

*Research*, part of a Special Feature on [Empirical based agent-based modeling](#)  
**Creating Agents and Landscapes for Multiagent Systems from Random  
Samples**

*Thomas Berger<sup>1</sup> and Pepijn Schreinemachers<sup>1</sup>*

---

**ABSTRACT.** An important goal of modeling human–environment interactions is to provide scientific information to policymakers and stakeholders in order to better support their planning and decision-making processes. Modern technologies in the fields of GIS and data processing, together with an increasing amount of accessible information, have the potential to meet the varying information needs of policymakers and stakeholders. Multiagent modeling holds the promise of providing an enhanced collaborative framework in which planners, modelers, and stakeholders may learn and interact. The fulfillment of this promise, however, depends on the empirical parameterization of multiagent models. Although multiagent models have been widely applied in experimental and hypothetical settings, only few studies have strong linkages to empirical data and the literature on methods of empirical parameterization is still limited. This paper presents a straightforward approach to parameterize multiagent models in applied development research. The parameterization uses a common sampling frame to randomly select observation units for both biophysical measurements and socioeconomic surveys. The biophysical measurements, i.e., soil properties in this study, are then extrapolated over the landscape using multiple regressions and a digital elevation model. The socioeconomic surveys are used to estimate probability functions for key characteristics of human actors, which are then assigned to the model agents with Monte Carlo techniques. This approach generates a landscape and agent populations that are robust and statistically consistent with empirical observations.

**Key Words:** *common sampling frame; generation of statistically consistent agent populations; integrated modeling; interdisciplinary data collection; Monte Carlo approach; Uganda.*

---

## INTRODUCTION

In applied development research we often find "critical triangle situations" of poverty, resource depletion, and decreasing productivity (Vosti and Reardon 1997). Poor farm households are compelled to apply unsustainable farming practices that erode their natural resource base, reduce crop yields, and in turn, promote poverty. In many developing countries, for example in Uganda, the case study reported on here, farm households are trapped in such a downward spiral of interacting biophysical and socioeconomic forces (Pender et al. 2004). A major research issue in agricultural and development economics is to explore policy options to overcome this critical triangle and, in particular, to analyze the likely impacts of innovations on the livelihood of farm households and their natural resource conditions. Whereas in the past, research

mainly focused on technical innovations, more emphasis is currently paid to institutional innovations such as land rental and labor sharing arrangements, resource user associations, and catchment management boards.

Multiagent system models (MAS) have a large potential to improve the understanding of critical triangle situations, to learn about the uncertainties related with natural resource management and to explore new policy options (Parker et al. 2003). They may also provide a collaborative learning framework in which scientists, policy makers, and stakeholders may interact (Roeling 1999, Hazell et al. 2001, van Paassen 2004). To fulfill their potential, MAS need to be carefully parameterized and validated with empirical data. Only then they may provide relevant information about boundary conditions of rural development and the

---

<sup>1</sup>University of Hohenheim

uncertainties involved, although MAS have been widely applied in hypothetical and experimental settings. However, few studies have tried to build empirically based MAS, and the literature on methods of empirical parameterization is therefore limited (Berger and Parker 2002).

In this paper, we present a straightforward approach for empirical parameterization of MAS in applied development research. The paper starts with an introduction to integrated modeling in agricultural sciences and highlights the challenges of building empirical multiagent models. A subsequent section discusses the peculiarities of developing country research and proposes the use of "common sampling frames" as a suitable method for organizing data collection in interdisciplinary research projects. We then present a statistical approach based on spatial interpolation and Monte Carlo techniques to parameterize MAS with empirical data. Taking the example of ongoing research in Uganda, we show how landscapes and agent populations can be generated from field measurements and farm household survey data. The last two sections discuss the validity of our approach and conclude.

## **INTEGRATED MODELING AND EMPIRICAL MULTIAGENT MODELS**

The use of computer models has a long tradition in agricultural sciences, and high standards for model parameterization and validation have been established. Computer models in agricultural sciences have always been tailored to provide "practical" results for applied development problems. Crop growth models, for example, are used to simulate the fertilizer response of major food crops, thereby substituting for costly field experiments, and linear programming models are used to derive improved farm management plans. There is also vast experience with the coupling of these computer models, for example in bioeconomic models (Barbier 1998, Woelcke 2003, Holden and Shiferaw 2004) and in integrated river basin models (Rosegrant et al. 2000, Fisher et al. 2002).

### **Integrated modeling of natural resource use**

In general, integrated models in agricultural sciences have been very instrumental in capturing the technical or engineering aspects of human-nature interactions and in highlighting the economic

consequences of resource use changes (Kuyvenhoven et al. 1998). They may elucidate the tradeoffs that farm households face in crop choice and farming practices, assess the profitability of various land-use options, and capture the internal costs of adjusting to changes in environmental and marketing conditions (see also the more recent work of Holden et al. 2004). However, they face also limitations when it comes to analyzing critical triangle situations, in which heterogeneity of actors and landscapes is large and increasing (Berger et al. 2006). In general, this is the case when farm households differ considerably in terms of factor endowments and decision-making processes and when resources are exchanged locally or in networks. Another challenge for integrated modeling is to allow for a sufficient degree of spatial and temporal complexity, since changes in the natural environment, the market environment, and the introduction of improved technologies typically involve long-term interacting processes.

Heterogeneity and interactions clearly fall into the core competence of multiagent models, which may therefore extend the scope of computer modeling in applied development research. In the field of natural resource use, MAS were applied to a variety of research questions (for an overview see Janssen 2002, Parker et al. 2003). MAS have been used to theorize about social and spatial dynamics (Gotts et al. 2003, Parker and Meretsky 2004), to simulate land-use changes (Huigen 2004) and diffusion of innovations (Weisbuch 2000, Berger 2001, Deffuant et al. 2002), to assess the impact of agricultural policies (Balmann 1997, Happe 2004), to accompany role-playing games (Barreteau et al. 2003) and in game theory applications (Bousquet et al. 2001). According to the classification proposed by Berger and Parker (2002), most of these applications are abstract or experimental; only few studies have tried to build empirical multiagent systems. Moreover, there are only few studies that have attempted to exploit the potential of MAS in integrating biophysical and socioeconomic model components (Parker and Berger 2002)

### **Challenges to building empirical multiagent system model**

When parameterizing MAS with empirical data, the researcher moves into the problem domain of empirical modeling for which an extensive literature has been produced over the years. As an example,

we mention the influential work of Bockstael (1996) who addressed fundamental issues such as sample selection and spatial autocorrelation. One major challenge is to represent, in a statistically consistent way, a real-world situation of typically heterogeneous biophysical and socioeconomic conditions such as various soil types, crop and vegetation growth, land holdings, social networks, and human actors. Here, we will only focus on the parameterization of the cellular and the agent-based component of MAS and present one possible approach to generate landscapes of grid cells and agent populations. A detailed discussion of agent decision making, error propagation, and sensitivity analysis is, however, beyond the scope of this paper.

## COLLECTION OF EMPIRICAL DATA

As mentioned in the previous section, multiagent system models (MAS) have a large potential for integrated modeling of land-use changes based on empirical data. Although the capabilities of present-day technology, such as GIS and related data processing tools, have made more biophysical information accessible, there are still some challenges for data collection especially in interdisciplinary research projects.

### Common sampling frame

One drawback of doing applied research in developing countries is data scarcity although the situation has much improved in the last two decades. Under data scarcity, scientists in interdisciplinary research projects do best to concentrate their primary data collection on selected sites. However, one tendency is that scientists from various disciplines use different criteria for selecting their units of observation. As a result, data collection activities are often scattered over the study region, soil scientists are taking samples in one set of villages, whereas the social scientists are surveying another set of villages, plant growth models are calibrated for selected varieties, yet not the ones grown by trial farmers etc. Under these conditions, which are unfortunately rather the rule than the exception, it becomes very difficult to integrate empirical data within MAS.

One way out is to apply a statistical procedure to select representative units and subunits of observation using a common sampling frame. This

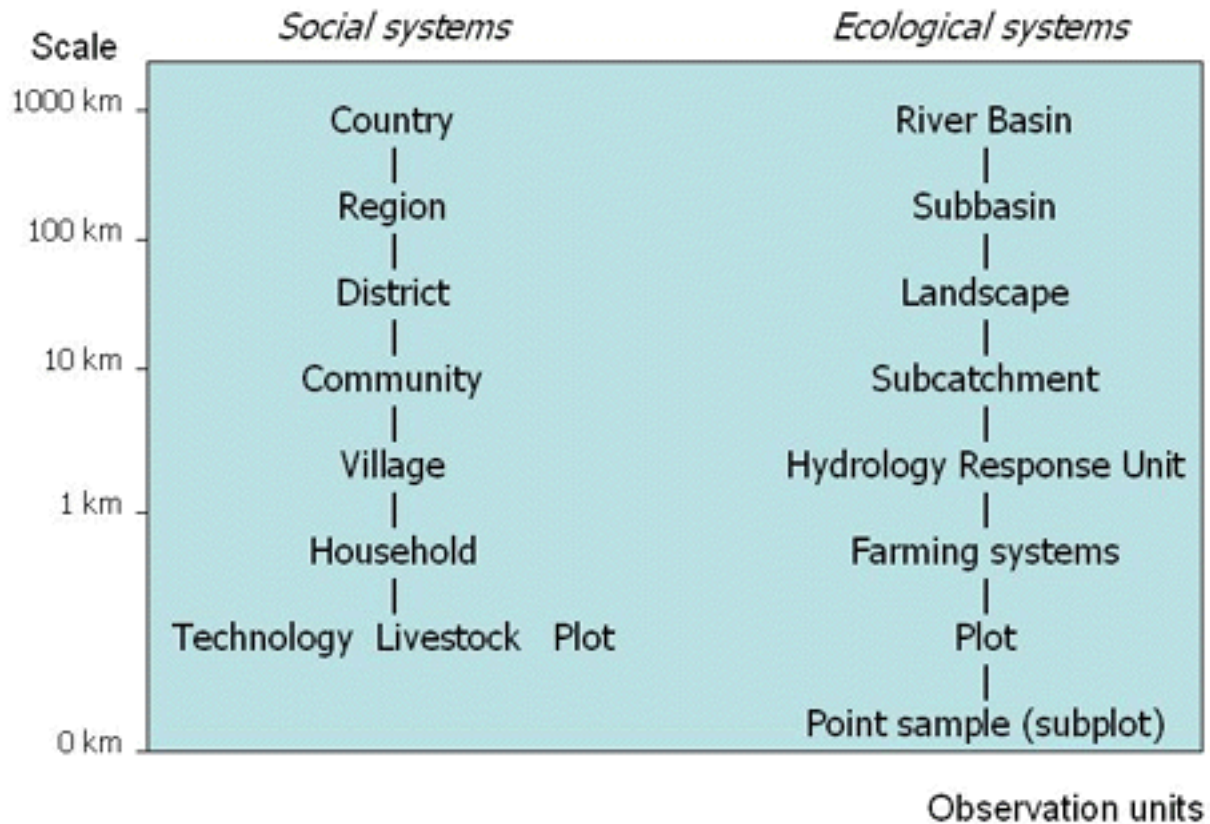
approach was employed for data collection in the Ghanaian Volta basin in West Africa as described in Berger et al. (2006) and van de Giesen et al. (2006). First, data on living standards and geospatial data were compiled from secondary sources and merged into one data set. Project scientists then developed a hierarchy of observation units (Fig. 1) and defined a priori selection criteria that would potentially capture all key research questions of the various scientists involved. Based on preliminary analysis, seven categories of selection criteria at the community level were identified: agro-ecological conditions, agricultural and fishing intensity, market orientation, household welfare, health and water use, social capital, and migration. Correlation analysis of the merged data set revealed interdependencies among some of the selection criteria, and principal component analysis was, therefore, used to derive a relatively small number of linear combinations of the original variables that retain as much statistical information as possible. Eight new variables, explaining about 70% of the total variance in the data, were used for a subsequent cluster analysis. The communities closest to the cluster centroid were then selected as representative communities. Scientists of the various disciplinary subprojects then randomly selected their subobservation units within these sample communities, for example households, water sources, or sample plots.

The advantage of this interdisciplinary approach, apart from the logistical benefits that accrue from the concentration of field activities at certain communities, is that it provides representative and integrated data sets for the empirical parameterization of MAS.

## GENERATION OF LANDSCAPES IN MULTIAGENT SYSTEM MODELS

The remainder of this paper describes one possible approach to parameterize multiagent system models in applied development research. The parameterization combines predictive soil maps to generate a landscape from soil samples, and Monte Carlo techniques to generate agent populations from a sample of farm households. The method is described on the basis of a case study for two village communities in Southeastern Uganda, which were selected to analyze the downward spiral of declining soil fertility and increasing poverty and to assess ex ante the impacts of alternative policy interventions.

**Fig. 1.** Hierarchy of observation units in interdisciplinary research.



The MAS software used for this case study builds on earlier work of Balmann (1997) and Berger (2001). It is coded in C++ and runs on Windows and Unix/Linux platforms (Appendix 1). Since this paper discusses only how to initialize the MAS, but does not present results from using this software, we will not provide a detailed description here. Some indications on functionality and where to download more information are given in the appendix. One integral component of this software is a crop growth and nutrient cycling module that was calibrated for agro-ecological conditions in Uganda using the "Tropical Soil Productivity Calculator" (see Aune and Lal 1995). Soil measurements and predictive soil maps were needed to parameterize this module.

### Research area

The two village communities are Magada and Buyemba in the Mayuge District; Magada counts 374 households and Buyemba counts 247 households. Both communities are densely populated with an average of 436 and 383 people/km<sup>2</sup> in Magada and Buyemba, respectively, and are relatively well connected to market towns.

Climatic conditions allow the cultivation of two sequential crops in a year. Main food crops are cassava, sweet potato, and beans; the main cash crop is coffee, whereas maize and plantain are both sold and home consumed. Farm households predominantly rely on the hand hoe in their crop management; the use of external inputs such as fertilizers, pesticides,

and improved seeds is rare. Intercropping is common and farm households usually allocate only small parts of a plot to a single crop combination. Soil fertility is generally low but varies across locations. The landscape is fairly sloping with large flat areas, and erosion levels are moderate (Brunner et al. 2004). A spatially explicit model was chosen to capture this heterogeneous landscape with local interactions.

### **From soil samples to continuous soil maps**

In the MAS, a landscape of grid cells represents the biophysical environment that farm households manage. In the Uganda study, the landscape model contains 12 layers, including soil chemical and physical properties, village boundaries and the location of farmsteads and agricultural plots. Layers are composed of grid cells of 71 x 71 m (0.5 ha), which is the smallest amount of land cultivated by a single farm household.

Empirical information about soil properties is obtained from soil samples. The challenge was to create continuous soil maps by interpolating soil sample values. There are various approaches to this such as a kriging interpolation using semivariogram models or distance weighting algorithms (e.g., Ruecker 2005). The Uganda study used predictive soil mapping based on stepwise multiple regressions of soil properties on terrain parameters and/or other soil properties (Rhew et al. 2004). Terrain parameters were derived from a digital elevation model (DEM) with a 30-m grid size, and included elevation, slope, upslope area, plan curvature, profile curvature, curvature, wetness index, streampower index, and aspect.

In a first stage, the prediction was based on 285 soil samples and 910 GPS measurements from a single hillslope in the village of Magada collected in 2000 (Ruecker 2005). Predictive models were initially estimated from these data and scaling effects were explored to find robust estimators at different scales. In a second stage, a new round of 120 soil samples and GPS measurements were collected in two villages in 2003. These data were used to validate the predictive models of the first stage, and to subsequently modify these (Rhew et al. 2004).

### **Distribution of agents into the landscape**

The next challenge was to populate the landscape model with agents. The location of farmsteads and agricultural plots can be obtained through GPS measurements, maps of the land registry office, or aerial photography, yet such information is sometimes unavailable. In the Uganda study, hand-drawn maps were available on which most of the sample households had been marked during the survey (1999–2000) but not the exact location of their plots. The maps, however, showed that farmsteads were mainly located along community roads with scattered farmsteads in the other parts of the villages. The remaining nonsample households and all plots were to be assigned using these qualitative patterns. In ongoing research in Chile, complete georeferencing of household surveys and land registry maps is used.

When complete data sets are lacking, random assignment is an option and subsequent sensitivity testing to repeated random assignments can reveal the impact of this unknown factor on simulation outcomes. Figure 2 shows the different stages in generating the spatially located agents and farm plots for the village of Magada; the same procedure was applied to Buyemba. The left upper panel (Fig. 2-A) depicts the sample points within the village boundary of Magada. The figure shows spatial dependency as sample farm households are not evenly distributed in the landscape but are clustered around the road network.

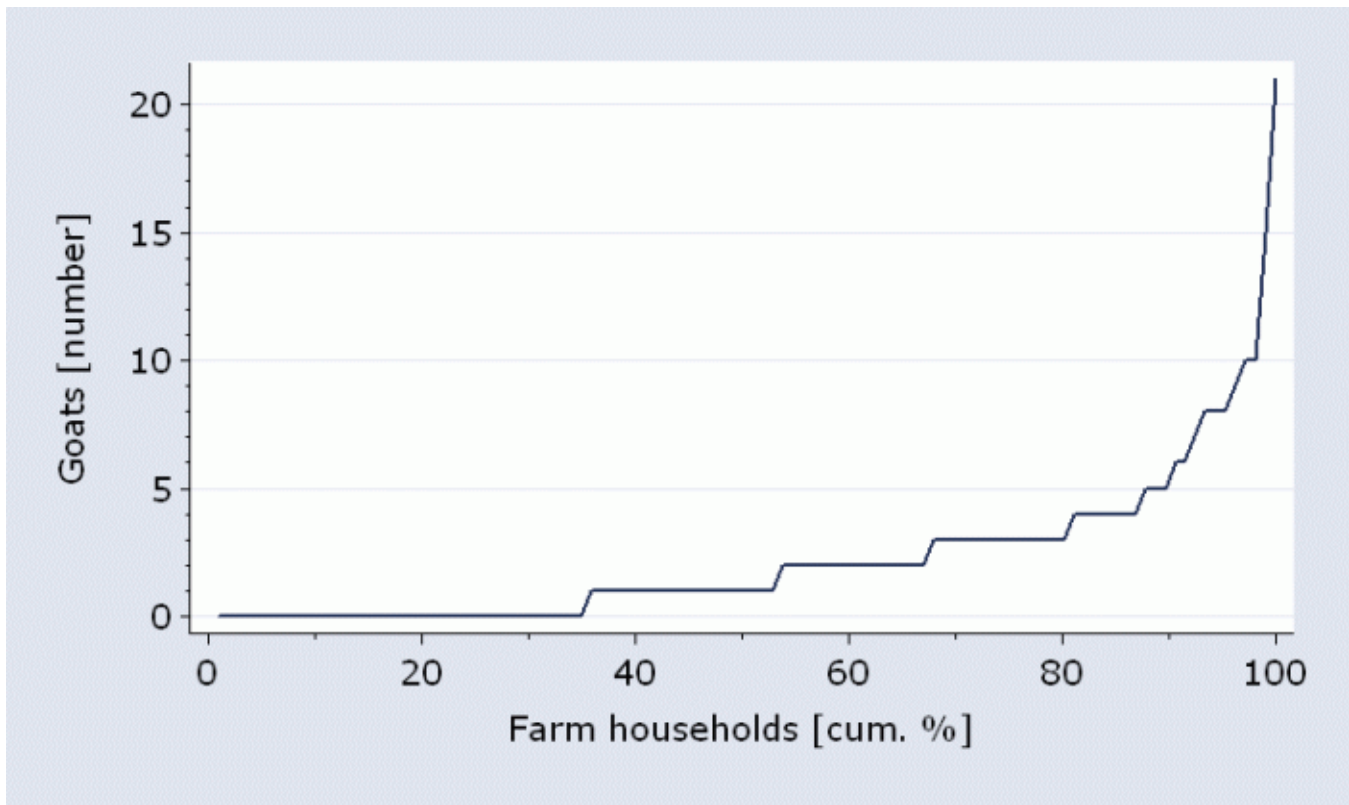
Two different areas according to population density were therefore first demarcated: areas alongside the road network were designated as of high population density, and all other areas were of low population density (Fig. 2-B). Because the sample was random, the geographical distribution of the sample households represents the distribution of the total population. In Magada, for instance, 84% of the sample households lived in the high-density area, which accounted for 40% of the total village area. Of the remaining nonsample households, 84% was, thus, allocated in the high-density area and 16% in the low-density area. A standard routine for spatial random allocation, DNR Sampling Tool extension in ArcView GIS, was used for this purpose.

All allocated farmsteads were then converted into grid cells, as shown in Fig. 2-C. Finally, using the estimated sample distribution from the survey, agricultural plots were allocated to the agents. A

**Fig. 2.** Spatial generation of agent population and agricultural plots from a sample of farm households.



**Fig. 3.** Empirical cumulative distribution of goats over all households in the sample.



random spatial allocation was not used at this stage, as this would have produced an unrealistically scattered pattern of farm plots. The allocation was therefore done manually based on available qualitative information (Fig. 2-D).

### GENERATION OF AGENT POPULATIONS IN MULTIAGENT SYSTEM MODELS

When generating an empirically based multiagent system model, (MAS), every computational agent must represent a single real-world farm household. In applied development research, random samples are typically preferred to population censuses or censuses of agriculture. Censuses do capture all farm households in the study area but, for financial and time constraints, cannot provide in-depth data of high quality (Carletto 1999). In the Uganda study, data were collected together with soil samples in 1999–2000 for 106 households, which is about 17% of the population. The challenge, hence, was to

extrapolate the sample population to parameterize the remaining 83% of farm households.

### Monte Carlo approach

Monte Carlo studies are generally used to test the properties of estimates based on small samples. It is thus well suited to this study, in which data about a relatively small sample of farm households are available, but the interest goes into the properties of an entire population. The first stage in a Monte Carlo study is modeling the data generating process, and the second stage is the creation of artificial sets of data.

The methodology applied here is based on empirical cumulative distribution functions. Figure 3 illustrates such a function for the distribution of goats over farm households. The figure shows that 35% of the farm households in the sample have no goats; the following 8% have one goat, etc. This

function can be used to randomly distribute goats over model agents. For this, a random integer between 0 and 100 is drawn for each agent and the number of goats is then read from the y-axis. Applying this procedure for all agents recreates the depicted empirical distribution function for goats. By varying the random seed number, the procedure yields different endowments at the individual agent level.

When repeating the Monte Carlo procedure for all other resources, each resource would be allocated independently at the agent level, excluding the event of possible correlations between resources. However, actual resource endowments typically correlate, for example, larger households have more livestock and more land. To include these correlations in the agent populations, first the resource that most strongly correlates with all other resources is identified and used to divide the survey population into a number of clusters. Empirical cumulative distribution functions are then calculated for each cluster of sample observations.

In the Uganda study, the sample was divided into clusters defined by household size because extensive data analysis showed that this was the variable most strongly correlated with all other variables. Nine clusters were chosen because this captured most of the different sizes in households and gave a minimum of five households per cluster. Cluster analysis can also be used for this purpose if several variables show strong correlations, but clusters produced this way are more difficult to interpret, especially when many variables are used. Figure 4 illustrates the approach for the distribution of goats.

Each agent was allocated quantities of up to 80 different resources in the Monte Carlo procedure. These resources include 68 different categories of household members, i.e., 34 age groups of two sexes, four livestock types, i.e., goats, young rams, cows, and young bulls, area under coffee plantation, female head of household, liquidity, ratio of equity and debt capital, plus innovativeness. Agents were generated sequentially, that is, agent No. 1 first drew 80 random numbers in 80 different cumulative distribution functions before agent No. 2 did the same.

## Consistency checks

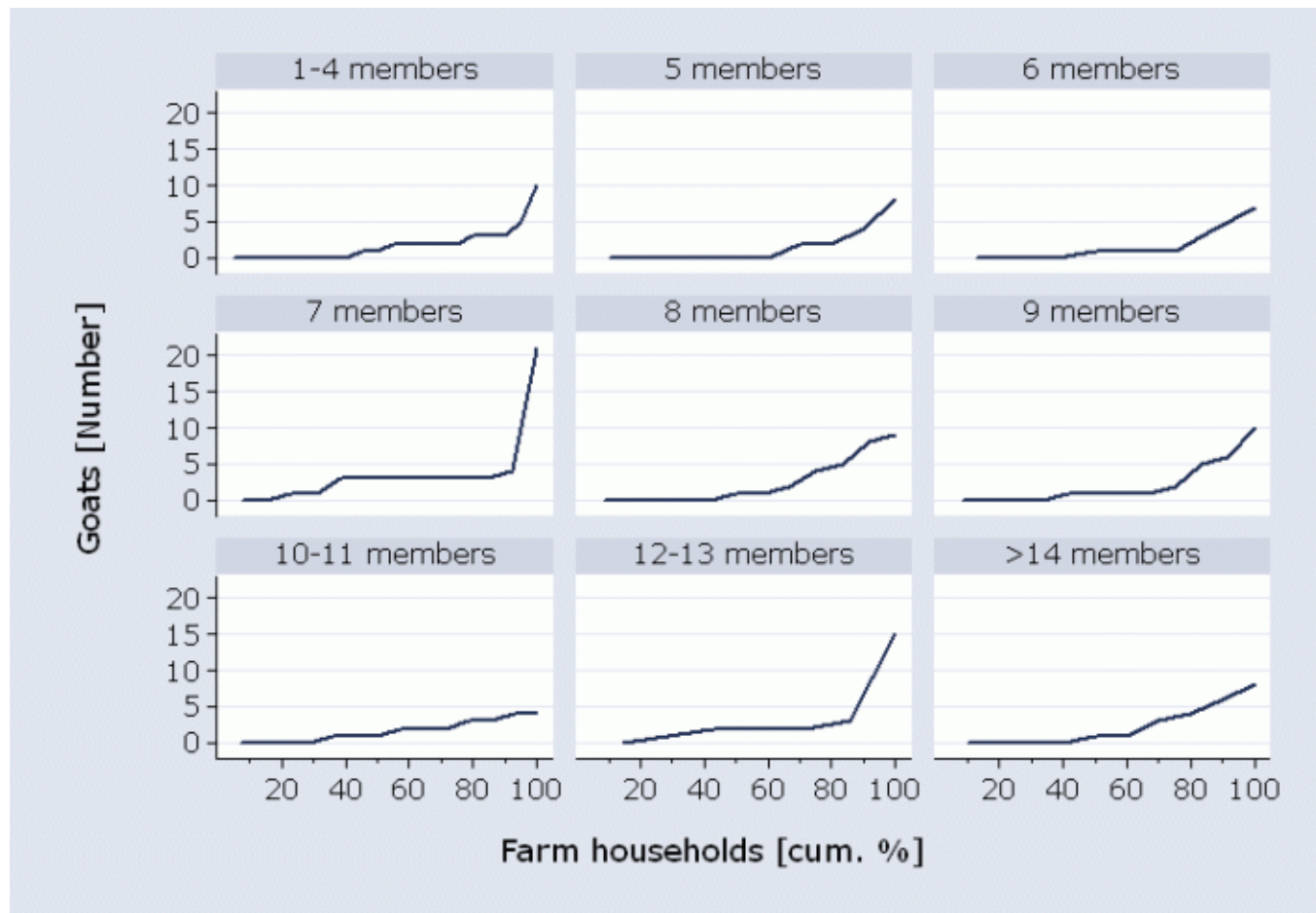
In order to get both statistically consistent and realistic agents, the generated agent populations were submitted to three tests at various levels of aggregation:

- 1. Checks for inconsistencies at the population level:** The average resource endowments of the agent population have to lie within the confidence intervals of each estimated sample mean. If not, the agent population generated from this seed value has to be rejected, and the agent random assignment is repeated with a new seed.
- 2. Checks for inconsistencies at the cluster level:** The correlation matrix of agent resource endowments has to reflect the correlation matrix of the sample population. Otherwise, the agent population has to be rejected.
- 3. Checks for inconsistencies at the agent level:** An agent with 20 household members is very unlikely to have only one plot of land. However, because of the randomness of the resource allocation, unrealistic settings can occur in the agent population. By defining a lower and/or upper bound for some critical combinations, this problem can be overcome. If a resource combination lies outside such bound the generated agent has to be rejected, and the random assignment of this particular resource combination is repeated. Two sets of bounds are included. The first set defines minimum land requirements for livestock and the second set defines demographic rules to ensure realistic family compositions.

The Monte Carlo approach outlined here works well if correlations among agent characteristics are not too tight. If individual agents, clusters of agents or entire agent populations are continuously being rejected on one of the above three criteria, then the cluster-specific distribution functions have to be fine tuned. By skewing the distribution functions towards otherwise underrepresented combinations of agent characteristics, as was necessary in ongoing



**Fig. 4.** Empirical cumulative distribution functions of goat numbers over nine clusters by household size.



research in Chile, the random assignment may then still yield statistically consistent agent populations.

## RESULTS

To test the methodology, a large number of agent populations was generated by applying different random seed values. Their properties were analyzed at three levels: (1) the population level, (2) the cluster level, and (3) the level of the individual agents. Each of these analyses is discussed. Within the scope of this paper, the entire variation between and within agent populations cannot be shown. Instead, the results are illustrated with a few examples and snapshots from the agent populations.

### Population level

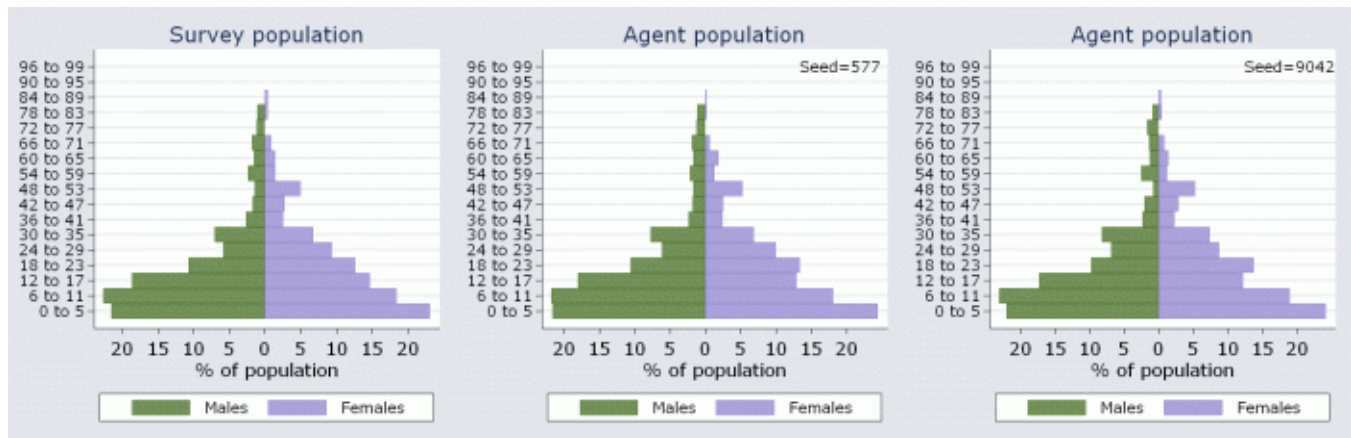
At the population level, we checked whether the averages in the agent population resembled those of the survey population. For this purpose, average resource allocations for 100 generated agent populations were calculated (Table 1). For all resources, the average resource endowments in the agent population fell within the 95% confidence interval of the survey average, and the difference between the two averages was generally small. The random agent generator hence reproduces population averages.

To get more detail about the demographic structure of the population, a population pyramid for the

**Table 1.** Resource endowments of the survey population compared to meta-averages of the agent population. Agent population is an average over 100 different agent populations. SE is the standard error of the average referring to the average within the survey population and SD is the standard deviation of the average referring to the average across agent populations.

Resource	Population	Average	SE SD <sup>1</sup>	Confidence interval	
Household members	Survey	7.87	0.45	6.99	8.75
	Agent	7.89	0.11		
% children	Survey	55.06	2.47	50.22	59.91
	Agent	54.87	0.75		
Cows	Survey	0.81	0.18	0.45	1.17
	Agent	0.81	0.02		
Young bulls	Survey	0.08	0.04	0.01	0.16
	Agent	0.09	0.01		
Goats	Survey	1.29	0.16	0.98	1.61
	Agent	1.23	0.04		
Young rams	Survey	0.14	0.04	0.06	0.23
	Agent	0.14	0.02		
Coffee, ha	Survey	0.31	0.10	0.11	0.51
	Agent	0.31	0.02		
Plots, 0.5 ha	Survey	4.58	0.51	3.58	5.58
	Agent	4.34	0.00		
Innovativeness	Survey	3.88	0.17	2.35	3.03
	Agent	3.85	0.04		

**Fig. 5.** Population pyramids for survey and one agent population compared.



survey population was drawn and compared with those of two agent populations in Fig. 5. This form of presentation may not be very suitable for comparing age groups exactly but does illustrate the high degree of similarity between the two populations.

### Cluster level

The above has shown that the sample population was well replicated at the aggregate population level, but this might not necessarily be so at lower levels of aggregation. The following graphs and figures, thus, look at the cluster and agent level.

Figure 6 depicts four box plots comparing the distribution of household size, area under coffee, and goats and cows in the sample with an agent population with seed value 577. Each box ranges from the 25th to the 75th percentile, i.e., the interquartile range, with the 50th percentile, or median, also marked within. Clusters were based on household size, which is why there is a strong correlation between these two variables in the left upper pane. The figure shows that median values do not differ much between the survey population and the agent population. In addition, most interquartile ranges are of comparable width, except for household size, but that is because this variable was used to define the clusters.

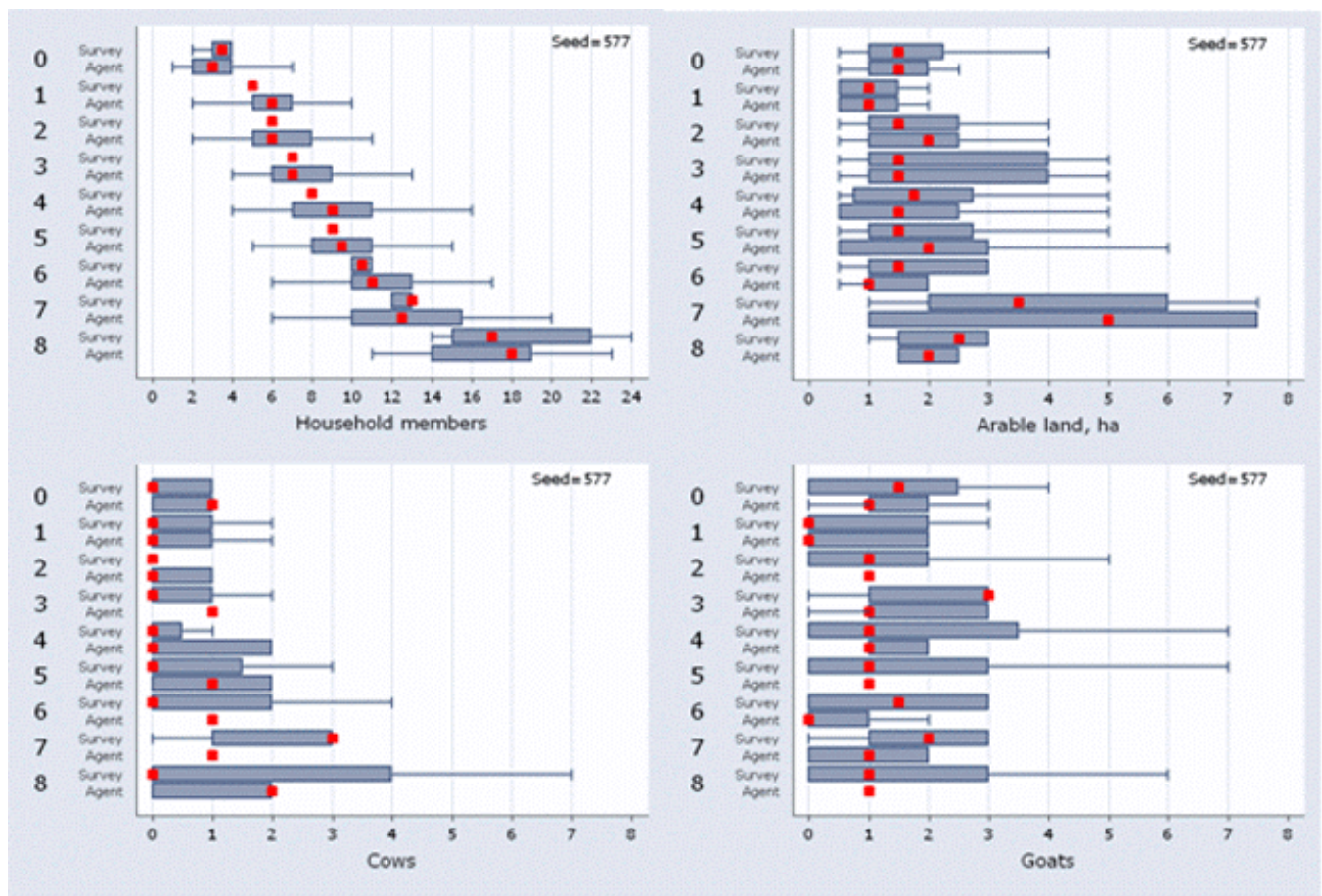
### Agent level

In the Ugandan case study, assigning the spatial location of farmsteads and farm plots is not part of the Monte Carlo procedure. Because of the lack of georeferenced data, we could not meaningfully redraw and statistically crosscheck the agent location maps. The land endowment was therefore constant for each agent in all agent populations. Only the nonland resources were randomly allocated to the agents based on cluster-specific distribution functions. Figure 7 plots household size against the number of plots per agent.

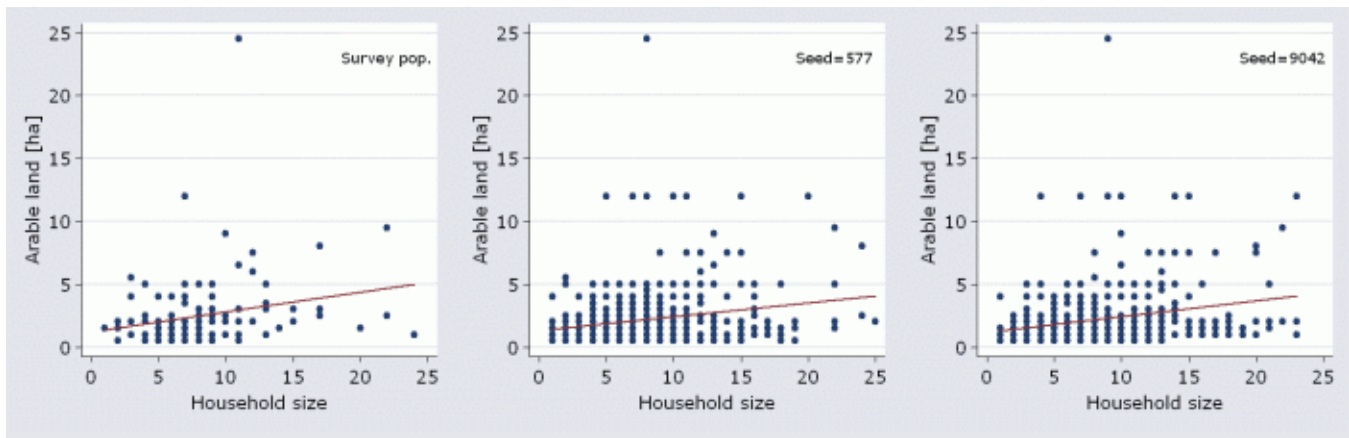
One objective of generating agents randomly was to endow each agent differently in alternative agent populations. The success of the approach is illustrated with Fig. 8. This box plot shows the variation in resource endowments for agent No. 100 over 100 alternative populations. Agent No. 100 has a fixed location for farmstead and plots in the landscape, as can be seen from the zero variance in the agent's land area of 1.5 ha. The variation in resource endowments is high with, for instance, the household size varying between one and eight members and the number of goats varying between zero and three.

The reproduction of correlations was the third objective in the random agent generation. The left diagram in Fig. 9 plots the number of adults against the number of children in the survey population,

**Fig. 6.** Boxplots for the distribution of the four major resources over clusters.



**Fig. 7.** Correlation between household size and amount of arable land.



whereas the two right panes do the same for two generated agent populations. The figures show that correlation between adults and children within the household, as observed in the survey, is well replicated in the agent populations, ensuring that the agents created are demographically consistent in this respect.

## DISCUSSION

A less demanding approach to parameterize model agents based on the Uganda survey data would be to multiply every farm household in the sample by a factor of six, or if the sample is not random, by the inverse of every observation's probability weight. Average values in such agent population would exactly equal those of the sample survey. This cloning procedure, however, is unsatisfactory for two reasons.

First, it reduces the variability in the model agent population. A sampling fraction of about 17% gives six identical agents, or clones, in the agent population. This might affect the simulated system dynamics, as these agents are likely to behave analogously. It becomes difficult then to interpret, for instance, a structural break in simulation outcomes: is the structural break endogenous, caused by agents breaking with their path dependency, or is the break simply a computational artifact resulting from the fact that many agents are

the same? This setback becomes the more serious the smaller the sampling fraction is, because a higher share of the agents is identical.

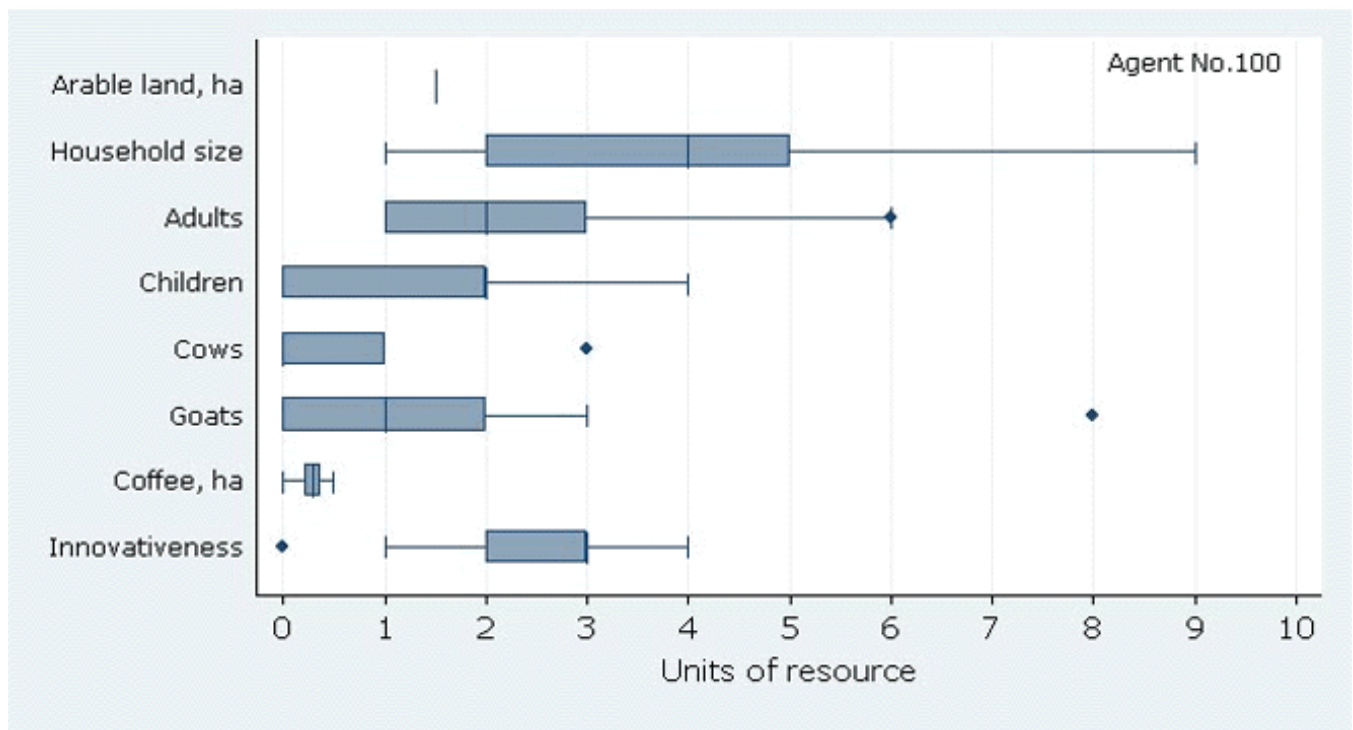
Second, the random sample contains a sampling error of unknown magnitude, which is also multiplied in the procedure. When using the cloning procedure, only a single agent population can be created, whereas for sensitivity analysis a multitude of alternative agent populations is needed.

For these reasons, we developed a straightforward Monte Carlo approach that generates populations that are both statistically consistent and contain high levels of agent heterogeneity. The literature suggests alternative methods for Monte Carlo sampling and correlation control such as Latin hypercube sampling and restricted pairing (Saltelli et al. 2000). An improvement to our method could be to assign values to each of the model variables that conform to both their own empirical distributions and their correlations with other variables. More detailed testing will be needed to compare the performance of different approaches.

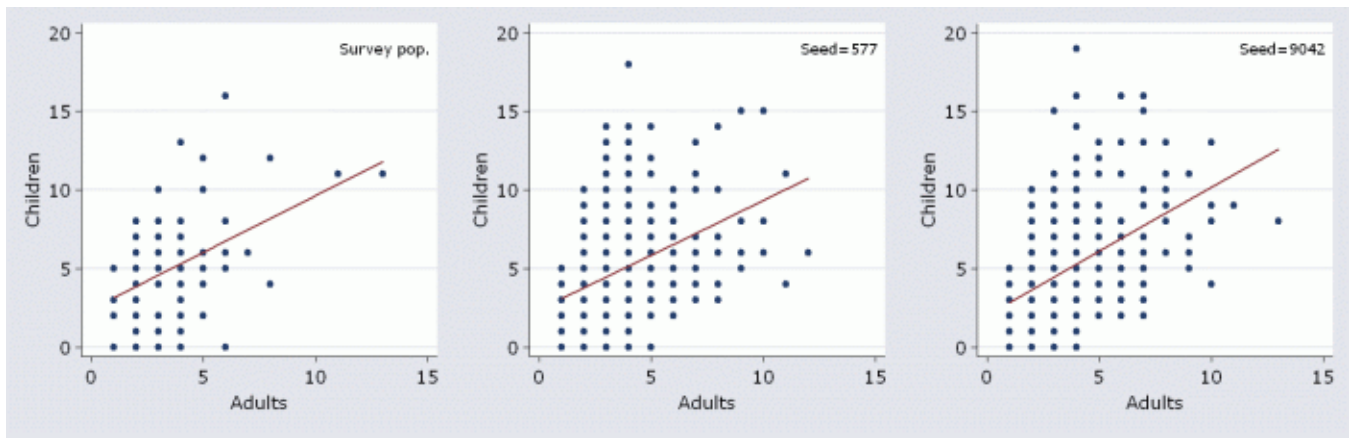
## CONCLUSION

Most multiagent systems have been based on hypothetical settings using artificial rather than real data. This paper showed that MAS could also reproduce real-world conditions by empirical

**Fig. 8.** Boxplot illustrating the variation in agent endowments in alternative agent populations.



**Fig. 9.** Scatter plots correlating the number of children and adults, with regression line.



parameterization of agents, representing real-world farm households. The paper outlined a method to generate agent populations from farm household survey data in combination with spatial data on the location of plots and farmsteads and the quality of soils.

Responses to this article can be read online at:  
<http://www.ecologyandsociety.org/vol11/iss2/art19/responses/>

### Acknowledgments:

*The authors gratefully acknowledge the financial support of the Robert Bosch Foundation and of Senator Dr. Hermann Eiselen who endowed the Josef G. Knoll Visiting Professorship at Hohenheim University. The research in Uganda was conducted under scientific agreements among the Center for Development Research (University of Bonn), the Institute for Regional Studies (Seoul National University), and Noragric (Centre for International Environment and Development Studies). We thank our colleagues Soojin Park, Hosang Rhew, Gerd Ruecker, and Jens Aune for their inputs to the biophysical model components.*

### LITERATURE CITED

- Aune, J. B., and R. Lal.** 1995. The tropical soil productivity calculator—a model for assessing effects of soil management on productivity. Pages 499-520 in R. Lal and B. A. Stewart, editors. *Soil management: experimental basis for sustainability and environmental quality*. CRC Press, Lewis, Boca Raton, Florida, USA.
- Balman, A.** 1997. Farm-based modelling of regional structural change: a cellular automata approach. *European Review of Agricultural Economics* 24:85-108.
- Barbier, B.** 1998. Induced innovation and land degradation: Results from a bioeconomic model of a village in West Africa, *Agricultural Economics* 19:15-25.
- Barreteau, O., C. Le Page, and P. D'Aquino.** 2003. Role-playing games, models and negotiation processes. *Journal of Artificial Societies and Social Simulation* 6(2). Available online at: <http://jasss.soc.surrey.ac.uk/6/2/10.html>.
- Berger, T.** 2001. Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. *Agricultural Economics* 25 (2/3):245-260.

- Berger, T., and D. C. Parker.** 2002. Examples of specific research—introduction. Pages 26-25 in D. C. Parker, T. Berger, and S. M. Manson, editors. *Agent-based models of land use/land cover change*. LUCS Report Series 6, Louvain-la-Neuve. Available online at: [http://www.indiana.edu/%7Eact/focus1/ABM\\_Report6.pdf](http://www.indiana.edu/%7Eact/focus1/ABM_Report6.pdf).
- Berger, T., P. Schreinemachers, and J. Woelcke.** 2006. Multi-agent simulation for development of less-favored areas. *Agricultural Systems* 88:28-43.
- Bockstael, N. E.** 1996. Modeling economics and ecology: the importance of a spatial perspective. *American Journal of Agricultural Economics* 78 (5):1168-1180.
- Bousquet, F., R. Lifran, M. Tidball, S. Thoyer, and M. Antona.** 2001. Agent-based modelling, game theory and natural resource management issues. *Journal of Artificial Societies and Social Simulation* 4(2). Available online at: <http://www.soc.surrey.ac.uk/JASSS/4/2/0.html>.
- Brunner, A. C, S. J. Park, G. R. Ruecker, R. Dikau, and P. L. G. Vlek.** 2004. Catenary soil development influencing erosion susceptibility along a hillslope in Uganda. *Catena* 58:1-22.
- Carletto, C.** 1999. *Constructing samples for characterising household food security and for monitoring and evaluating food security interventions: theoretical concerns and practical guidelines*. Technical Guide 8, International Food Policy Research Institute, Washington, D.C., USA. Available online at: <http://www.ifpri.org/themes/mp18/techguid/tg08.pdf>.
- Deffuant, G., S. Huet, J. P. Bousset, J. Henriot, G. Amon, and G. Weisbuch.** 2002. Agent-based simulation of organic farming conversion in Allier département. Pages 158-187 in M. A. Janssen, editor. *Complexity and ecosystem management: the theory and practice of multi-agent systems*. Edward Elgar, Cheltenham, U.K.
- Fisher, F. M., S. Arlosoroff, Z. Eckstein, M. Haddadin, S. G. Hamati, A. Huber-Lee, A. Jarrar, A. Jayyousi, U. Shamir, and H. Wesseling.** 2002. Optimal water management and conflict resolution: the Middle East Water Project. *Water Resources Research* 38(11):1243-1260.
- Gotts, N. M., J. G. Polhill, and A. N. R. Law.** 2003. Agent-based simulation in the study of social dilemmas. *Artificial Intelligence Review* 19 (1):3-92.
- Happe, K.** 2004. Agricultural policies and farm structures: agent-based modelling and application to EU-policy reform. *Studies on the agricultural and food sector in central and eastern Europe* 30, IAMO. Available online at: [http://www.iamo.de/dok/sr\\_vol30.pdf](http://www.iamo.de/dok/sr_vol30.pdf).
- Hazell, P. B. R., U. Chakravorty, J. Dixon, and R. Celis.** 2001. *Monitoring systems for managing natural resources: economics, indicators and environmental externalities in a Costa Rican watershed*. EPTD Discussion Paper 73. International Food Policy Research Institute. Available online at: <http://www.ifpri.org/divs/eptd/dp/papers/eptdp73.pdf>.
- Holden, S., and B. Shiferaw.** 2004. Land degradation, drought and food security in a less-favored area in the Ethiopian highlands: a bio-economic model with market imperfections. *Agricultural Economics* 30(1):31-49.
- Holden, S., B. Shiferaw, and J. Pender.** 2004. Non-farm income, household welfare, and sustainable land management in a less-favored area in the Ethiopian highlands. *Food Policy* 19(4):369-392.
- Huigen, M. G. A.** 2004. First principles of the MameLuke multi-actor modelling framework for land use change, illustrated with a Philippine case study. *Journal of Environmental Management* 72 (1-2):5-21.
- Janssen, M. A., editor.** 2002. *Complexity and ecosystem management: the theory and practice of multi-agent systems*. Edward Elgar, Cheltenham, U. K.
- Kuyvenhoven, A., H. Moll, and R. Ruben.** 1998. Integrating agricultural research and policy analysis: analytical framework and policy applications for bio-economic modeling. *Agricultural Systems* 58(3):331-349.
- Parker, D. C., and T. Berger.** 2002. Synthesis and discussion. Pages 79-88 in D. C. Parker, T. Berger, and S. M. Manson, editors. *Agent-based models of land use/land cover change*. LUCS Report Series 6, Louvain-la-Neuve. Available online at: [http://www.indiana.edu/%7Eact/focus1/ABM\\_Report6.pdf](http://www.indiana.edu/%7Eact/focus1/ABM_Report6.pdf).



**Parker, D. C., S. M. Manson, M. A. Janssen, M. Hoffmann, and P. Deadman.** 2003. Multi-agent system models for the simulation of land-use and land-cover change: a review. *Annals of the Association of American Geographers* 93 (2):314-337.

**Parker, D., and V. Meretsky.** 2004. Measuring pattern outcomes in an agent-based model of edge-effect externalities using spatial metrics. *Agriculture, Ecosystems, and Environment* 101:233-250.

**Pender, J., P. Jagger, E. Nkonya, and D. Sserunkuuma.** 2004. Development pathways and land management in Uganda: causes and implications. *World Development* 32(5):767-792.

**Rhew, H., S. Park, and G. Ruecker.** 2004. *Predictive soil mapping at regional scale in Iganga District, Uganda, Final report.* University of Seoul, South Korea.

**Roeling, N.** 1999. Modelling the soft side of the land: the potential of multi-agent systems. Pages 73-97 in C. Leeuwis, editor. *Integral design: Innovation in agriculture and resource management.* Mansholt Institute, Wageningen, The Netherlands.

**Rosegrant, M. W., C. Ringler, D. C. Cai, C. McKinney, A. Keller, and G. Donoso.** 2000. Integrated economic-hydrologic water modeling at the basin scale: the Maipo River Basin. *Agricultural Economics* 24(1):33-46.

**Ruecker, G. R.** 2005. Spatial variability of soils on national and hillslope scale in Uganda. *Ecology and Development Series* 24. Available online at: [http://www.zef.de/fileadmin/webfiles/downloads/zefc\\_ecology\\_development/ecol\\_dev\\_24\\_abstract.pdf](http://www.zef.de/fileadmin/webfiles/downloads/zefc_ecology_development/ecol_dev_24_abstract.pdf).

**Saltelli, A., K. Chan, and M. Scott, editors.** 2000. *Sensitivity analysis.* Probability and Statistics Series, Wiley, New York, New York, USA.

**van de Giesen, N., T. Berger, M. Iskandarani, S. J. Park, and P. L. G. Vlek.** 2006. Integrative water research in the Volta Basin. Pages 169-186 in E. Ehlers and T. Krafft, T., editors. *Earth system science in the Anthropocene: emerging issues and problems.* Springer-Wissenschaftsverlag, Berlin, Germany.

**van Paassen, J. M.** 2004. *Bridging the gap: computer model enhanced learning about natural resource management in Burkina Faso.* Dissertation, Wageningen University, Wageningen, The Netherlands. Available online at: <http://www.gcw.nl/dissertation/s/3530/dis3530.pdf>.

**Vosti, S. A., and T. Reardon, editors.** 1997. Introduction. Pages 1-5 in S. A. Vosti and T. Reardon. *Sustainability, growth, and poverty alleviation: a policy and agroecological perspective.* Johns Hopkins University Press, Baltimore, Maryland, USA.

**Weisbuch, G.** 2000. Environment and institutions: a complex dynamical systems approach. *Ecological Economics* 34:381-391.

**Wilson, D. G., and B. D. Rudin.** 1992. Introduction to the IBM optimization subroutine library. *IBM Systems Journal* 31(1):4-10.

**Woelcke, J.** 2003. *Bio-economics of sustainable land management in Uganda.* F. Heidhues and J. von Braun, editors. Peter Lang, Frankfurt am Main, Germany.

## **APPENDIX 1. Software**

### **NOTE ON THE MP-MAS SOFTWARE**

MP-MAS was encoded in C++ by Thomas Berger with contributions from Pepijn Schreinemachers and Thorsten Arnold. It is freeware that can be downloaded from <http://www.uni-hohenheim.de/mas/software>. A manual is available electronically from the same location. The software is a single executable file that does not need installation. Both a Windows and a Unix version are available and the program runs on any personal computer. For very large models, the use of Unix OS is recommended as the software runs more stable.

MP-MAS uses mathematical programming software, which needs to be preinstalled. Currently, the Optimization Solutions Library (OSL) is used, which gives a high performance on very large models and can handle many integers (Wilson and Rudin 1992). IBM, the producer of OSL, has stopped the development of this software and transferred parts of the code to an open source community (<http://www.coin-or.org/resources.html>). This new solver platform, called COIN, is currently being implemented in MP-MAS to replace IBM-OSL.

The number of agents in the downloadable version is limited to 50. MP-MAS is not yet open source but interested academic software developers can join the group's effort and contribute to the software's development. People interested in applying the software to their own research area are also encouraged to get into touch and potential extensions of the software to a particular application can be discussed.

The software is constantly being improved: additional features are added with each new application, the input files contain more explanatory information, the error handling of the program is improved, and more powerful solvers (all freeware) are being integrated. Although the model has integrated many new features in recent years, empirical applications will often require additional ones.

### **USE OF THE MANUAL**

The purpose of the manual is to make MP-MAS more widely accessible and increase its use by other researchers. The MP-MAS does not have an own graphical user interface (GUI). Instead, Microsoft Excel workbooks are used to organize the data and to setup simulation experiments while the MP-MAS can be run using simple command line functions.

The advantage of using Microsoft Excel workbooks is that most users are familiar with it, that it is relatively easy to make changes, and that explanatory notes are easily inserted into the files. Workbooks are easily linked and can contain separate sheets for calculations and documentation of the model. The disadvantage is that 'small changes can have big consequences'; that is, accidentally entering a value in the wrong place can make the program crash.

To avoid this, the user is advised to start with a basic set of default input files that contain most of the features of the model and then to stepwise adjust this to the own application while trying to run the model at each step. For instance, the MP model can be gradually expanded to include more livestock and crop enterprises. Running the model after each significant change helps to locate possible errors more easily as the program's error calls can sometimes be cryptic and it is therefore difficult to pinpoint an error after making many changes at a time.

---