Comparison of Two Common Empirical Methods to Model Land-Use Choices in a Multi-Agent System Simulation of Landscape Transition: Implication for a Hybrid Approach

Quang Bao Le[1] and Flávia F. Feitosa[2]

[1] *Swiss Federal Institute of Technolog Zurich (ETH Zurich), Natural and Social Science Interface (NSSI), Switzerland (quang.le.@env.ethz.ch)*
[2] *Earth System Science Center (CCST), National Institute for Space Research (INPE), Brazil (flavia@dpi.inpe.br)*

Abstract: Land-use choice routines embedded in a human-environment system (HES) model must meet more requirements than those in models typically presented in purely economic or psychological studies. This study compares the strengths and shortcomings of two common empirical methods - multi-nominal logistic (MNL) regression and classification tree (CT) analysis – for specifying land-use choices in a multi-agent system simulation framework (**L**and **U**se **D**yn**a**mics **S**imulator - LUDAS). First, we described design concepts of land-use decision-making mechanism in the LUDAS framework in which household's land-use choice is a component. Next, we compared two common methods for modeling the land-use choice with respect to pre-established criteria: a MNL model was specified to represent assumed rational behavior of human agents, while the CT model used a data-fit hierarchical rule set to represent heuristic process of reflex behavior. The study was conducted based on an intensive household-farm survey in a Central Vietnam's mountainous catchment. Based on the comparative analysis, we recommended explicit strategies for developing structurally realistic models that utilizes the complementarities of the both techniques to better represent bounded rational, yet adaptive, land-use choices in a HES model in the face of uncertainty.

Keywords: land use, human-environment system, adaptive choice analysis, multi-nominal logistic analysis, tree classification analysis, agent-based model, Vietnam

1   INTRODUCTION

The representation of land-use decisions that co-evolve with social-ecological environments over time and space remains as a key challenge of coupled human-environment system (HES) modeling for supporting sustainable land management. To address this challenge, land-use choice models embedded in HES must meet more requirements than those models typically presented in purely economic or psychological studies. As human decision process is genuinely causal, it is desirably built on causal relationships between decision's divers and activities. Causal decisions on resource uses are crucial for substantiating social-ecological feedback loops in HES that are important for understand system vulnerability and resilience. Thus, empirical formalism for agent-based processes in Multi-agent systems (MAS) should be based on causal analyses that identify relevant social and environmental determinants of land-use decisions and quantify their relative influence. Creditable and defensible causal analysis models reflecting the agent-based processes within MASs (mostly inferential statistics) are expected to have both *sufficient explanation capability* and *prediction power*. In agent-based models (ABM) for HES, land-use decisions are expected to be explained by a wide range

of socio-ecological drivers and, consequently, contribute to represent the co-adaptation and co-evolution between the human and environmental system [An and López-Carr 2012]. However, it is still challenging to obtain both interdisciplinary explanation and prediction power in a causal analysis.Moreover, ABM modelers often have to deal with the trade-off between needing an empirically-grounded decision model with easily measured data and having a model with strong theoretical foundations. The later requirement is particularly important, since the human decision-making is genuinely a cognitive and social process. However, it is difficult to measure psychological factors underlying the decisions of heterogeneous human agents, such as attitudes, perceived behavior control, personal and subjective norms. Given the wide range of methods available for the empirical modeling of land-use choices, comparative studies should be done to explore their potential and limitations in reasoning hybrid approaches able to provide a better comprehension of adaptive human behaviors in a HES context.

In this paper we firstly provide an overview of the decision-making mechanism of an existing multi-agent system model. In this overview, we emphasize conceptual aspects of land-use decision-making agents and how they stand in relation to the social-ecological feedback looks that ensure the co-adaptation and co-evolution of human and environment systems. Based on that, we identified important criteria required for land-use choice models to operationally represent the co-adaptation between individual behavior and social-ecological dynamics. Afterwards, we used these criteria to investigate the potential and shortcomings of two common empirical methods in land-use choice analysis.

2   DESIGN CONCEPT OF DECISION-MAKING MECHANISM IN LUDAS MODEL

LUDAS [Le 2005, Le et al. 2008, 2010] is a multi-agent system model for spatial-temporal simulation of a coupled human–landscape system. The model falls into the class of *all agents*, where the human population and the landscape environment are all self-organized interactive agents. The human community is represented by household agents that integrate household, environmental and policy information into their land-use decisions. The natural landscape was modelled as landscape agents representing land units that host natural processes and change their state in response to local conditions, exerting influence on the ecological condition of each unit of land and its immediate neighbourhood. Detailed description of the LUDAS model can be found in Le et al. [2008, 2010]. Here, we use the ODD+D protocol [Mueller et al. 2012] to describe the design concepts of the decision-making sub-model in the LUDAS model.

Theoretical Background: The Human–Environment System (HES) framework [Scholz 2011] can be used as an analytical guide for conceptually understanding the adaptive land-use decision-making modeled by LUDAS. It points out different types of environmental feedback loops that represent perception, evaluation, and adaptation of human systems regarding environmental changes. Following Piaget's theory in developmental psychology, the HES-framework defines adaptation of human decision-making to environmental change as the human agent's learning with respect to the adjustment of their decision rules, depending on their static internal model of the human–environmental interactions (i.e. behavioral program).

To represent adaptation, LUDAS followed the main assumption of social psychology that people often become aware of new behaviors by using information about the attractive behavior of others – so-called social processing/learning [Vygotsky 1978, Bandura 2001]. When people are uncertain about their decision outcomes, they tend to engage in social learning. Given high uncertainty, human agents with high need for satisfaction tend to imitate the behavior of others having similar characteristics [Jager et al. 2000]. As in other decision-making models, LUDAS represent this as a typology to economize cognitive efforts and minimize risks of failures with the alternative strategy. Imitation is an automatic social process, which can be explained by social learning theories. A fundamental principle of imitation is that the process is facilitated by favoring some similarity between the imitator and the group to be imitated. This similarity is often viewed in

terms of (1) social profile (e.g., altitudes, beliefs, education, and social status), and (2) owned resources or ability of accessing resources. 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 the imitation would be.

LUDAS used the Sustainable Livelihood Framework (SLF) [Bebbington 1999, Ashley and Carney 1999] for specifying livelihood-relevant variables of the household's decision-making model. The livelihood framework includes five core household asset categories: *human*, *social*, *financial*, *natural* and *physical* assets. This spectrum of livelihood assets is the basis of a farmer's capacity to generate new activities in response to needs and opportunities. LUDAS used this framework for selecting attributes that represent the livelihood typology of households. Based on this typology, the model uses livelihood similarity as a criterion to facilitate imitative learning: the social learning process is facilitated by favoring some similarity between the imitator and the group to be imitated. This theoretical (social-psychological) construct for the decision-making formalization in LUDAS was applied to a case study in Central Vietnam uplands and the results are presented in Le et al. [2012].

Individual Decision Making: LUDAS explicitly represents goal-seeking in the land-use decisions made by household agents. For that, 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, land-use decisions in LUDAS are bounded-rational rather than purely rational. Given high uncertainty of decision outcomes, this bounded optimization holds the risk that some households choose a land-use type that may not be the optimal alternative.
The household's land-use decision model takes inputs from the household profile (including attributes representing the five livelihood assets), its perceived landscape information, and sometimes information from other household agents (e.g. when checking land ownership of a given patch). The procedure is a logical process that includes both reflex and bounded-rational decision-making mechanisms. It assumes that household agents behave reactively according to production rules when deciding where to collect forest products, and that they seek optimal utility when looking for a location for cultivation. The decision procedure is universal with all household agents in terms of its logical sequence. However, because the agent's state, parameters and structure of utility functions are individual-specific, decision outcomes are diverse.

Interaction: In LUDAS, household 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 two (or more) households find their optimal 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.

Feedback Loops, Learning and Adaptation: Adaptive decision-making of household agents comprises primary and secondary feedback loop learning. The primary feedback loop involves direct information and material flows between household agents and their farm environment. Household agents perceive the biophysical state of their farm and the past performance of their production activities. They use that knowledge to anticipate and compare the benefits of different alternatives and, based on that, make their decisions. This primary feedback learning does not alter the goal-oriented decision rules of agents. The secondary feedback loop learning is defined by household-driven cumulative changes in household livelihood and farm environment on larger scales and in the longer term (possibly unintended), leading to the reframing of the agent's behavioral program. Household agents can change their behavior pattern (i.e., shift to another land-use choice model) by imitating the strategy of the most similar household group. Because secondary feedback loops involve changes in the agent's cognitive structures (i.e., internal behavior models), their functions may

induce qualitative changes in human actions (e.g. triggering the adoption of new classes of farming technology or new farm types) [see Le et al. 2012].

Individual Sensing: To evaluate land-use alternatives, household agents perceive the landscape characteristics and livelihoods of households in their surroundings. The extent of the limited land space that a household considers in its land-use decision, named here as landscape vision, is household-specific. Anticipated crop yield is also a household-specific variable as it is a function of both site condition (potentials of soil erosion and nutrient/water accumulation) and household's inputs (labor and agrochemicals).

Heterogeneity: Regarding land-use decision-making, LUDAS represents heterogeneity in three dimensions: multi-dimensional decision space (multiple alternatives x multiple farm locations), household-specific decision space (each household has its own decision space), household type-specific decision models.

Stochasticity: LUDAS uses stochasticity to represent (1) initialization of household population, (2) the locations of the land properties owned by the newly "born" households, (3) the preference coefficients of land-use choice functions and (4) some status variables not affected by agent-based process (all defined by even distribution and predefined bounds).

3   EMPIRICAL BACKGROUND AND TWO ALTERNATIVE METHODS FOR MODELLING LAND-USE CHOICES

3.1   Empirical Background and Data

The LUDAS model was calibrated for the Hong Ha commune, located in the forested mountains of central Vietnam. Data inputs include landscape and household attributes. Landscape data were land use/cover, terrain indices (slope, flow accumulation), proximities to rivers/streams and roads, holdings, village territory, and protection zoning. Household data cover socio-economic attributes (educational status, size, labor, land endowment, income) and household's access to policies or developmental programs [see Le 2005]. Household classification identified three distinct livelihood types in the study area, namely "paddy rice-based", "upland crops and livestock", and "off-farm and better-off" farmers. Characteristics of these livelihood types were shown in Le [2005]. In accordance with the above-mentioned design concepts (see section 2), a land-use choice model is statistically estimated for each livelihood type. In a LUDAS simulation run, household agents update/review micro and macro social-ecological patterns as a result of their earlier land-use decisions (social-ecological feedback loops are therefore incorporated) and decide what to do in the next time step.

3.2   Two alternative choice modelling methods

In separate to LUDAS simulation, we statistically estimated the livelihood type-specific choice models using two alternative methods that are commonly used in choice analysis: multi-nominal logistic (MNL) regression and classification tree (CT) analysis. MNL regression is specified to represent the rational behavior of human agents. The method involves the computation of a multi-attribute utility function for each 'land-use type x location' pair, then assumes random-utility maximization in the prediction of choices (see McFadden [1974]). In contrast, the CT approach is assumption-free. It is a non-parametric and non-linear statistic technique that helps discover a data-fit hierarchical rule set to represent heuristic process of reflex (*if-then*) behavior. In general, CT determines a set of *if-then* logical (split) conditions that permit accurate prediction or classification of categorical cases [Breiman et al. 1984, De'ath and Fabricius 2000, Vayssieres et al. 2000].

3.3   Criteria for comparative analysis

We compare the two methods using the following criteria:

- Assumption independency: This refers to the robustness of explanation and/or prediction [Levin 1966]. The criterion is considered in two contexts: the context when the choice models are estimated, and when the estimated choice models are used for generative simulation over time.
- Explanatory power: including (1) the capacity to representing the ordered and interactive effects among explanatory variables, (2) capacity for proving/inference social-ecological causalities (see causation criteria in Rothman et al. [2008]), (3) the ease for applying relevant theory, and (4) the potential for knowledge discovering.
- Vulnerability of the method with respect to data shortage
- Prediction power (percentage of correct prediction, determinant coefficient)
- Implementing in generative modeling: uncertainty representation and sensitivity to floating-point errors [Izquierdo and Polhill 2006].

4   RESULTS AND DISCUSSION

We conducted MNL regression and CT analyses for three livelihood household groups in Hong Ha. Here, we only show the results obtained for "upland crop and livestock" farmers, i.e. the largest group (Table 1 and Figure 1).

Table 1. MNL estimation of land-use choices by household type "upland crop and livestock farmers" (n = 165 plots), using non-agricultural land as a base case

Variable	Upland crop	Paddy rice	Fruit-based agroforestry
(constant)	3.027	6.511*	-4.350
	(4.315)	(4.464)	(5.064)
Environmental attributes of holding plots:			
Distance to roads (P_{road})	0.000	0.002	-0.006
Distance to house (P_{dhouse})	0.001*	0.001	-0.004***
Distance to rivers (P_{driver})	-0.023***	-0.028***	-0.004
Slope angle (P_{slope})	0.039	-0.233*	0.050
Wetness index ($P_{wetness}$)	0.180	0.084	0.241
Household characteristics:			
Age (H_{age})	-0.079*	-0.098**	-0.041
Leadership (H_{leader})	-2.385	-1.313	-1.855
Education status (H_{edu})	-2.012	-1.872	-2.925*
Labor availability (H_{labor})	0.692	0.639	-0.671
Dependency ratio (H_{depend})	1.994	1.674	3.323*
Holding/person ($H_{holding/pers}$)	0.000**	-0.001**	0.000
Income/person ($H_{income/pers}$)	-0.000	-0.000	-0.002*
Policy variables:			
Acces to extension (H_{extens})	0.166	-0.437	0.645
Fertilizer subsidy ($H_{subsidy}$)	0.171***	0.171***	0.171***

Fitness and accuracy assessment of the model:
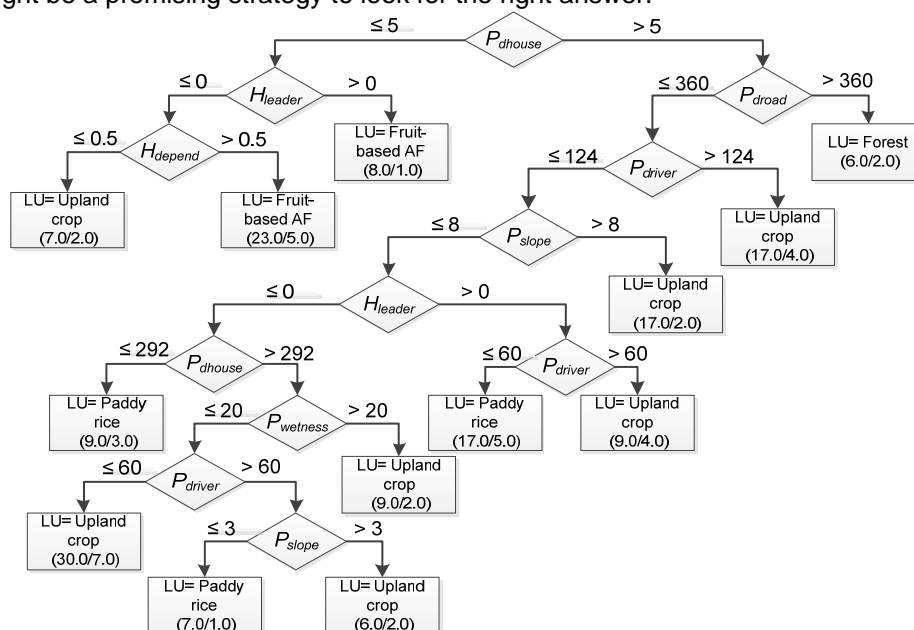Likelihood ratio (chi-square statistics): 226.730***   *df* = 42   *p* = 0.000
Pseudo r^2 = 0.676 (Nagelkerke); 0.609 (Cox and Snell); 0.406 (McFadden)
Percentage correct prediction: 73%

Note: Signs ***, **, and * indicate statistical significance at the 99%, 95%, and 90% levels, respectively.

Based on the results, our comparative assessment of strengths and weaknesses of the two empirical choice analysis methods are shown in Table 2. As MNL models depend on the assumption of rationality that may not correspond to real household behavior, it should be associated to some bounded-rational decision-making algorithm such as the ordered choice routine in LUDAS model [Benenson and Torrens 2004, Le et al. 2008, 2010]. Although a CT is assumption-free in its estimation, implementing a complicated decision-tree like the one presented in Figure 1 for a long simulation period (e.g. 3-5 decades) is likely not plausible. In addition, the structural complication of the CT is highly sensitive to the selection CT parameters (e.g. the minimal number of instances per tree leave). Which tree structures are parsimonious is also a difficult question to answer. Pattern-oriented

modeling (POM) process or robustness analysis [Railsback and Grimm 2012] might be a promising strategy to look for the right answer.

Figure 1. Classification tree of land-use choices by household type "upland crop and livestock farmers" (n = 165 plots), Tree "pruned" with confidence factor = 0.25, minimal number of instances/leave = 6. Percent of correctly predicted land-use types = 77.9%. Note: Meanings of variable abbreviations are shown in Table 1.

CT analyses were found to be useful for understanding the hierarchical and interactive effects of predictors in toward the prediction results, which cannot be provided by MNL. The CT method can reveal simple relationships between a few variables that could have easily gone unnoticed using parametric and general linear techniques like MNL. For instance, the variables of leadership (H_{leader}) and proximity to road (P_{droad}) are very important for determining household's land-use choices as the variables act as the first ordered factors (see Figure 1), which strongly corroborates with our field observations and household interviews (Le 2005]. We think that CT methods are particularly well-suited for data mining tasks in developing decision-making models embedded in HES. Here, there is little *a priori* knowledge or psychological theories that deal well with ecological variables such as surface slope, proximities to infrastructure and wetness.

CT method is found to be robust to datasets with small sizes and/or many qualitative variables (i.e. dummy variables), while MNL results are fragile to data shortage. With the "off-farm and better-off" farmers, i.e. the group with the smallest size (n=65 plots), MNL regression cannot be performed unless the two dummy variables of leadership, access to agricultural extension are omitted. Because these variables are actually important for triggering land-use choices, the causal effects captured by MNL become less robust.

The major limitation of CT methods is that it provides no statistic inference for causal effects in decision-making. It gives only a non-causal hierarchical association between explanatory variables and the choices [also see Vayssieres et al. 2000]. Therefore, there is likely no scientific evidence that CT-based decision-making models help to build social-ecological feedbacks, which are important in understand HES dynamics. One direction to overcome this limitation is to build Bayesian Networks (BN) based on the data-fit tree structure (which can be modified to be more theoretically sound) revealed by CT. With BN, it is possible to infer the land-use choices [Needham et al 2006]. The transition to BN also helps to improve the representation of uncertainty in the decision models, which is a weak point of CT (see Table 2).

Table 2. Strength (+) and weakness (-) of the two methods of choice analysis

Criteria	MNL regression	CT
Assumption independency		
• At the time the model estimated	Assumption of rationality, and linearity of utility function (-)	Assumption free (+)
• On the course of generative simulation	Less suffering of detailed hierarchical assumptions (+)	Assumption of fixed hierarchical structure (-)
Explanatory power		
• Capacity for representing ordered and interactive effects	No (-)	Strong (+)
• Capacity for proving causalities		
- Strength	Strong (+)	No (-)
- Consistency	Strong (+)	No (-)
- Plausibility	Strong (+)	Weak (-)
- Robustness	Sometimes weak (-)	No (-)
• Ease for applying existing theories	Easy, proactive (+) Structural implicit (-)	Difficult, passive (-) Structural explicit (+)
• Potential and type of pattern/knowledge discovering	Low (-) Confirmative (hypothesis testing) type	High (+) Data mining type
Representing uncertainty	Strong with choice probability (+)	No (-)
Vulnerability to data shortage	High (-)	Low (+)
Prediction power	Good (+)	Good (+)
Application in generative modeling (e.g. ABM)		
• Uncertainty representation	Good (+)	Weak (-)
• Sensitive to floating point error	High (-)	
Ease for stakeholder participation	Low (-)	High (+)

6 CONCLUSION

To be relevant with the HES modeling, decision-making sub-models in ABM should substantially help form social-ecological feedback loops across the human and environment systems. We showed a real-case, explicit and HES-relevant demonstration for comparative analysis of two common empirical methods, in line with two contradicting traditions: rational vs. heuristic approach, for modeling decisions in ABM's for natural resource use.

We conclude the following: (1) when the social-ecological dataset for model parameterization is small and/or contain many qualitative variables, the CT should be firstly used for revealing the variable relationships related to the resource-use choices. Then, it makes a strong sense to transit CT-based patterns to BN model for possible inferring the choices with better representation of uncertainty. (2) When the dataset is large, we recommend using both MNL and CT methods, then apply robustness analysis and/or pattern-oriented modelling (POM) [Railsback and Grimm, 2012] to identify a "structurally realistic" model that include key structural elements of real decision-making mechanism, but is with the most parsimonious representation. Our follow-up efforts in both directions are underway.

REFERENCES

An, L., López-Carr, D., Understanding human decisions in coupled natural and human systems*, Ecological Modelling*, 10.1016/j.ecolmodel.2011.10.023, 2012.

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

Bandura, A., Social cognitive theory: An agentic perspective, *Annual Review of Psychology*, 52, 1-26, 2001.

Bebbington, A., Capitals and capabilities: a framework for analyzing peasant viability, rural livelihoods and poverty, World Development, 27, 2021-2044, 1999.

Benenson, I., Torrens, P., *Geosimulation: Automata-based Modeling of Urban Phenomena*, John Willey & Sons, New York, USA, 2004

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J., *Classification and regression trees,* Monterey, Wadsworth & Brooks/Cole Advanced Books & Software, CA, 1984.

De'ath, G., Fabricius, K.E., Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology, 81(11), 3178–3192, 2000.

Izquierdo, L.R. and Polhill, J. G., Is your model susceptible to floating-point errors? Journal of Artificial Societies and Social Simulation, 9(4), 4 <http://jasss.soc.surrey.ac.uk/9/4/4.html>, 2006.

Jager, W., Janssen, M.A., De Vries, H.J.M., De Greef, J., Vlek, C.A.J., Behaviour in commons dilemmas: *Homo economicus* and *Homo psychologicus* in an ecological-economic model, *Ecological Economics*, 35, 357-379, 2000.

Le, Q.B., *Multi-agent System for Simulation of Land-use and Land-cover Change: A Theoretical Framework and Its First Implementation for an Upland Watershed in the Central Coast of Vietnam*. Ecology and Development Series 29, Cuvillier Verlag Göttingen, Göttingen, 2005.

Le, Q.B., Park, S.J., Vlek, P., 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, 2010.

Le, Q.B., Park, S.J., Vlek, P.L.G., Cremers, A.B., 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, 2008.

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

Levin, R., The strategy of model biulding in population biology, *Americant Scientist*, 54, 421-431, 1966.

McFadden, D., Conditional logit analysis of qualitative choice behavior, In: Zarembka, P. (Ed.), *Frontiers in Econometrics.* Academic Press, New York, pp. 105 – 142,1974.

Mueller, B. and 10 co-authors, Describing human decisions in agent-based social-ecological models - ODD+D an extension of the ODD protocol, Working Paper Prepared for A Workshop Session in iEMSs 2012, Leipzig, Germany, 2012.

Needham, C.J., Bradford, J.R., Bulpitt, A.J., Westhead, D.R., Inference in Bayesian networks. *Nature Biotechnology*, 24, 51-53, 2006.

Piaget, J. *The child's conception of the world.* Paladin, London, 1973.

Railsback, S., Grimm, V., Agent-based and Individual-based Modeling: A Practical Introduction. Princeton University Press, Princeton, NJ, USA, 2012.

Rothman, K.J., Greenland, S., Lash, T.L., *Modern Epidemiology (Third Edition)*, Lippincott, Williams & Wilkins, Philadenphia, PA, 2008.

Scholz, R.W., *Environmental Literacy in Science and Society: From Knowledge to Decisions*, Cambridge University Press, Cambridge, 2011.

Scholz, R.W., Tietje, O., *Embedded Case Study Methods: Integrating Quantitative and Qualitative Knowledge*, Sage Publication, Thousand Oaks, California, 2002.

Vayssières, M.P., Plant, R.E., Allen-Diaz, B.H., Classification trees: An alternative non-parametric approach for predicting species distributions. Journal of Vegetation Science, 11, 679-694, 2000.

Vygotsky, L.S., 1978. *Mind in Society*. Harvard University Press, Cambridge, MA.