Progress Report

Agent-based model for assessing ex-ante impacts of farm management innovations on performance of GL/DC-based smallholder systems: Design framework, theoretical parameterization

Cairo, December 2019

Progress Report

Agent-based model for assessing ex-ante impacts of farm management innovations on performance of GL/DC-based smallholder systems: Design framework, theoretical parameterization

Quang Bao Le[1]*

[1] Resilient Agricultural Livelihood Systems Program (RALSP), International Center for Agricultural Research in the Dry Areas (ICARDA), 2 Port Said, Victoria Sq., Ismail El-Shaaer Building, Maadi, Cairo, Egypt. Phone: +202 2359 1138. E-mail: Q.Le@cgiar.org

* Corresponding

CAIRO, DECEMBER 2019

CITATION:

Le, Q.B. (2019). Agent-based model for assessing ex-ante impacts of farm management innovations on performance of GL/DC-based smallholder systems - Design framework and theoretical parameterization. Progress Report. International Center for Agricultural Research in the Dry Areas (ICARDA) and CGIAR Research Program on Grain Legumes and Dry Cereals (CRP GLDC). Cairo, Egypt.

Table of Contents

Abstract

Viable management and policy options for sustaining smallholder farming systems in grain legume and dry cereal production regions need special attention. Although a great deal of knowledge on ways to efficiently agronomic measures exists, too few studies seek to understand how agricultural policy, financial services, farming technologies, local capabilities interactively affect smallholders' decision about farming system management. As a methodological opportunity, multi-agent system (MAS) or agent-based model (ABM) has been recently recognized as a promising approach for explaining complex human-environment interactions in agroecosystems. This report presents the concept, framework and theoretical parameterization of a MAS/ABM for the typical coupled community-landscape system that can be used for ex-ante assessment of long-term impacts of management and policy options on soil fertility, food productivity and profitability of smallholder agroecosystems in different geographic regions. The goal is to provide insights into appropriate strategies for promoting the viability of smallholder agricultural livelihoods over the long term.

Keywords: Agricultural livelihood system, socio-ecological system, complex system, grain legumes, dry cereals, multi-agent system, agent-based modelling, decision making

1. Needs and methodological opportunities in modelling research for supporting sustainable agricultural livelihood system

1.1. A shift to socio-ecological system investigations with a focus on decision making

Given that human and environmental dimensions of agricultural livelihood system are inextricably intertwined, efforts to promote sustainable system management needs to examine the socio-ecological system (SES) at hand. Farmers' decisions are affected by a portfolio of political, social, economic and biophysical driving forces. Changes in farming patterns, management and practice affect soil fertility, food productivity and profitability. Accumulation of these effects over space and time will in the long-term lead to changes in the greater social-ecological landscape that in turn reshape decisions of smallholders and other involved actors. Recognizing that substantial socio-ecological systems relationships, researchers have been advocating for more interdisciplinary research about nutrient recycling and management issues that explores interactions between different components of farm system, such as livestock, crops, and soil, and also human actors' decision making at different hierarchical levels, such as policy makers, fertilizer suppliers, and farmers (Craswell *et al.*, 2004) (Matthews and Selman, 2006); (Vitousek *et al.*, 2009). Attempts to transform these smallholders from the current difficulties to more sustainability require the consideration of not only technical nutrient management options, but as well the social and policy framing conditions that are beyond of the farm domain. Such interactions explain for ecosystem regime shifts, thus must be focused in diagnosing problems of human-environment systems (Scheffer and Carpenter, 2003; Foley *et al.*, 2011).

Although a great deal of agronomic knowledge exists, so far not many studies seek to assess combining effects of different conditions/drivers (e.g. technological intervention, agricultural policy, financial and extension services, environmental conditions and farmer profile) on their decision on livelihood strategy and farm management. Improved understanding of farmers' decisions will provide insight into what types of appropriate coping strategies might be taken at the local and regional scale. At the same time, assessment of impacts that changing farm management and practices have on the environment and on overall productivity and profitability is also crucial (Schlecht and Hiernaux, 2004).

1.2 The need for a conceptual framework for investigations of complex social-ecological interactions in smallholder agricultural livelihood system

An increasing number of studies are integrating social and natural sciences (Belcher, 2004; Matthews and Selman, 2006; Gaube *et al.*, 2009). These interdisciplinary efforts integrate tools and techniques from ecological and social sciences, such as geographic information systems (Grimm *et al.*, 2006), system modelling and simulation, and survey research, to anticipate long-term outcomes, identify threshold points for agro-ecosystem regime shifts, pinpoint system feedback loops and time lags, and characterize numerous non-linear human-environment relationships (Liu *et al.*, 2007).

A major challenge with human-environment investigations is to effectively integrate knowledge (epistemic) and methods of various scientific disciplines, and from different societal domains. As suggested by (Scholz *et al.*, 2011), a conceptual framework for considering socio-ecological system in-transition, such as smallholder agricultural livelihood systems, would be at least based on the following postulates: human and environment sub-systems are coupled; system hierarchies exist that often experience interferences; feedback loops, human decision-making and human awareness of their biophysical and social environments are parts of the system.

1.3 Multi-agent systems (MAS) as a methodological opportunity for modelling social-ecological interactions and system transition with a focus on adaptive decision/behaviours

There is an emerging recognition in the scientific community of the capability and utility that multi-agent simulations (MAS), or agent-based models (ABM) offer for understanding the complexity of energy, nutrient and material flows that result from rich interactions and feedback among social and natural processes (Bousquet and Le Page, 2004; Gaube *et al.*, 2009). In an MAS/ABM model, the human-environmental system is described through autonomous 'agents', which can be defined to represent human actors such as a farm household, a company, or biophysical entities such as a crop field. Compared to other modelling approaches, MAS/ABM can better support interdisciplinary, long-term, multi-level, participatory studies with prominent uncertainties (Boulanger and Brechet, 2005).

Because it can incorporate the heterogeneity of social and ecological components, MAS represent individual behaviors, enabling, for example, simulation of collective use of common resources.

Great progress has been made in applying MAS to study land use change over the last 10 years. Recent versions of some MAS models are being used to examine specific real-world questions that can inform, for example, land use policy decisions ((Le *et al.*, 2008; Gaube *et al.*, 2009; Le *et al.*, 2010), technology diffusion (Schreinemachers and Berger, 2011). Le et al. (2005, 2008, 2010, 2012) developed LUDAS (Land-Use Dynamics Simulator), an agent-based model to anticipate outcomes of alternative land use policies in in forest margins and agrarian community-landscape systems, offering one of the most integrated couplings of human drivers and natural constraints to date. Gaube et al. (2009) employed agent-based modeling and stakeholder involvement in the Reichraming area of Upper Austria to assess outcomes of alternative future land use scenarios, and demonstrate innovative scenario development. Berger et al. (2010) introduced MP-MAS (Mathematical Programming-based Multi-Agent System) as a tool for simulating sustainable resource use in agriculture and forestry, and uses mathematically programming to simulate human decision-making. Matthews (2006) reported People And Landscape Model (PALM) that simulates water, carbon, nitrogen, labour, and financial flows through a rural subsistence community using agent-based modelling.

However, the MAS/ABM model is still in an early stage of development for smallholder agricultural livelihood systems: the farmers' decision-making is modelled in a too simple way that has not yet sufficiently supported system transition options. Still, long-term and secondary feedback loops are often not adequately conceptualized and modelled.

1.4 The need of multi-stakeholder involvement in SES modelling research

To allow for well-informed modelling of the human agent (e.g. farmers decision on P use), close interaction with respective real-world stakeholders is indispensable: firstly to gain insights into their decision behaviour, secondly to verify the developed model and thirdly to allow for a potential adaption of their action given their exposure to the results of the modeling exercise (Voinov and Bousquet, 2010; Le *et al.*, 2012). Stakeholder involvement in a research process must be guided by functionality, i.e. level of involvement (information, consultation, collaboration, and empowerment) must appropriately fit the issue, step in the process, and concrete project task (Krütli *et al.*, 2010). Transdisciplinarity (Td), a protocol of processes for collaborative science-society research, offers ways to engage stakeholders in studying human-environment systems, in integrating layperson and scientific knowledge, to improve environmental awareness and learning about the issue, as well as support to policy decisions (Scholz, 2011). Experts agree that stakeholders involvement and input leads to more sound, more acceptable policy decisions. The scientific challenge includes how to organize multi-stakeholder dialogues (transdisciplinary discourses) and mutual learning along the model development process (a) to improve the contextual validity of the model and (b) to increase system understanding of key stakeholder groups that need for developing rational decisions.

2. Conceptual MAS/ABM model design

We develop a MAS/ABM by adapt the LUDAS (Land Use DynAmic Simulator) to the GL/DC production landscapes in SSA. The detailed description and theoretical specification of LUDAS model are shown in Le et al. (2008, 2010, 2012) (first version), Villamor et al. (2014) and Miyasaka et al. (2017) (other variants). We describe systematically the concept, structure and detailed specification of the model using the ODD (Overview, Design concepts, and Details), a standard protocol for describing agent-based models (Grimm et al., 2006). ODD has been more widely applied to ensure the descriptions of such complex models are readable and complete for the purposes of understanding the model application and replication (Müller *et al.*, 2014). The MAS/ABM model described here is named as GLDC-LUDAS. The model description along seven elements of the ODD protocol is as follows.

2.1. Purpose

Primarily, the GLDC-LUDAS model is designed to support management and policy options for smallholder GL/DC-based community-landscape based on its capabilities as follows:

- To explore the magnitude of possible socio-ecological changes over space and time driven from different management/policy interventions,
- To identify the most affected system's components (*what*), locations (*where*), actor groups (*who*) and periods (*when*) with respect to specific policy interventions,

- To highlight sound management/policy interventions that likely enhance environmental, socio-economic benefit at least cost in a long run, and
- To explore the potential trade-offs and synergies of the management/policy interventions over different objectives and social groups.

2.2. Agents, their state variables and scales

At agent scale:

GLDC-LUDAS comprises a human-environment landscape consisting of (1) human agents that represent farming households, and (2) landscape agents that are congruent autonomous land pixels (Figure 1, the lower part).

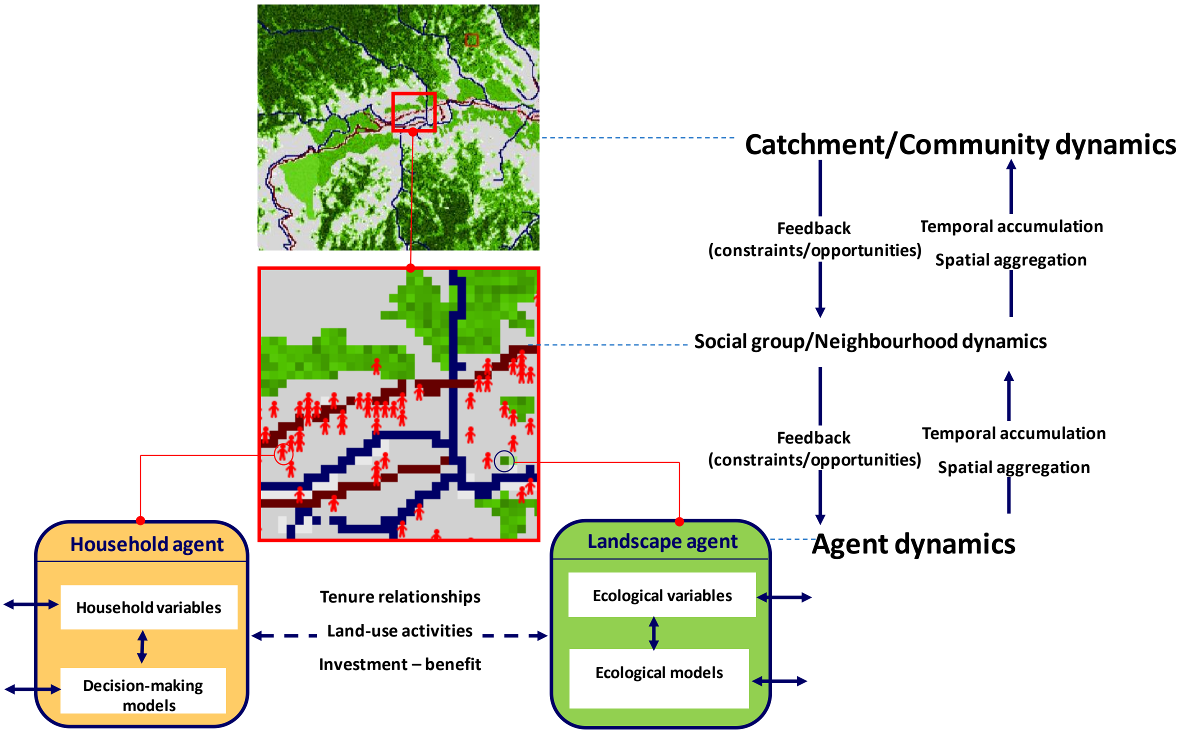


Figure 4. Agents, its variables and sub-models, interactions and feedbacks over across scales presented in GLDC-LUDAS model. Source: Adapted from Le et al. (2010).

The state variables of these agents are as given below.

(1) Human agents: The state variables of these agents capture the sustainable livelihood capitals of each household. This includes social identity (or simply the identification number), age, group membership, and human resources (e.g., household size, dependency ratio and education), land and natural resources (e.g., land holdings and land structures), financial capital (e.g., gross income and gross income per capita),

physical capital (e.g., access to market and distance to town), and households' accessibilities to certain rural services (e.g. extension services, credit schemes), agricultural institutions, and policy (e.g. subsidies of farming inputs) (see Figure 2).

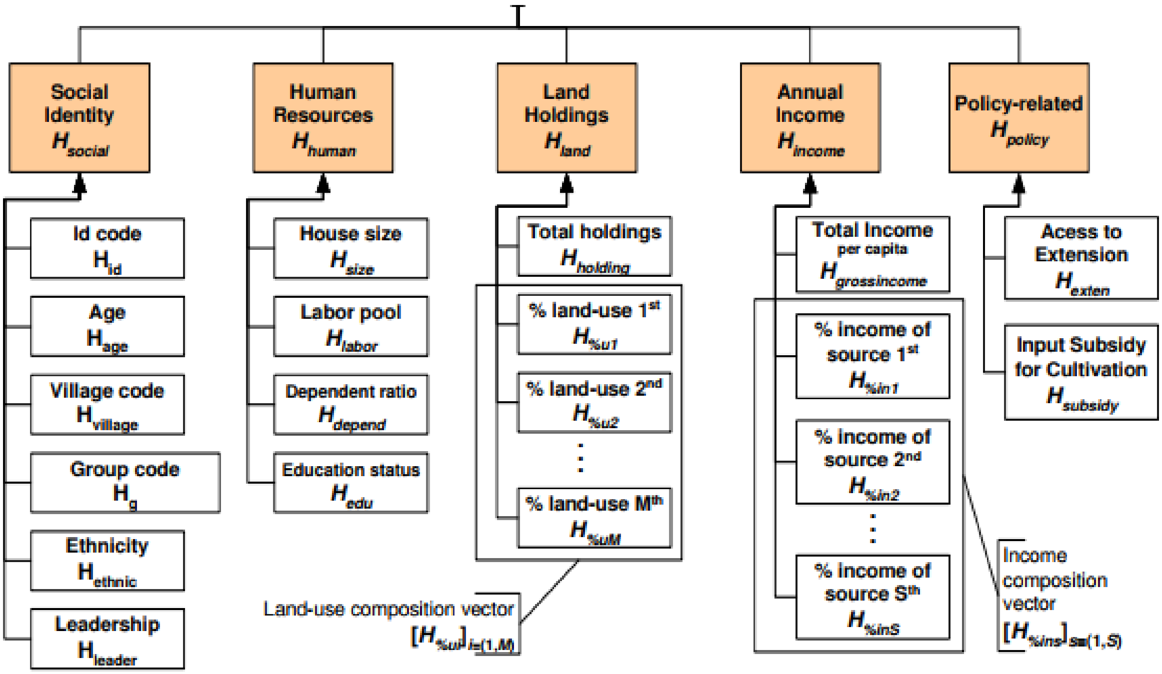


Figure 2. Diagram shows state variables of household agents. Source: adapted from Le (2005).

(2) Landscape agents: State variables of landscape agents are corresponding to GIS-raster layers of biophysical spatial-variables (e.g. land cover, topographical attributes, soil quality level, cropping type and responsive yield), economical spatial-variables (e.g. proximate distance to roads), institutional spatial-variables (e.g. owner, village territory, protection zoning class), and histories of particular patch properties (see Figure 3).

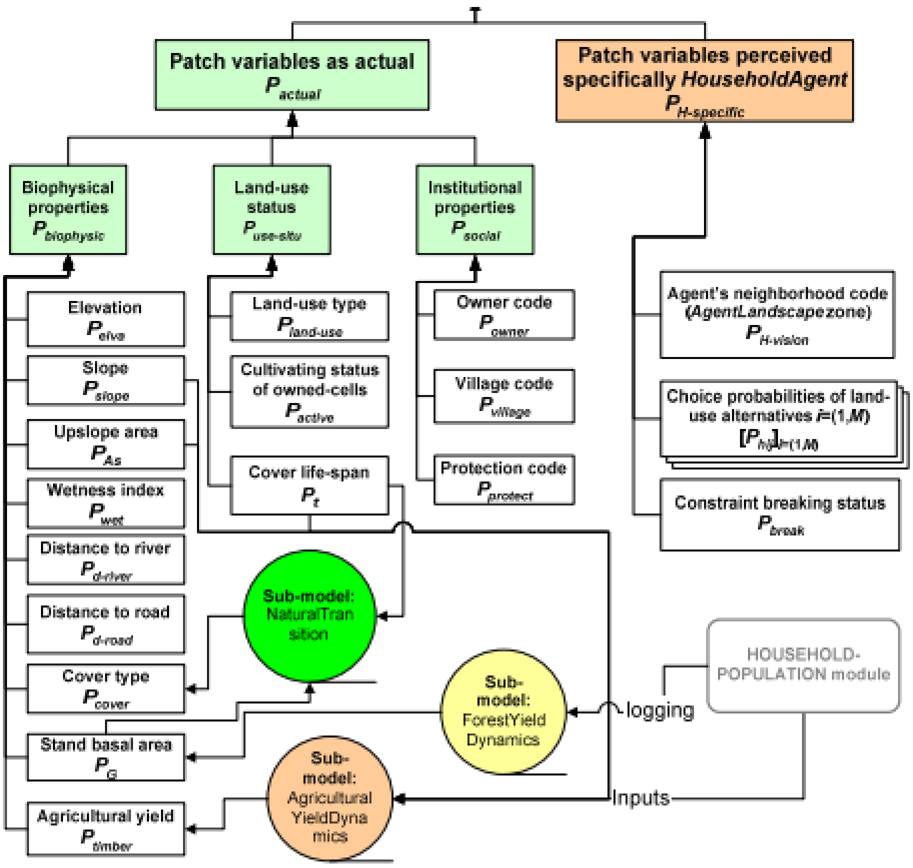


Figure 3. Diagram shows state variables of landscape agents (patch variables). Source: adapted from Le (2005)

Regarding their dynamics, there are two types of state variables of agents: (1) static variables and (2) dynamic variables.

(1) Static variables, including some landscape variables such as topographical attributes, proximities/distances to the nearest town and road.

(2) The dynamic variables have two sub-types:

- dynamic variables driven by natural process that beyond human's control, such as the age of the household agents and natural vegetation growth of naturally vegetative pixels,
- dynamic variables induced by household decisions or policy interventions, such as land holdings and income of household agents; cover type, crop yield and protection code of land pixel. Dynamic variables can act simultaneously as causes for some particular processes and consequences from other processes.

In GLDC-LUDAS, because the behavioral strategy of a household agent can changes over time, parameters specifying household behavior are also treated as state variables characterizing household behavior and stored in the memory of household agents. These behavioral variables include:

(1) a set of preference coefficients (or weights) reflecting the relative importance of various environmental, socio-economic and policy factors in household decision about land uses, and
(2) a set of ratios determining the amount of labor allocated for each branch of livelihood activities.

At agent groups' scale (i.e. landscape neighborhoods and household groups) and whole population-landscape scale:

State variables for the entire (or sub-sets) of landscape and human community are emerged from interactions between human and landscape agents. Since neighborhood interactions are taken into account in the modelling of farming decision and farm type transition, changes in landscape and/or community status feedback to household- and pixel-based processes (Figure 1, the right part).

2.3. Process overview and scheduling

In a GLDC-LUDAS run, the coupled human-environment system is annually successive. The initial grid-based landscape is given by GIS raster files of corresponding variables. The initial population is generated from data of sampled households. For each time step, sequential tasks are done within the annual production cycle (Fig. 4), in which household decisions are linked with concurrent processes in landscape agents. In most cases, all household and landscape agents are called upon on perform tasks in parallel (i.e. synchronizing actions). The current version of GLDC-LUDAS is coded using NetLogo 6.0 (Wilenski, 1999).

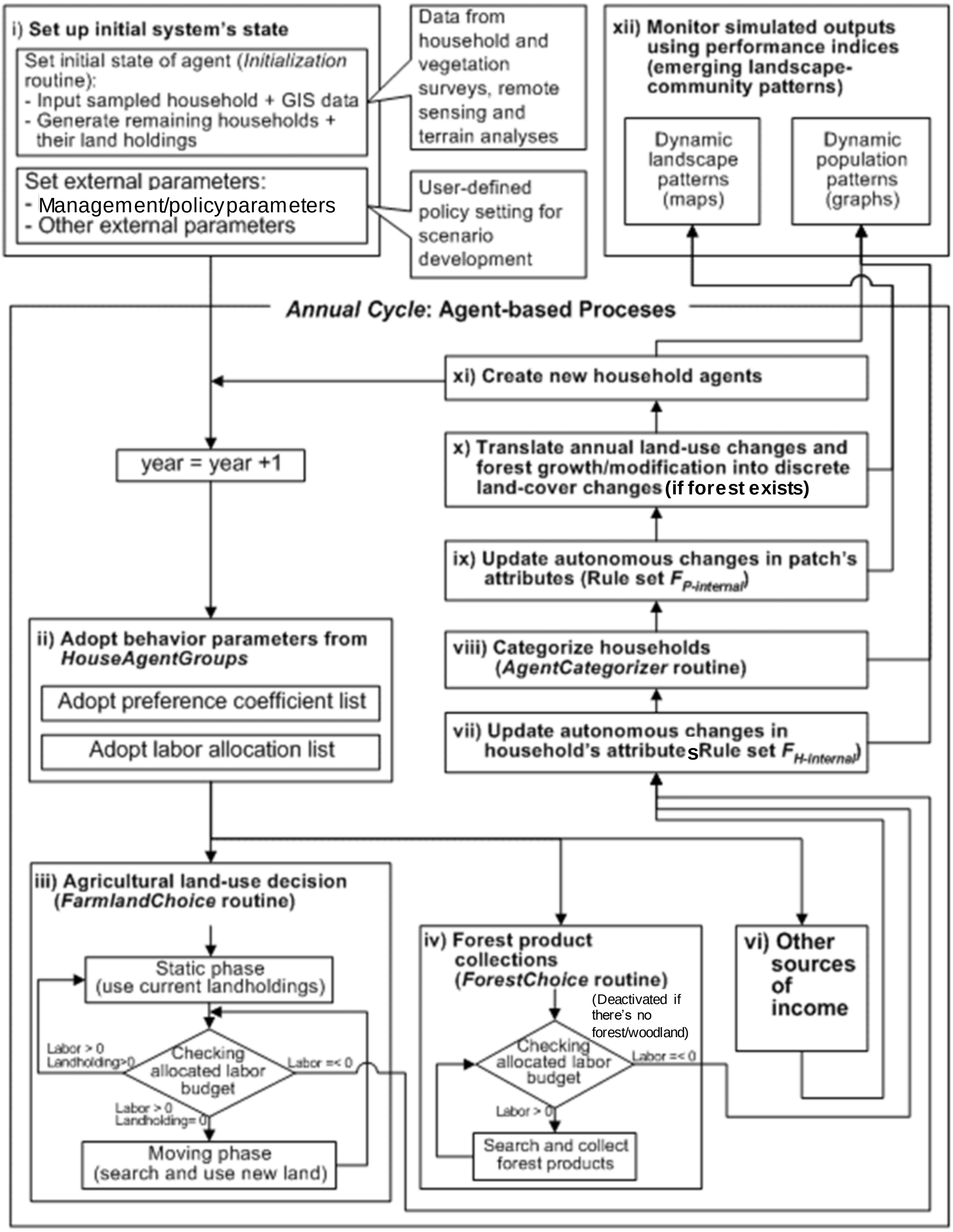


Figure 4. Flow chart showing main steps of the simulation/scheduling program in GLDC-LUDAS (Source: modified from Le et al. (2008, 2010)).

2.4. Design concepts

The modelled units in GLDC-LUDAS are landscape and household agents. Landscape agents are regular land grids with their own attributes and ecological response mechanisms to environmental changes and human interventions. Household agents have their own state and decision-making mechanisms about land uses.

Emergence

Livelihood performance of the entire household population or social groups (e.g. average household income, income structure and inequality) truly emerges from household's decisions that integrate household characteristics, surrounding dynamic environment and policy information (Fig. 1, the right part). However, population dynamics are modelled by an empirical equation estimated from historical data, thus are not an emergence property from micro interactions. In general, agent's interactions generate community and landscape changes and such macro changes create new opportunities or constraints for the agent-based processes, thus forming cross-scale feedback loops (Fig. 2, the right part).

Adaptation

The GLDC-LUDAS model includes human adaptation mainly in the land-use decision process and the change of behavioral strategy.

(1) Household agents adapt to current socio-ecological condition by choosing the best land-use in the best location in term of utility.
(2) Household agents can changes their behavior strategy (i.e. structure of labor allocation and preference coefficients in land-use/crop/management choice functions) by imitating the strategy of the household group who is most similar to it.

Fitness (goal orientation)

Goal-seeking in land-use decisions by household agents is explicitly modelled, in which households calculate utilities - expressed in a probability term - for all land-use and location alternatives and "likely" select the alternative with the highest utility. However, by applying an ordered choice algorithm described in Benenson and Torrens (2004), and Le et al. (2008), concrete household decisions in GLDC-LUDAS are bounded-rational rather than purely rational. This bounded optimization holds the risk that some household agents select a land-use/crop/management type that may not be the best alternative, but the chance for choosing the best alternative is high.

Interactions

In GLDC-LUDAS, agents interact indirectly or directly. Indirect interactions among household agents involve the fact that land-use conversions caused by households can lead to changes in the decision space of other agents in the next time step. Household agents interact directly when neighbor households also look for their best land-use alternative in the same location. In this case, in a random manner one of them will have to leave that location and search for another place.

Stochasticity

GLDC-LUDAS uses stochasticity to in the following processes:

(i) Initializing household population,

(ii) searching the locations of the landholding of households generated in the system initialization (framed by land-use polygons) or newly "born" during the simulation period,

(iii) generating the preference coefficients of land-use choice functions that are around empirically estimated values and bounded by confidence intervals, and

(iv) generating some status variables not affected by agent-based process (all defined by even distribution and predefined bounds).

Observations includes annually successive maps of land use/cover, agricultural yield and land holdings; and graphs that describe temporal patterns of land-use/cover coverage (calculated based on the whole or partial landscape), average farm size and income (mean, composition and equality).

2.5. Initialization

The initial landscape of the model is deterministically given by importing GIS raster files of landscape variables that are either secondary data or produced by separate spatial analyses. The model has a deterministic function to create protection zone in according to the zoning parameter defined by users.

The initialization of the household population includes the following sequential steps.

(1) Data of a household sample (N_s households) are imported and users set the size of the total population (N_t households).

(2) The regeneration of the remaining fraction of the total population is based on this true equation: $N_t = Ns \times int(N_t/N_s) + mod(N_t/N_s)$, where $int(N_t/N_s)$ is the integer part of the ratio N_t/N_s and $mod(N_t/N_s)$ is the remainder after N_t is divided by N_s. Assuming that the sample (Ns) and the total population (N_t) have the same distributions of the status variables, household subset $Ns \times int(N_t/N_s)$ (the majority) is exactly generated by multiplying the household sample by the integer component of the ratio N_t/N_s. Household subset $mod(N_t/N_s)$ is generated by a random selection of households from the sample (N_s).

(3) The last step is the creating of land parcels hold by the newly generated households using spatially bounded random rules. According to these rules, given a state variable representing the number of land parcels (with explicit land-use types) of new generated households, the corresponding locations of such parcels are randomly selected among the pixel set bounded by the polygons of the household's village territory and the corresponding land-use types.

3.6. Inputs

Inputs for simulations with GLDC-LUDAS include two types: data and parameters.

Calibrated data

In LUDAS, data for initializing the coupled human-landscape system include GIS data in forms of text files, and household data as worksheets. Because of being path-dependent, behavior of the complex human-environment systems is sensible to the initial state of the system. Therefore, the better the quality of input data is, the higher creditability the model's outcomes have. All data used by the GLDC-LUDAS are calibrated and/or processed in separate studies to adequately represent the reality of the coupled human landscape system.

As human-environment interactions are modelled at pixel and household levels, the linkages between the spatial and household datasets are crucial in GLDC-LUDAS. Both human (household) and landscape (land parcel) agents have a shared identical variable indicating the household-pixel link.

Parameters

Within the GLGC-LUDAS, two types of parameters are distinguished: calibrated parameters and user's defined parameters. Most calibrated parameters are coefficients representing behaviors of household and landscape agents, such as ratios of labor allocation to livelihood activities and preference coefficients of land-use choice functions. These parameters are estimated from sample data to reflect the real world. User's input parameters are mainly policy parameters, which enable users to set their own policy options for development of land-use change scenarios.

2.7. Sub-models

Main sub-models and calculation routines in GLDC-LUDAS are summarized in Table 1. Important sub-models that substantially constitute complex human-environment interactions and adaptation are *FarmlandChoice*, *ForestChoice*, *AgriculturalYieldDynamics*, *AgentCategorizer* and *NaturalTransition*. Sub-models *ForestChoice* and *NaturalTransition* exist in the model code, but activated only if the study landscape contains considerable areas of natural vegetations such as forest/woodland and shrubland. Detail descriptions of these sub-models and routines are shown in Le et al. (2008).

Table 1. Main sub-models/procedures of GLDC-LUDAS

Name	Brief functionalities/tasks (references for more detailed information)	Involved agent (HA: human/house hold agent, LA: landscape agent)
Initialization [a]	Import GIS data and sampled household data, generate the remain population, create household-pixel links (see Section 2.5 of this report)	HA LA
SetLaborBudget	Annually set the labor list of the household (See Appendix 1)	HA

FarmlandChoice [a]	Perform agricultural land-use choices, including bounded-rational choice, nested with rule-based decision algorithms (see Appendix 2)	HA and LA
ForestChoice [a]	Perform forest-use choices, mainly rule-based algorithms (activated only if there are woody/forest vegetation areas in the study landscape)	HA and LA
GenerateOtherIncome	Generate non-crop and non-timber incomes	HA
UpdateHouseholdState	Annually update changes in household profile	HA
AgentCategorizer	Annually categorize household into the most similar group (see the next sub-section	HA
GenerateHouseholdCoefficients	Generate behaviour coefficients of household, allow variants within groups, but stabilize behaviour structure of the group	HA
AgriculturalYieldDynamics	Empirical production functions that calculate the economic yield of farmlands in response to human investment (e.g. labor, agrochemicals) and site conditions (e.g. slope, soil conditions) (see Appendix 3).	HA and LA
ForestYieldDynamics	Calculate forest stand basal area in response to human interventions (activated only if there are natural vegetation areas in the study landscape)	LA
NaturalTransition	Perform natural transition among vegetation types based on accumulated vegetation growth and ecological edge effects (activated only if there are natural vegetation areas in the study landscape)	LA
CreateNewHousehold	Create a young new household, controlled by an empirical function of population growth	HA

DrawGrapths	Draw different graphs of system performance indicators	HA and LA

[a] Complex procedure: procedure that contains one or more other procedures.

Imitative learning, livelihood typology and AgentCategorizer sub-model

A fundamental principle of imitation is that the process is facilitated by favors some similarity between the imitator and the group to be imitated. In agricultural land use, Schmit and Rounsevell (2012) suggested a highest probability of imitation between farmers of a *similar typology*. For instance, a farmer specialized in field cropping is more likely to imitate a farmer with the same typology rather than someone specialized in livestock grazing. It is possible that a potentially imitating farmer would assess the extent to which a 'model' farmer's situation is similar to his own in order to determine how valuable imitation would be (Polhill *et al.*, 2001; Gotts and Polhill, 2009; Le *et al.*, 2012).

We used the Sustainable Livelihood Framework (SLF) concept (Ashley and Carney, 1999) for selecting criteria that represent the livelihood typology of households, and incorporating a livelihood similarity comparison component in the model. The SLF includes five core asset categories: human, social, financial, natural and physical capital (Ashley and Carney, 1999). This spectrum of livelihood assets is the basis of people's capacity to generate new activities in response to needs and opportunities. The concept forms a theoretical basis for deriving indicators for multi-dimensional assessment of the livelihood performance and similarity, helping to avoid bias selection of indicators from one particular discipline (Campbell *et al.*, 2001).

As in other models using the generic LUDAS framework, in GLDC-LUDAS there is an automatic classification algorithm, called *AgentCategorizer*, to annually update the livelihood typology of household agents by evaluating the temporal cumulative changes in variables of five main household capitals, namely natural, physical, social, human, and financial capitals (Le et al., 2008, 2010, 2012). These variables – such as land-use structure of household land, agricultural income and so on – are the results of cumulative impacts caused by land-use actions of the considered households and his/her neighbor. *AgentCategorizer* annually compares and ranks dissimilarities between the considered household and all livelihood groups in the population, and then assigns each household into the most similar livelihood group. Thus, an imitative learning behavior is assumed. Details of the algorithm are shown in the following.

The algorithm is similar to the K-mean clustering procedure, except that the group centroids here were predefined outside the simulation model by descriptive statistics of household groups, and thus fixed during the simulation runs. The categorizing process consists of the following steps:

(1) A given household h measures dissimilarities in livelihood typology, based on grouping criteria. He compares between himself and all defined household groups in the population:

$$D_{hg} = \sum_{c=1}^{C} w_c \frac{(H_{h,c} - \overline{H}_{g,c})^2}{\left|H_{h,c} + \overline{H}_{g,c}\right|} \quad (1)$$

where D_{hg} is the Squared Chi-squared Distance from household h to the centroids of group g (g = 1, 2, … , K). $H_{h,c}$ is the instant value of criterion c (c = 1, 2, ... ,C) of household h. Criteria $H_{h,c}$ are household livelihood variables, many of which change as the results of micro household-land interactions during the simulation. Parameter w_c is the weight coefficient of the criteria explaining the discrimination of household groups. The default value of w_c is 1/C.

(2) Household h assigns itself into the most similar livelihood group (g*):

$$g^* = \arg\min (D_{h1}, D_{h2}, \dots , D_{hK}) \quad (2)$$

where g* is the most similar group to household h. D_{h1}, D_{h2}, … , D_{hK} are distances from household h to groups 1, 2, … , K, respectively.

(3) If the livelihood group of a household h has changed, it will ask to delete the old land-use decision model and to adopt the decision model of the new group (*imitative strategy*). Otherwise, the household will repeat its former land-use/crop/management choices (*repeating strategy*). When adopting a new land-use decision model, there are not only changes in parameter values but possibly also in the behavior structure: some decision variables and production components are added or deleted.

Thus, with the *AgentCategorizer* sub-model/algorithm, a household (encoded as a human agent in GLDC-LUDAS) follows its previous behavior as long as it perceives itself as similar (enough) to other households of its group compared to other household groups. If the accumulative changes in household's livelihood variables are large enough and really make the household belonging to another livelihood type, the household will shift to this new type and adopts new behavior by imitating the groups behavior model. However, this change of the livelihood typology happens only in a long-term perspective.

2.8. Model testing

Traditional statistical methods are proved to have a limited capacity in testing integrated dynamic models (Forrester and Senge, 1980; Nguyen and de Kok, 2007). Measuring goodness-of-fit between the simulated and observed data is sometimes considered to be the only legitimate test for model validation. However, this test *alone* is argued to be unable to demonstrate the logical validity of the model's scientific contents (Oreskes *et al.*, 1994; Rykiel, 1996), to have a poor diagnostic power (Kirchner *et al.*, 1996) and even to be inappropriate for the validation of deterministic system dynamics models (Forrester and Senge, 1980). In fact that integrated systems models do not strive for prediction of future values (Nguyen and de Kok, 2007).

Because GLDC-LUDAS belongs to the class of complex human-environmental systems models, we argue that its validity cannot be achieved by only a single test such as point-to-point history matching, but rather a series of tests that could increase the user's confidence in the usefulness of the model. Similar to the extents of validation methodology for integrated systems model proposed by the authors cited above, we follow the multi-criteria validation approach that includes:

(1) Evaluate the fitting of the model to the questions it is meant to answer,

(2) Evaluate the plausibility of the assumptions and theories forming the model (construct validity),

(3) Validate elementary causal relations used for constructing the model (e.g. behavioral rules and sub-models) (internal validity),

(4) Evaluate input data, and

(5) Evaluate of model outputs: it includes (i) the testing of model outputs' behavior/pattern against independent reference data or knowledge, and (ii) sensitivity/uncertainty analyses

3. Scenarios of driving conditions (including management/policy interventions) for ex-ante assessment using GLDC-LUDAS (under developing)

Scenarios for analyzing the smallholder farming systems will be designed for each study area through discussions and interviews with local, national, and regional key informants. Like Gaube et al. (2009) and Le et al. (2010), scenario will represent external framing conditions, local and regional policy and market conditions, and farm household preferences and will draw from well-known global and regional scenarios (Annan, 2000; International Assessment of Agricultural Knowledge, 2008; IFPRI, 2010; World Bank, 2010). Table 2 shows a preliminary formulation of scenarios for Satiri district in Burkina Faso. The elaboration and finalization of input scenarios to be assessed by GLDC-LUDAS will be done in 2020 through participatory processes.

Table 2: Preliminary input scenarios in Satiri district, Burkina Faso

Drivers	Scenario BASE	Scenario TREND	Scenario INT	Scenario TECH
Socio-economic, policy:				
• Input prices (fertilizers, other agricultural inputs)	Default	Max	Min	Max
• Agricultural subsidies	Default	Default	Low/Med	High
• Marketing for crop products	Default	Default	Strongly enhanced	Default
Capacity development:				
• Farmer-to-farmer extension (with/without)	Default	Default	High	Default
• Technical training/education	Default	Default	High	High
Alternative technologies:				
• Mixed fertilizer application	Default	Default	NPK mixed with liming, manure	NPK mixed with liming, manure + bio-fertilizer (e.g. mycorrhiza)
• Improved seeds	Default	Default	Optimal crop mix (local + improved seeds)	Improved seeds for GL/DC crops
• Recycling nutrients in farming system	Default	Default	Traditional	Compost + Mechanized transportation of

				manure/fertilizer to fields
• Intercropping	Default	Default	Increased intercropping area, but still minority	Increased intercropping area, to be majority

Note: BASE: Conditions of 2019 remain constant over 30-year simulation period.
TREND: Follows current conditions and trajectories and continued use of the maximum amount of fertilizer per household when available.
INT: Assumes intervention policies and programs that satisfy farmer requirements while being profitable for the market, and support local and regional farmer assistance and education programs that promote soil conservation and P residue recycling. Recycling of nutrients in animal and crop residues.
TECH: Assumes policies and market forces that drive investment in mechanized recycling and invention of high technology solutions such as genetically improved crop varieties. Local support and empowerment is not a priority. Recycling of nutrients in crop and animal residues, for example, occur via a mechanized measures (e.g. transportation of materials to fields).

References

Annan, K.A., 2000. New Century, New Challenges: Millenium Report of the Secretary-General of the United Nations. United Nations, New York.

Ashley, C., Carney, D., 1999. Sustainable Livelihoods: Lessons From Early Experience. DFID, London, UK.

Belcher, K.W., M.M. Boehm, M.E. Fulton, 2004. Agroecosystem sustainability: a system simulation model approach. Agricultural Systems 79, 225–241.

Benenson, I., Torrens, P., 2004. Geosimulation: Automata-based Modeling of Urban Phenomena (1st Edition). John Wiley & Sons, West Sussex, England.

Boulanger, P.M., Brechet, T., 2005. Models for policy-making in sustainable development: The state of the art and perspectives for research. Ecol. Econ. 55, 337-350.

Bousquet, F., Le Page, C., 2004. Multi-agent simulations and ecosystem management: a review. Ecological Modelling 176, 313-332.

Campbell, B., Sayer, J.A., Frost, P., Vermeulen, S., Ruiz Pérez, M., Cunningham, A., Prabhu, R., 2001. Assessing the performance of natural resource systems. Conservation Ecology 5, http://www.consecol.org/vol5/iss2/art22.

Craswell, E.T., Grote, U., Henao, J., Vlek, P.L.G., 2004. Nutrient Flows in Agricultural Production and International Trade: Ecological and Policy Issues. Discussion Papers on Development Policy. ZEF Center for Development Research, Bonn, Germany.

Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., Mueller, N.D., O/'Connell, C., Ray, D.K., West, P.C., Balzer, C., Bennett, E.M., Carpenter, S.R., Hill, J., Monfreda, C., Polasky, S., Rockstrom, J., Sheehan, J., Siebert, S., Tilman, D., Zaks, D.P.M., 2011. Solutions for a cultivated planet. Nature 478, 337-342.

Forrester, J.W., Senge, P.M., 1980. Tests for building confidence in system dynamics models. In: Jr., L.A.A., J.W., F., J.M., L. (Eds.), System Dynamics. North-Holland Publishing Company, Amsterdam, pp. 209–228.

Gaube, V., Kaiser, C., Wildenberg, M., Adensam, H., Fleissner, P., Kobler, J., Lutz, J., Schaumberger, A., Schaumberger, J., Smetschka, B., Wolf, A., Richter, A., Haberl, H., 2009. Combining agent-based and stock-flow modelling approaches in a participative analysis of the integrated land system in Reichraming, Austria. Landscape Ecology 24, 1149-1165.

Gotts, N.M., Polhill, J.G., 2009. When and how to imitate your neighbours: Lessons from and for FEARLUS. Journal of Artificial Societies and Social Simulation 12, http://jasss.soc.surrey.ac.uk/12/13/12.html.

Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S.K., Huse, G., Huth, A., Jepsen, J.U., Jørgensen, C., Mooij, W.M., Mueller, B., Pe'er, G., Piou, C., Railsback, S.F., Robbins, A.M., Robbins, M.M., Rossmanith, E., Rueger, N., Strand, E., Souissi, S., Stillman, R., Vabø, R., Visser, U., DeAngelis, D.L., 2006. A standard protocol for describing individual-based and agent-based models. Ecological Modelling 198, 115-126.

IFPRI, 2010. Strategies and Policies for Adapting to Global Change. International Food Policy Research Institute, Washington DC.

International Assessment of Agricultural Knowledge, S., and Technology for Development, IIASTD, 2008. Synthesis Report. Approved by the Governments attending the IIASTD Intergovernmental Plenary in Johannesburg, South Africa (7-11 April 2008), Washington, D.C.

Kirchner, J.W., Hooper, R.P., Kendall, C., Neal, C., Leavesley, G., 1996. Testing and validating environmental models. Sci Total Environ 183, 33-47.

Krütli, P., Staufacher, M., Flueler, T., Scholz, R.W., 2010. Functional-dynamic publish participation in technological decision-making: Site selection process of nuclear waste repositories. Journal of Risk Research 13, 861-875.

Le, Q.B., Park, S.J., Vlek, P.L.G., 2010. Land Use Dynamic Simulator (LUDAS): A multi-agent system model for simulating spatio-temporal dynamics of coupled human–landscape system: 2. Scenario-based application for impact assessment of land-use policies. Ecological Informatics 5, 203-221.

Le, Q.B., Park, S.J., Vlek, P.L.G., Cremers, A.B., 2008. Land-Use Dynamic Simulator (LUDAS): A multi-agent system model for simulating spatio-temporal dynamics of coupled human–landscape system. I. Structure and theoretical specification. Ecological Informatics 3, 135-153.

Le, Q.B., Seidl, R., Scholz, R.W., 2012. Feedback loops and types of adaptation in the modelling of land-use decisions in an agent-based simulation. Environmental Modelling & Software 27-28, 83-96.

Liu, J.G., Dietz, T., Carpenter, S.R., Alberti, M., Folke, C., Moran, E., Pell, A.N., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C.L., Schneider, S.H., Taylor, W.W., 2007. Complexity of coupled human and natural systems. Science 317, 1513-1516.

Matthews, R., Selman, P., 2006. Landscape as a focus for integrating human and environmental processes. Journal of Agricultural Economics 57, 199-212.

Matthews, R.B., 2006. The People and Landscape Model (PALM): towards full integration of human decision making and biophysical simulation models. Ecological Modelling 194, 329-343.

Miyasaka, T., Le, Q.B., Okuro, T., Zhao, X., Takeuchi, K., 2017. Agent-based modeling of complex social–ecological feedback loops to assess multi-dimensional trade-offs in dryland ecosystem services. Landscape Ecology 32, 707-727.
Müller, B., Balbi, S., Buchmann, C.M., de Sousa, L., Dressler, G., Groeneveld, J., Klassert, C.J., Le, Q.B., Millington, J.D.A., Nolzen, H., Parker, D.C., Polhill, J.G., Schlüter, M., Schulze, J., Schwarz, N., Sun, Z., Taillandier, P., Weise, H., 2014. Standardised and transparent model descriptions for agent-based models: Current status and prospects. Environmental Modelling & Software 55, 156-163.
Nguyen, T.G., de Kok, J.L., 2007. Systematic testing of an integrated systems model for coastal zone management using sensitivity and uncertainty analyses. Environmental Modelling and Software 22, 1572-1587.
Oreskes, N., Frechette, K.S., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263, 641-646.
Polhill, J.G., Gotts, N.M., Law, A.N.R., 2001. Imitative versus nonimitative strategies in a land-use simulation. Cybernetics Systems 32, 285 - 307.
Rounsevell, M.D.A., Arnette, A., Brown, D.G., de Noblet-Ducoudre, N., Ellis, E., Finnigan, J., Galvin, K., Grigg, N., Harman, I., Lennox, J., Magliocca, N., Parker, D., O'Neil, B., Verburg, P.H., Young, O., 2012. Incorporating Human Behaviour and Decision Making Processes in Land Use and Climate System Models. GLP Report No. 7. GLP-IPO, São José dos Campos.
Rykiel, E.J., 1996. Testing ecological models: the meaning of validation. Ecological Modelling 90, 229-244.
Scheffer, M., Carpenter, S.R., 2003. Catastrophic regime shifts in ecosystems: linking theory to observation. Trends Ecol. Evol. 18, 648-656.
Schlecht, E., Hiernaux, P., 2004. Beyond adding up inputs and outputs: process assessment and upscaling in modelling nutrient flows. Nutrient Cycling in Agroecosystems 70, 303-319.
Scholz, R.W., Binder, C.R., Lang, D.J., Smieszek, T., Stauffacher, M., Brand, F., 2011. A framework for investigating human-envitonment systems (HES). In: Scholz, R.W. (Ed.), Environmental Literacy in Science and Society: From Knowledge to Decisions. Cambridge University Press, Cambridge, pp. 405-521.
Schreinemachers, P., Berger, T., 2011. An agent-based simulation model of human–environment interactions in agricultural systems. Environmental Modelling & Software Volume 26, 845-859
Villamor, G.B., Le, Q.B., Djanibekov, U., van Noordwijk, M., Vlek, P.L.G., 2014. Biodiversity in rubber agroforests, carbon emissions, and rural livelihoods: An agent-based model of land-use dynamics in lowland Sumatra. Environmental Modelling & Software 61, 151-165.
Vitousek, P.M., Naylor, R., Crews, T., David, M.B., Drinkwater, L.E., Holland, E., Johnes, P.J., Katzenberger, J., Martinelli, L.A., Matson, P.A., Nziguheba, G., Ojima, D., Palm, C.A., Robertson, G.P., Sanchez, P.A., Townsend, A.R., Zhang, F.S., 2009. Nutrient Imbalances: Pollution Remains Response. Science 324, 1519-1520.
Voinov, A.A., Bousquet, F., 2010. Modelling with stakeholders. Environmental Modelling and Software 25, 1268 - 1281.
World Bank, 2010. Rising Global Interest in Farmland: Can it Yield Sustainable and Equitable Benefits?

Appendix 1. SetLaborBudget routine

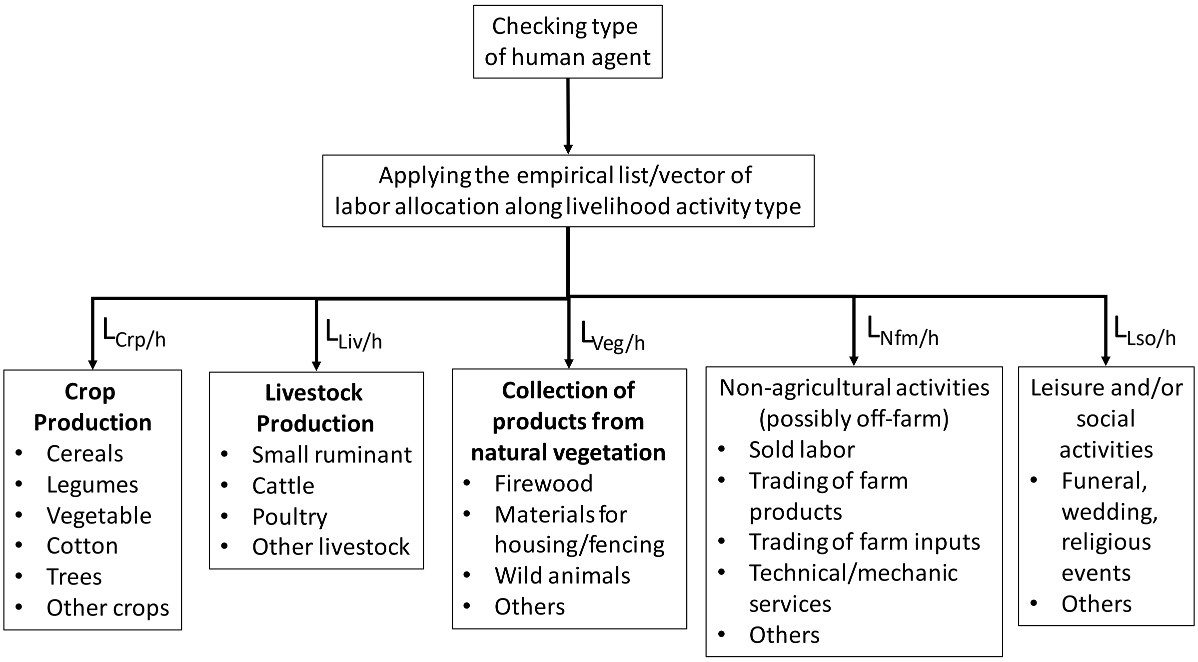


Figure S1. Illustration of SetLaborBudget routine in GLDC-LUDAS. Note: $L_{Crp/h}$, $L_{Liv/h}$, $L_{Veg/h}$, $L_{Nfm/h}$ and $L_{Lso/h}$ are household labor allocated respectively to crop production, livestock production, natural vegetation product collection, non-farm activities and leisure/social activities of household type h. The values of these parameters are empirically estimated for household type h.

Appendix 2. FarmlandChoice sub-model

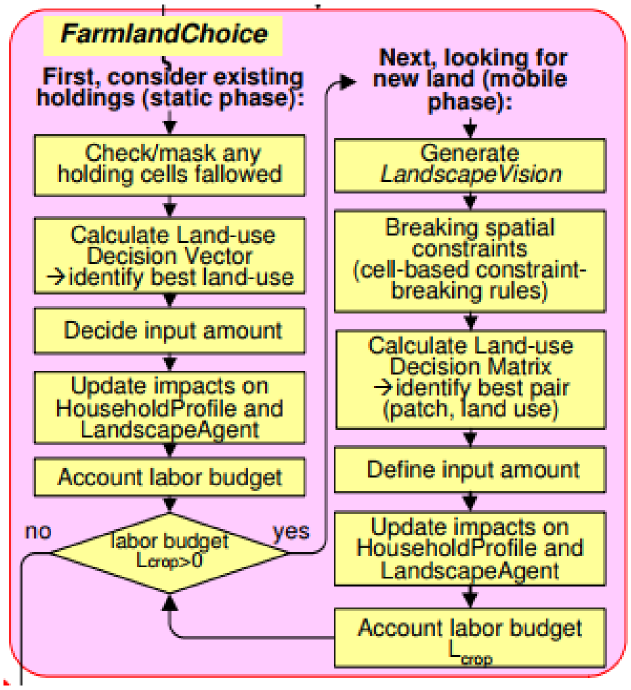


Figure S2.1. Pseudo algorithm of FarmlandChoice sub-model in GLDC-LUDAS. Note: LandscapeVision are a defined subset of the landscape perceived by the household agent, which are considered in searching more land for farming. Source: Le (2005).

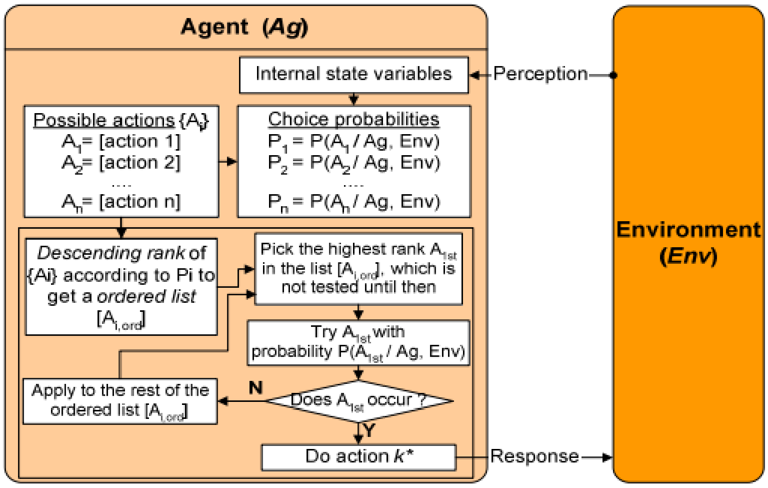


Figure 2.2S. Illustration of the ordered choice algorithm used in the FarmlandChoice sub-model of GLDC-LUDAS. The choice is bounded rational. With the use of ranking procedure, the household prioritizes to select options having the highest utility. However, as Monte Carlo simulation based on choice probabilities used, there is some chance that the household does not select the optimal choice. Source: Le (2005), Benenson and Torrens (2004).

Appendix 3. AgriculturalYieldDynamics sub-model

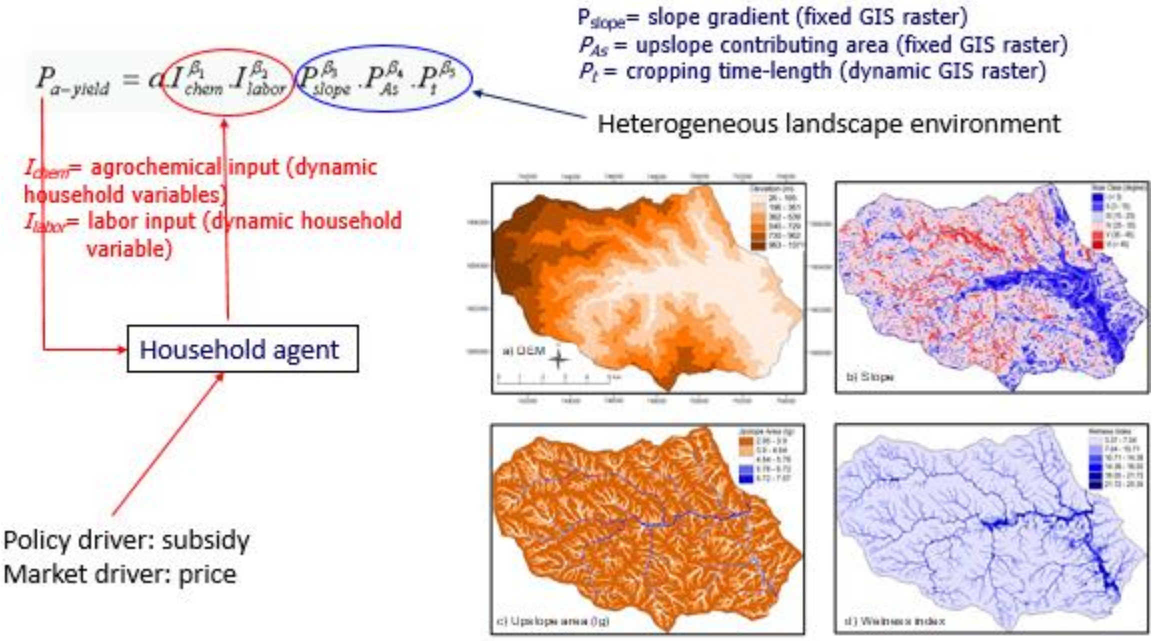
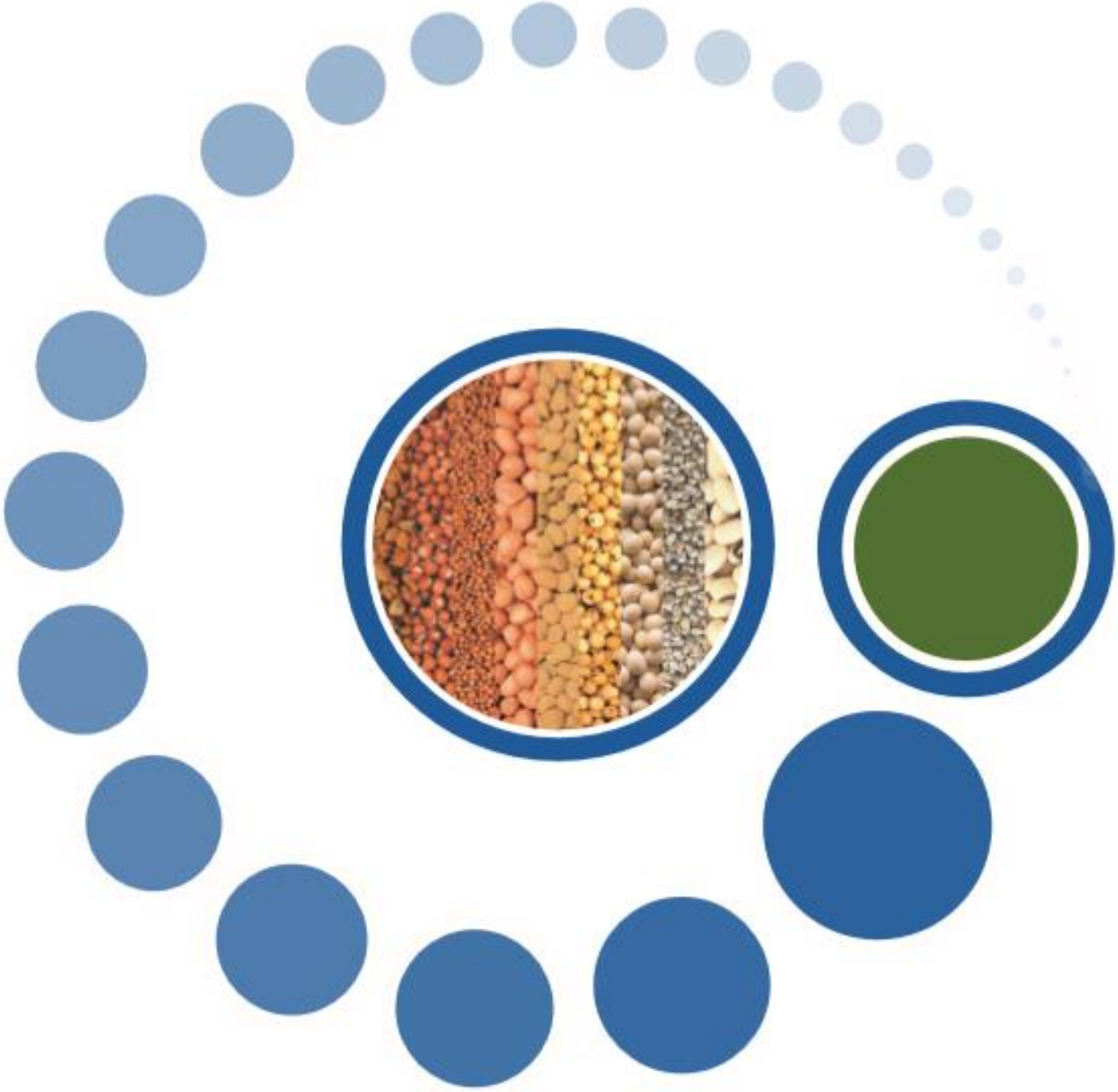


Figure S3. AgriculturalYieldDynamics uses extended Cobb-Douglas production function to combine household decision variables (e.g. amount of labor and agrochemical used) with landscape variables (e.g. site conditions affecting crop growth).



CGIAR

RESEARCH PROGRAM ON Grain Legumes

LED BY ICRISAT

IN PARTNERSHIP WITH CIAT ICARDA IITA

and public and private institutes and organizations, governments, and farmers worldwide