeral constant	0.5
Program length: maximum genetic program length	500

Table 5.
Genetic program frequency and direction analysis of key factors

Factor	$U(t)$
Environmental	
Soils	3.67
Precipitation	1.63
Elevation	– 1.88
Slope	0.01
Aspect	0.01
Institutional	
Road cost	– 4.24
Market proximity	4.66
Village proximity	5.64
Tenure (Open)	4.77
Tenure (Forest)	– 3.94
Direct payments	1.67
Population density (1980)	3.16
Δ (1980-1990)	4.48
Land use/cover	
Lowland forest	– 3.92
Upland forest	3.85
Secondary Succession	4.62
Agriculture	5.00
Land (distance to)	
Lowland forest	5.44
Upland forest	– 3.46
Secondary Succession	– 6.90
Agriculture	– 5.65
Fragmentation	
Lowland forest	1.94
Upland forest	3.23
Secondary Succession	3.99
Agriculture	12.27

Figures

Figure 1. Southern Yucatán Peninsular Region

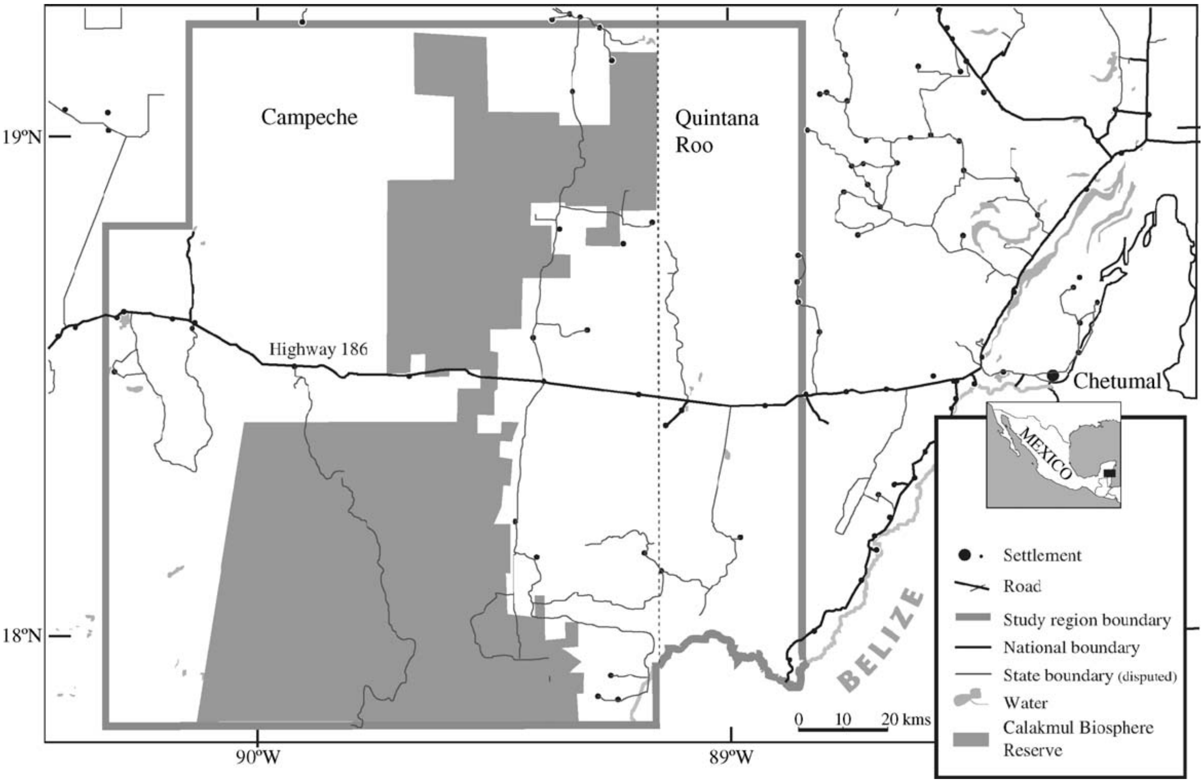

Figure 2. Ejidos and actor placement within study region

Figure 3. Genetic programming system with steady state population

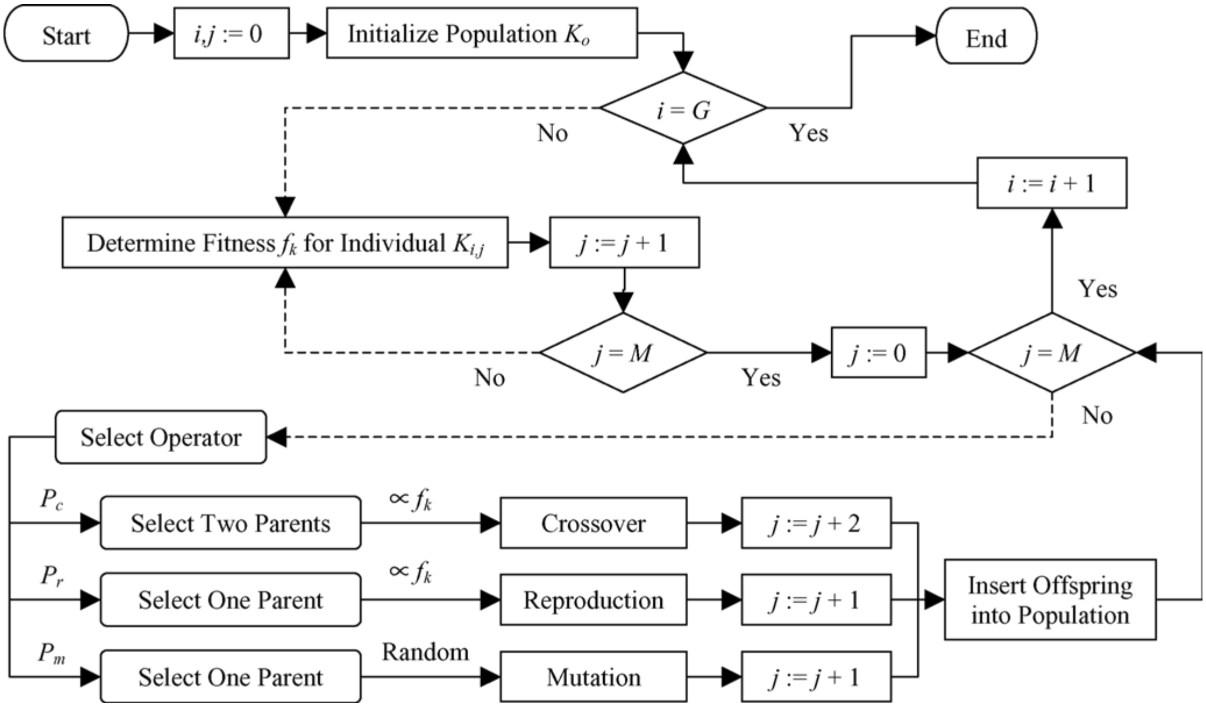

Figure 4. Spatial residuals for projected vs. actual agriculture demonstrating correctly and incorrectly projected agriculture and non-agriculture (1995)

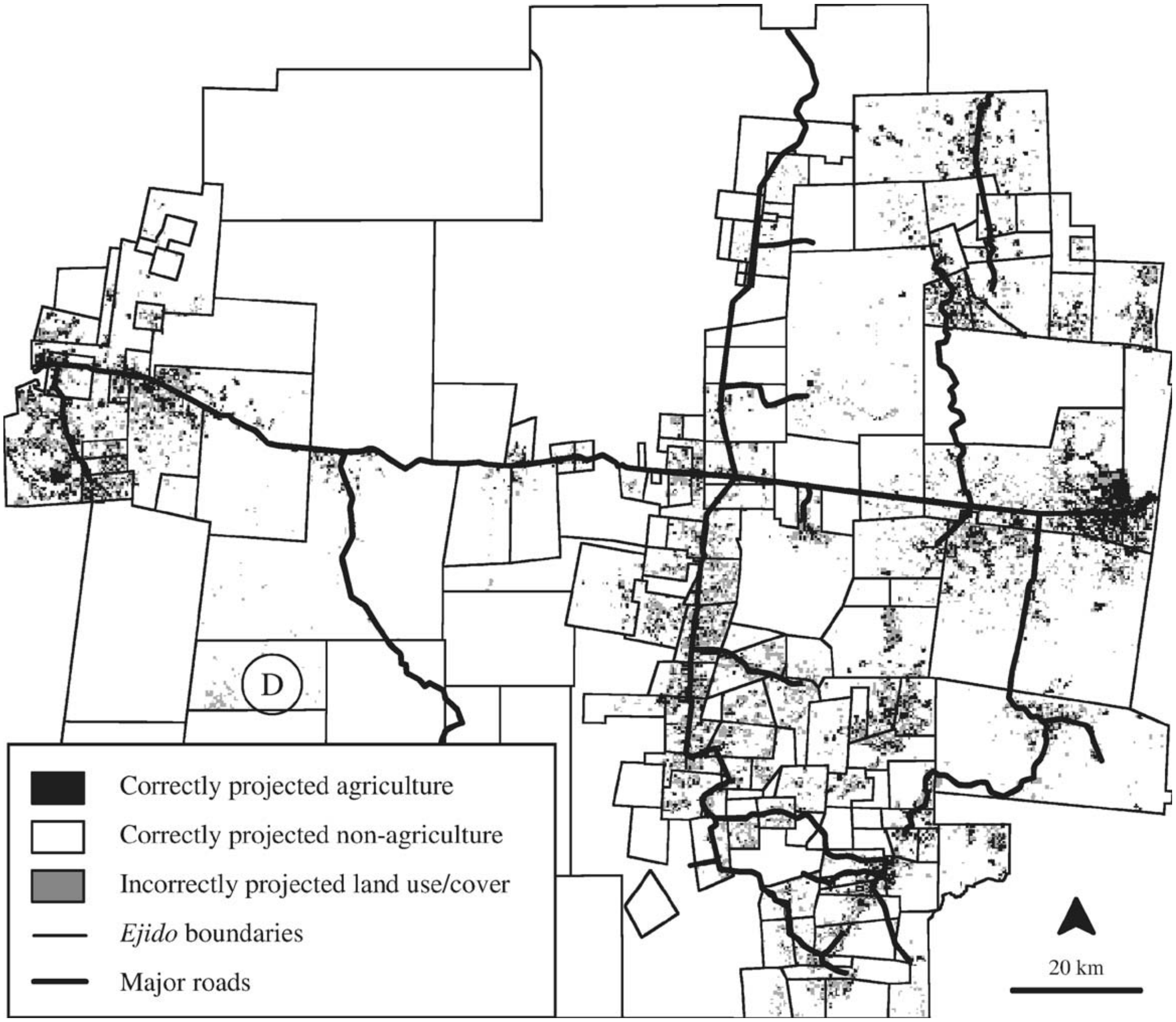

Figure 5. Spatial residuals for projected vs. actual agriculture demonstrating correctly and incorrectly projected agriculture and non-agriculture (1995), inset