HUMAN DECISION MAKING IN EMPIRICAL AGENT-BASED MODELS: PITFALLS AND CAVEATS FOR LAND-USE CHANGE POLICIES

Grace B. Villamor
Center for Development Research
University of Bonn
53113 Bonn, Germany
Email: gracev@uni-bonn.de

Meine van Noordwijk
World Agroforestry Centre
Southeast Asian Regional Office
Bogor, Indonesia
Email: m.vannoordwijk@cgiar.org

Klaus G. Troitzsch
Institut für Wirtschafts- und Verwaltungsinformatik
Universität Koblenz-Landau
56016 Koblenz, Germany
Email: kgt@uni-koblenz.de

Paul L.G. Vlek
Center for Development Research
University of Bonn
53113 Bonn, Germany
Email: p.vlek@uni-bonn.de

KEYWORDS

Human decision making, empirical agent-based model, process-based, causality, cross-sectional data, decision algorithm

ABSTRACT

This paper describes three fundamental pitfalls or caveats of empirical modeling of land-use decision making in agent-based models for land-use/cover change. A case study in the villages of Jambi Province (Sumatra), Indonesia, is presented to demonstrate the construction of empirical decision-making models using utility functions while taking into account these caveats. Incorporating the decision process as an option to deal with the drawbacks of cross-sectional data is recommended to better specify agents' behavior in the decision-making models.

INTRODUCTION

Decision making in a social, economic and environmentally interacting context is a major research focus in agent-based/multi-agent simulation (AB/MAS) modeling. Indeed, simulating various decision-making processes in their interaction is one of its main advantages (Matthews et al. 2005). Hence, the AB/MAS framework has been increasingly applied for understanding coupled human-natural systems (An 2011), land-use dynamics and policies (Matthews et al. 2007; Parker et al. 2003), and recently in assessing ecosystem services tradeoffs driven by policy interventions (Villamor, 2012). Although there are many ways to explicitly formalize simple to complex human decision making (An 2011), utility-seeking agents using preference functions calibrated with econometric techniques are the most common framework for land-use change studies (Benenson and Torrens 2004; Parker et al. 2008; Villamor et al. 2011). Newer concepts of behavioral economy such as the prospect theory as a step beyond 'utility' have not yet been routinely incorporated (Kahneman and Tversky, 1979). However, one of the fundamental issues is to represent in a statistically consistent way a real-world situation of typically heterogeneous biophysical and socio-economic conditions (Berger and Schreinemachers 2006). This paper presents some challenges of modeling human decision making in the context of land-use change studies. It is subdivided into three main sections. Section 1 describes three often ignored caveats of empirical modeling of human decision making using utility functions, section 2 presents alternative ways to address these caveats, and section 3 demonstrates in a case study, the construction of a process-based decision-making model.

CAVEATS OF MODELING LAND-USE DECISION MAKING

Caveat 1: Causal relationship

Utility functions provide a formal framework for specifying agent-choice behavior (Benenson and Torrens 2004) in which a weight is attached to a particular choice amongst a set of choices or opportunities. In the context of land-use change studies, for instance, a farm household deciding whether to expand its farm plot will consider various factors such as the market price of certain crops and its own labor capital. With a wide range of possible choices, the household will assess each utility of the choices. The utility values are then transformed into choice probabilities. Statistical tools like logistic regression are commonly used to calculate probabilities and correlate particular actor attributes with specific land-use decisions either reported in a survey or observed from remotely sensed imagery (Evan et al. 2006). This approach identifies a statistically significant relationship between actor or

Proceedings 26th European Conference on Modelling and Simulation ©ECMS Klaus G. Troitzsch, Michael Möhring, Ulf Lotzmann (Editors)
ISBN: 978-0-9564944-4-3 / ISBN: 978-0-9564944-5-0 (CD)

landscape attributes and land-cover change. However, the results do not necessarily provide clear insight into actual decision process such as how an agent evaluates the benefits of a land-use change, the risks involved, and time frames considered for decision-making (Evans et al. 2006; Galvin et al. 2006). Even if the results obtained from simulation match those from the target, there may be some aspects of the target that the model cannot reproduce (Gilbert and Troitzsch 2005) and the underlying causes remain unknown (Galvin et al. 2006; Heckbert et al. 2010; Janssen and Ostrom 2006; Windrum et al. 2007).

Missing confounder

In a statistical context, caution should be taken when using observational studies alone to model agents' decision making, i.e., especially when predicting land-use decisions, since these kinds of models (i.e., the mere regress of observed decisions on observed data) "*do not carry the burden in the causal argument nor give much help in controlling for confounding variables*" (Freedman 2010, p.46). According to Rothman et al. (2008), given the observable nature of association measures, it is tempting to substitute them for effect measure and even more natural to give causal explanations for observed associations in terms of obvious differences. It is a well known textbook fact that - outside the realm of experiments - observed associations of variables do not automatically imply causality (Moore and McCabe 2004, p.160) - and therefore cannot establish a law that could predict outcomes. Confounders as defined by Rothman et al. (2008) are the factors (e.g., exposures, interventions, treatment) that explain or produce all or parts of the difference between the measure of association and the measure of effect that would be obtained with counterfactual ideals. The methods of scientific studies therefore need to control these factors to avoid a false positive error or an erroneous conclusion that the dependent variables are in a causal relationship with the independent variable. According to Freedman (2010), it would be better to rely on subject-matter expertise, and to exploit natural variation to mitigate confounding and rule out competing explanations.

Caveat 2: Drawing causal inferences from cross-sectional data

Due to the difficulty of collecting empirical data, the researchers have to rely on the data from cross-sectional surveys. Cross-sectional data are used to estimate parameters of functional forms of agents' decisions or other relations of variables. In most cases, there are no other justifications for the selected variables used in the model as they are available, fairly reasonable and are a good fit with the data.

Generally, cross-sectional data are snapshots, and the decision whether we observed a punctuated equilibrium or a stasis cannot be made by merely using one observation in time. Therefore, using only one point in time, we cannot be sure whether parameters or functional forms are stable in time. The functional form could also be incomplete to produce reliable results from interpolation if relevant confounders are not accounted for. In principle, this unavoidable drawback of using an observational study could only be solved by either using an experimental study or practically by using subject matter's knowledge to justify the functional form. In this case, we would have a model suitable for interpolation. If, at best, the functional form would model a "sufficient cause" then this would justify inter- and extrapolation, as causality solves the matter of confounding, and a modeled causality would be suitable for extrapolation. In fact, models are of little use if they can only be trusted for interpolation in known domains rather than for extrapolation to new conditions. The level of trust in model predictions for extrapolated settings will always require judgment on appropriateness of assumptions rather than proven numerical track records alone.

Caveat 3: Functional form as a compressed description

In using a functional form as a compressed description, initially we have a sample that determines n points in a k-dimensional interval built by the extremes of the k variables (sample intervals). Then we could decide to use either (i) only those points observed as starting values for a new simulation step (bootstrap approach), or (ii) start with points from very small spheres around each observed point (adding noise approach), or (iii) use any point within the sample interval (interpolation). It seems that if we use decent functions in our model and stay in the sample interval in any of the three forms above we should be safe. However, leaving the sample interval and extrapolating might lead to overshooting, heavy oscillations or generally to unrealistic constellations. For example, we might assume that a sample domain has indicated or forced a simple form (e.g., linear) where a sigmoid relation would be more appropriate. In such a case, leaving the sample interval would clearly lead to unrealistic overshoots or lower deviates.

ALTERNATIVE APPROACHES

The best way to address these caveats is through a choice modeling approach using longitudinal data. Longitudinal studies enable one to accurately observe changes or patterns. However, collecting longitudinal data is time consuming and very expensive. Besides, some argue that the power to detect causal relationship in this way is less compared to experiments. A researcher may decide to use census data, yet the empirical information on the decision-making process of subjects is also often lacking. Some studies employ genetic or evolutionary programming for agent's decision making in which, its computational processes are similar to those in natural selection theory (An 2011). There are few empirical studies on this but nevertheless promising in providing reliable results (Manson and Evans 2007).

Various sources in literature suggest the use of process-based decision making. For example, in dealing with the uncertainty of assumptions in models and data, an accepted way of reducing uncertainty or showing the influence of uncertainty processes on model results is by modeling the actual processes (Barthel et al. 2008). Process-based decision models, accordingly, are those capturing the triggers, options, and temporal and spatial aspects of an actor's reaction in a relatively direct, transparent and realistic way. Thus, substantial efforts should be invested in process-based decision-making mechanisms or models to better understand the socio-ecological systems (An 2011). In the case study below, a process-based decision making model is constructed based on the preferred future land-use choices as part of the decision process of the household agents in the study site. The decision process includes a time element that is pertinent for establishing causal relationships (van Belle 2008) derived from a cross-sectional survey.

CASE STUDY

Study area

The study area is located in Jambi Province (Sumatra), Indonesia. The villages of Lubuk Beringin, Laman Panjang, and Desa Buat, which cover a total area of 157 km^2 - are near the foothills of Kerinci Seblat National Park. Except for Desa Buat, these villages are considered poor and have poor access to market roads and electricity infrastructure due to their distance from the district center. Rubber agroforest is the dominant land use in the province and is the major rural livelihood of the people living there while paddy rice is the main food source. However, due to the low latex productivity from the rubber agroforests, farmers are now forced to convert their farm lands into more profitable land use such as oil palm and monoculture rubber plantations.

Household survey

A household survey was conducted to elicit the agents' characteristics and behavioral responses. The survey was conducted with 95 households (out of 551 households) between February and March 2010. In the survey questionnaires, two main conditions are explored, namely 1) the current condition of the agent, the household profile, and the farm-holding characteristics from which the current land-use choice was generated, and 2) under certain conditions or situations in which the agent will likely perceive and behave as if the condition existed (i.e., if supported by financial investment in the next 5 to 10 years, under payments for ecosystem services or PES through conservation agreement scheme). We also asked the reasons for choosing the land use in order to understand the actual motivations and preferences behind the decision.

Data analysis

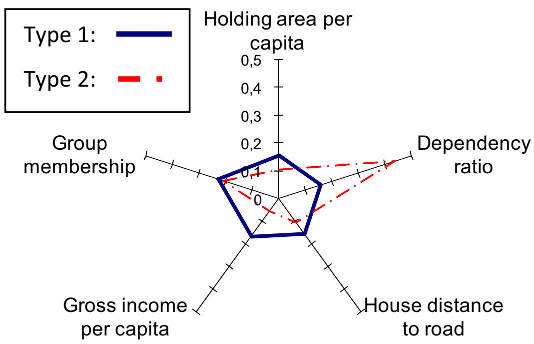
The data analysis was in two major steps: 1) household categorization using principal component analysis (PCA) and cluster analysis (i.e., K-means cluster or KCA), and 2) estimation of behavior of household types regarding land-use choices estimated using multinomial and binary logistic regression. We also compared the current land uses derived from the household survey to the land use maps of 1993 and 2002. Only the results of one household type are presented in the land-use choice section.

RESULTS

Household characterization

The summary of results derived from running PCA and KCA show that there are two types of households in the study area (Figure 1). These are rubber-rice households (type 1) as better-off households and rubber-based households (type 2) as relatively poor households, which explicitly shown in Figure 1.

Figure 1: Variation between household type 1 and 2 in terms of land holdings per capita, dependency ratio, and gross income per capita.



Land-use choices

Current land-use choices

The predictors and probabilities for land-use choices of household agents are summarized in Table 1 and Figure 2. There are 14 variables identified that significantly influence the choices of the household agent (Table 1).

Table 1: M-logit Model Estimation of Land-use Choices by Poor Households (type 2) that have Changed their Land Use between 1993 and 2005 (n = 74 plots); p = 0.000 and R^2 = 0.78.

Varia-ble	Definition	Rubber agroforest (β)	Paddy rice (β)
constant		36.56*	-271.54
$H_{\_age}$	Age of household head	-0.54**	-0.37
$H_{\_size}$	Household size	1.96	4.52**
$H_{\_dep}$	Household dependency ratio	-7.61**	8.98
$H_{\_edu}$	Education of household head	-4.13*	8.49
$H_{\_mem}$	Household number of memberships	2.26	10.77**
$H_{\_ACT}$	Household activities in regards to conservation agreement	-0.46	2.02*
$H_{\_CA}$	Household participation in conservation agreement	-2.59	-12.29*
$P_{\_wet}$	Plot wetness index	1.01**	4.28**
$P_{\_dt}$	Plot distance to town centre (m)	-2.88	-1.32*
$P_{\_dr}$	Plot distance to road (m)	2.62	-29.05*
$P_{\_F2}$	Neighborhood enrichment factor (NEF) of rubber agroforest	0.017**	0.19**
$P_{\_F45}$	NEF of other land uses	0.02*	0.36*
$P_{\_F6}$	NEF of rice field	0.001*	0.02*
$P_{\_F8}$	NEF of settlement	0.01**	0.09**

Note: ***, **, and * indicate statistical significance at the 0.01, 0.05 and 0.1 level, respectively. Other land uses (e.g., oil palm and rubber monoculture plantation) was selected as the base case for comparison.

Preferred land-use choices under certain condition

The predictors and probabilities for the land-use choice of household agents under the condition of "if supported by financial investment in the next 5 to 10 years" are summarized in Table 2 and Figure 3. There are 5 variables identified that significantly influence the choices of the household agent (Table 2).

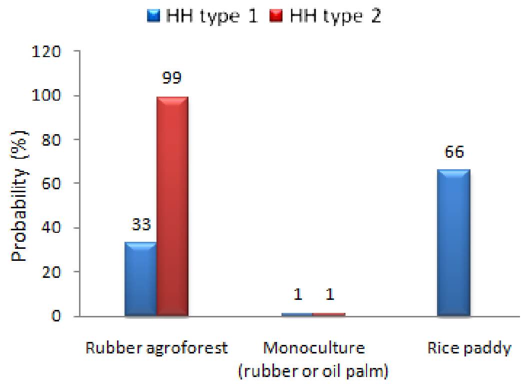


Figure 2: Probabilities for Current Land-use Choices

Table 2: Bi-logit Model Estimation of Land-use Choices by Poor Households (type 2) under Condition of Financial Support (n = 74 plots); p = 0.003 and R^2 = 0.29.

Variable	Definition	Rubber agroforest (β)
Constant		4.14**
$H_{\_age}$	Age of household head	-0.73*
$H_{\_edu}$	Education of household head	-1.12*
$H_{\_land}$	Household landholdings per person	0.67**
$H_{\_ACT}$	Household activities with regard to conservation agreement	0.17**
$P_{\_F45}$	NEF of other land uses	0.02*

Note: ***, **, and * indicate statistical significance at the 0.01, 0.05 and 0.1 level, respectively. Other land uses (e.g., oil palm and rubber monoculture plantation) was selected as the base case for comparison.

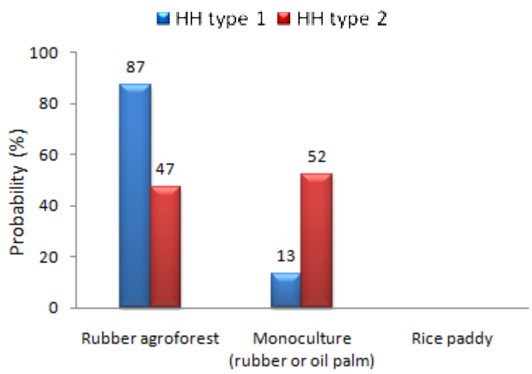


Figure 3: Probabilities of Preferred Land Use under Condition of Financial Support

The results suggest that under the current trend, 99% of the poor households are likely to stay with rubber agroforest. In terms of future land-use preferences if financial investments are provided, the probability to choose monoculture increases by 50% while the probability that the households will stay with rubber agroforest decreases by almost 50%. This could imply that half of the poor households would risk engaging in more profitable farming. It is also interesting that variables or factors influencing the land-use choice varies from current to future preferences.

It should also be noted that paddy rice was less preferred under condition of financial support (Figure 3) due to the fact that the survey was mainly done with male household heads, who are mainly responsible for rubber and oil palm production; females are solely responsible for rice production. Generally, only one decision maker was interviewed. Thus, the problem arises whether there is a gap between the expressed decision and the implementation of the expressed decision. The expressed intention could be just a wishful thinking, anticipated agreement with the other partner, or a decision that will be implemented without further consulting. In the given cultural environment, we could assume the latter two cases.

DISCUSSION

Choice heuristic algorithms

Based on the calculated choice probabilities, decision algorithms for the household agents were constructed. Using the decision-making routines of the Land Use Dynamic Simulator (LUDAS) model (Le et al. 2008), a decision-making choice algorithm for the current trend is presented in Figure 4. In the static phase, the land-use choice model under the current trend (Table 1) is specified. The same land-use choice model is also specified in the moving phase but with condition that the labor is more than zero and the land holding is equal to zero. If the functional parameters are estimated merely from cross-sectional samples (which is true in this case), we definitely cannot infer causality from the associated parameters (Caveat 1 and 2). At the same time we cannot do extrapolation (Caveat 3). Thus, we can only describe the association within the sample interval; we would otherwise risk unrealistic results. Nonetheless, we need to provide supporting information to justify the inferences from role-playing games, expert knowledge, and to identify missing confounders.

Drawing causal inferences from cross-sectional data can be done by incorporating time and confounding factors. The factors of the decision process itself can be seen as confounding factors between the observed socio-economic data and the production decision. Because such factors are associated with both the socio-economic status and the selected production, these factors could be the risk aversion of the decision maker, long-term decision, and alike. In order to estimate the central decision process more prospectively and so reduce confounding, one could estimate some parameters of the decision process directly. This leads us to the next decision algorithm (Figure 5).

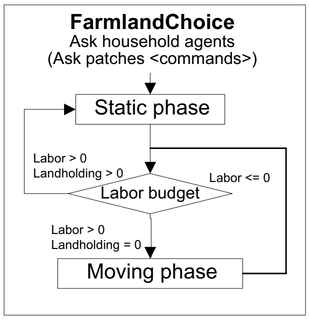


Figure 4: Flow chart of the basic decision-making routine for the farmland choice of the LUDAS model (Le et al. 2008).

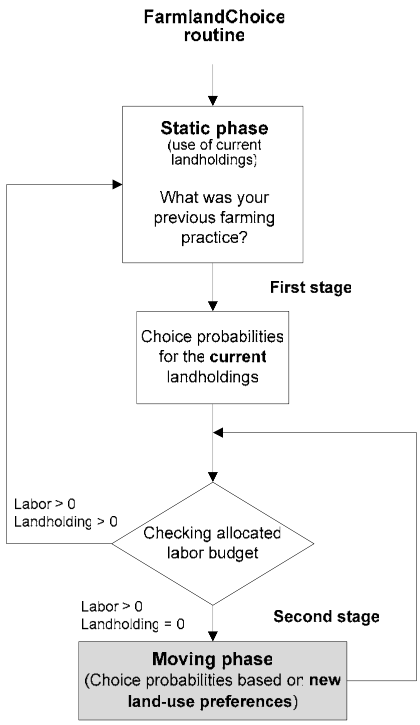


Figure 5: Flow chart of the process-based decision making routine under certain condition

A two-stage (or layered) decision-making routine was constructed to better incorporate the human behavior component (i.e., decision-making process). In Figure 5, the preferred land use under a certain condition (i.e., if supported by financial investment or subsidies for the next 5 to 10 years) is a new process-based decision-making routine. In the moving phase, the preferred land-use choice model (Table 2) is specified, while the same land-use choice (Table 1) is included in the static phase. In this way, the agents already have available capital (i.e., labor) to open new land as a form of resource-use efficiency (van Noordwijk et al. 2012). In the simulation, in spite of the use of cross-sectional data, inferences regarding the possible land-use change are justified by incorporation of the decision process (i.e., what will you do if supported by financial investment in the next 5 or 10 years). It also allows different decision strategies in different situations, which is consistent with the results in cognitive science (Gigerenzer and Selten, 2001). Although to an unknown extent, the confounding factor "risk aversion" of a given household under certain socio-economic conditions could also be captured.

One of the key issues addressed in the use of process-based decision making (as presented in the case study) is the introduction of a more adequate description of the underlying causal mechanisms, which is crucial in land-use and natural resources management decision making of household agents. By asking what agents will do or choose under certain conditions, including the temporal aspects (e.g., in the next 5 years), we could understand the 'sufficient cause' that would justify future choices. This also allows different decision strategies in different situations to be incorporated in the model. A multi-choice experiment household survey is being developed and designed as a recommended alternative to mend the caveats of using cross-sectional survey in empirical modeling of decision making (Villamor 2012).

CONCLUSIONS

This paper discusses the caveats of decision-making models and lists a number of fallacies (most of which have been discussed for decades in literature). The alternatives offered are experiments and longitudinal studies, which are often much too expensive for most projects. Experiments have their own drawbacks, as they exclude many of the aspects usually considered by decision makers; longitudinal studies might not find the same decision makers in subsequent waves of the study, as decision makers might have moved across the boundaries of the target region.

ACKNOWLEDGEMENTS:

We are grateful for the financial assistance from Federal Ministry for Economic Cooperation and Development (BMZ), German Academic Exchange (DAAD), and the Landscape Mosaic project of ICRAF. We also thank Guido Lüchters for the useful comments on the earlier version of this paper.

REFERENCES

An L. 2011. "Modelling human decisions in coupled human and natural systems: review of agent-based models". *Ecological Modelling* doi:101016/j.ecolmodel.2011.07.010.

Barthel R.; S. Janisch; N. Schwarz; A. Trifkovic; D. Nickel; C. Schulz and W. Mauser. 2008. "An integrated modelling framework for simulating regional-scale actor responses to global change in the water domain". *Environmental Modelling and Software* 23, 1095-1121.

Benenson I. and P. Torrens. 2004. *Geosimulation: automata-based modelling of urban phenomena*. John Willey & Sons New York, USA.

Berger T. and P. Schreinemachers. 2006. "Creating agents and landscapes for multiagent systems from random samples". *Ecology and Society* 11(2).

Evans T.P.; W. Sun and H. Kelley. 2006. "Spatially explicit experiments for the exploration of land-use decision-making dynamics". *International Journal of Geographical Information Science* 20, 1013-1037.

Freedman D.A. 2010. *Statistical models and causal inference*. Cambridge University Press, Cambridge.

Galvin K.A.; P.K. Thornton; J.R. de Pinho; J. Sunderland and R.B. Boone. 2006. "Integrated modeling and its potential for resolving conflicts between conservation and people in the rangelands of East Africa". *Human Ecology* 34, 155-183.

Gigerenzer G. and R. Selten. 2001. *Bounded rationality: the adaptive toolbox*. The MIT Press, Cambridge, MA.

Gilbert N. and K.G. Troitzsch. 2005. *Simulation for social scientists*. Open University Press, Berkshire.

Heckbert S.; T. Baynes and A. Reeson. 2010. "Agent-based modeling in ecological economics". *Annals of the New York Academy of Sciences* 1185, 39-53.

Janssen M. and E. Ostrom. 2006. "Empirically based, agent-based models". *Ecology and Society* 11, No.2 37.

Kahneman D. and A. Tversky. 1979. "Prospect theory: an analysis of decision under risk". *Econometrica* 47 No.2:263-92

Le, Q.B., S.J. Park, P.L.G. Vlek and A.B. Cremers 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,* 2:135-153.

Manson S.M. and T. Evans. 2007. "Agent-based modeling of deforestation in southern Yucatán, Mexico, and reforestation in the Midwest United States". Proceedings of the National Academy of Sciences 104: 20678-20683.

Matthews R.; N. Gilbert; A. Roach; G. Polhill; N. Gotts and L. Izquierdo. 2005. "Development of a rural economy and land use simulation modelling strategy". Macaulay Institute Craigiebuckler, Aberdeen, UK.

Matthews R.; N. Gilbert; A. Roach; J. Polhill and N. Gotts. 2007. "Agent-based land-use models: a review of applications". Landscape Ecology 22: 1447-1459.
Moore D.S. and G.P. McCabe. 2004. *Introduction to the practice of statistics*. Freeman, New York.
Parker D.C.; A. Hessl and S. Davis. 2008. "Complexity, land-use modelling, and the human dimension: fundamental challenges for mapping unknown outcome spaces". *Geoforum* 39: 789-804.
Parker D.C.; S.M. Manson; M.A. Janssen; M.J. Hoffmann and P.J. Deadman. 2003. "Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review". *Annals of the Association of American Geographers* 93: 314-337.
Rothman K.J.; S. Greenland; C. Poole and T.L. Lash. 2008. Causation and causal inference. *In* Rothman K. J., Greenland S. and Lash T. L. (eds.), *Modern Epidemiology*. Lippincott Williams & Wilkins, Philadelphia.
van Belle G. 2008. *Statistical rules of thumb*. Wiley, New York.
van Noordwijk M.; H.L. Tata; J. Xu; S. Dewi and P. Minang. 2012. Segregate or integrate for multifunctionality and sustained change through landscape agroforestry involving rubber in Indonesia and China. In: *Agroforestry: The Future of Global Landuse*. Nair PKR and Garrity DP (eds.), Springer, The Netherlands (in press)
Villamor, G., 2012. "Flexibility of multi-agent system models for rubber agroforest landscapes and social response to emerging reward mechanisms for ecosystem services in Sumatra, Indonesia". Unpublished manuscript, University of Bonn, Bonn.
Villamor G.B.; M. van Noordwijk; Q.B. Le; B. Lusiana, R. Matthews and Vlek P.L.G. 2011. "Diversity deficits in modelled landscape mosaics". *Ecological Informatics* 6: 73-82.
Windrum P.; G. Fagiolo and Moneta A. 2007. "Empirical validation of agent-based models: alternatives and prospects". *Journal of Articifial Societies and Social Simulations* 10 (2): 8.

AUTHOR BIOGRAPHIES

GRACE B. VILLAMOR was born in Bay, Laguna, Philippines. She is currently a researcher and PhD candidate at the Center for Development Research (ZEF), University of Bonn, Germany, and a research fellow at the World Agroforestry Center (ICRAF) in Indonesia. Prior to that, she was involved in various biodiversity-related research projects in Southeast Asia with the ASEAN Biodiversity Center and ICRAF. She obtained her Master degree at the Technical University in Dresden, Germany, in 2003.

MEINE VAN NOORDWIJK is the Chief Science Advisor of ICRAF. From 2002 to 2008, he was Regional Coordinator for Southeast Asia. Before joining ICRAF, he was a senior research officer at the Root Ecology Section at the DLO Institute for Soil Fertility Research in Haren, The Netherlands, focusing on models of the relationships between soil fertility, nutrient-use efficiency and root development of crops and trees.

KLAUS G. TROITZSCH has been professor of computer applications in the social sciences at the Computer Science Department of the University of Koblenz-Landau, Germany, since 1986. His main focus is on computational social science, in which he has published extensively. He has organized many workshops and conferences.

PAUL L.G. VLEK is the Executive Director of the West African Science Service Center for Climate and Adapted Land Use based in Ghana. He is also Director of the Department of Ecology and Natural Resources Management at ZEF and professor at the University of Bonn. Trained as a tropical soil scientist in The Netherlands and the USA, he conducts research on the sustainable use of natural resources in the tropics and how this is affected by or affects development processes.