Ecology and Development Series No. 29, 2005

Quang Bao Le

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

Cuvillier Verlag Göttingen

ABSTRACT

Land-use and land-cover change (LUCC) is an essential environmental process that should be monitored and projected to provide a basis for assessing alternatives for better land management policy. However, studies on LUCC processes are often challenged by the complex nature and unexpected behavior of both human drivers and natural constraints. A multi-agent simulation model (VN-LUDAS - Vietnam – Land Use Dynamics Simulator) was developed to model the interdependencies and feedback mechanisms between human agents and their environment. The aim of developing the model is to explore alternative policy scenarios to improve rural livelihoods and the environment, thereby providing stakeholders with support for making better-informed decisions about land resource management.

The VN-LUDAS model consists of four modules, which represent the main components of the coupled human-landscape system in forest margins. The human module defines specific behavioral patterns of farm households (i.e., household agents) in land-use decision making according to typological livelihood groups. The landscape module characterizes individual land patches (i.e., landscape agents) with multiple attributes representing the dynamics of crop and forest yields and land-use/cover transitions in response to both household behaviour and natural constraints. The policy module represents public policy factors that are assumed to be important for land-use choices. The decision-making module integrates household, environmental and policy information into land-use decisions. In the first development of the model, we nested the bounded-rational approach based on utility maximization using spatial multi-nominal logistic functions with heuristic rule-based techniques to represent the decision-making mechanisms of households regarding land use. The proposed agent-based architecture allows integration of diverse human, environmental and policy-related factors into farmers' decision making with respect to land use and presentation of subsequent accumulated outcomes in terms of spatiotemporally explicit patterns of the natural landscape and population. Although many features of the complex processes of human decision-making have not yet been included, the agent-based system has built-in flexibility for adaptation, upgrading and modification.

The developed model was applied to an upland watershed of about 100 km^2 in the A-Luoi district of the Central Coast of Vietnam. Spatially explicit data were obtained from LANDSAT ETM images, thematic maps, extensive forest inventory and intensive household surveys. Field data were used for calibrating the behavioural parameters of households and land patches, and to develop an initial database for simulations. Considered policy factors were watershed forest protection zoning, agricultural extension and agrochemical subsidies, which are the policy issues of local concerns (i.e., use cases) identified though interviewing local key informants and organizations. The model can potentially serve as a consistent tool to provide quick and relevant feedbacks in a form that allows stakeholders to revise and retest their ideas of policy interventions. Simulation outputs are spatiotemporally explicit, including multi-temporal land-use/cover maps of the landscape environment and basic socio-economic indices of the community at different aggregate levels of human/landscape agents. This enables efficient communication with stakeholders in land-use planning and management.

Preliminary simulation results for 10 different policy options suggest that reducing the current proportion of protected area from 90 % to 50 % and increasing the enforcement of protection, together with provision of extension services for a third of the total population and subsidizing 5 % of the population with agrochemicals ($ US 16 household^{-1} year^{-1}) would, on average, increase per capita gross income by 15 % and significantly reduce forest degradation compared to the current scenario (i.e., the policy setting in 2002). The simulated spatiotemporal data can be used for further analyses using standard GIS (geographic information system) and statistical packages. The simulated scenarios are rather scientific reasoning that provides information for stakeholders on policy options and their consequences.

Multi-Agent-System für die Simulation von Veränderungen in Landnutzung und Landbedeckung: Ein theoretischer Rahmen und erste Anwendung in einem Wassereinzugsgebiet im Hochland von Zentralvietnam

KURZFASSUNG

Landnutzungs- bzw. Landbedeckungsänderung (LUCC) ist ein außerordentlich wichtiger Umweltprozess, der überwacht und projektiert werden sollte, um eine Basis für die Abschätzung von Alternativen für eine bessere Landbewirtschaftungspolitik zu schaffen. Die komplexe Natur bzw. das unerwartete Verhalten der menschlichen Faktoren und natürlichen Randbedingungen stellen häufig eine große Herausforderung für die Untersuchungen der LUCC-Prozesse dar. Ein multi-agenten Simulationsmodel (VN-LUDAS - Vietnam - Land Use Dynamics Simulator) wurde entwickelt, um die gegenseitigen Abhängigkeiten bzw. die Rückkoppelungsmechanismen zwischen den Menschen und ihrer Umwelt zu modellieren. Das Ziel des Modells ist die Untersuchung alternativer Politikszenarien, die den Lebensunterhalt der ländlichen Bevölkerung und die Umwelt verbessern sollen. Dadurch soll den Beteiligten Entscheidungshilfen für das Management der ländlichen Ressourcen zur Verfügung gestellt werden.

Das VN-LUDAS-Model besteht aus vier Modulen, die die Hauptbestandteile des verknüpften Mensch-Landschaft-Systems der Waldrandzone darstellen. Das Modul ‚Mensch' definiert die spezifischen Verhaltensmuster der Farmhaushalte (Agent ‚Haushalt') bei der Entscheidungsfindung hinsichtlich Landnutzung entsprechend der typologischen Lebensunterhalt-Gruppen. Das Modul ‚Landschaft' charakterisiert einzelne Landflächen (Agent ‚Landschaft') mit Mehrfachattributen, die die Dynamik der Erträge der angebauten Pflanzen bzw. der Wälder sowie Landnutzungs-/Landbedeckungsübergänge als Antwort auf Haushaltsverhalten und natürlichen Randbedingungen darstellen. Das Modul ‚Politik' stellt Faktoren der öffentlichen Politik dar, von denen angenommen wird, dass sie für Landnutzungsentscheidungen von Bedeutung sind. Das Modul ‚Entscheidungsfindung' integriert die Informationen aus den Haushalten, aus der Umwelt und aus der Politik in Bezug auf Landnutzungsentscheidungen. Bei der ersten Entwicklung des Modells wurde der sogenanntet „bounded-rational“ Ansatz auf der Grundlage der Nutzungsmaximierung verschachtelt. Hierbei wurden räumliche multinominale logistische Funktionen mit heuristischen Techniken verwendet, um die Entscheidungsmechanismen der Haushalte hinsichtlich Landnutzung darzustellen. Die vorgeschlagene agenten-basierte Architektur erlaubt die Integration unterschiedlicher menschlicher, umweltrelevanter bzw. politischer Faktoren in die Entscheidungsfindung der Bauern sowie die Darstellung der anschließenden kumulierten Ergebnisse im Sinne von räumlich-zeitlich expliziten Mustern der natürlichen Landschaft bzw. der Bevölkerung. Obwohl zahlreiche Eigenschaften der komplexen Prozesse der menschlichen Entscheidungsfindung bisher noch nicht berücksichtigt wurden, besitzt das agenten-basierte System eine eingebaute Flexibilität für Anpassung, Ausbau bzw. Modifikation.

Das entwickelte Model wurde auf ein Hochland-Einzugsgebiet von ca. 100 km^2 im A-Luoi-Distrikt der Zentralküste von Vietnam angewandt. Räumlich explizite Daten wurden aus Landsat ETM Bildern, thematischen Karten, umfassenden Waldaufnahmen sowie Haushaltsbefragungen entnommen. Felddaten wurden eingesetzt, um die Verhaltensparameter der Haushalte bzw. der Landflächen zu kalibrieren und eine erste Datenbasis für die Simulation zu erhalten. Lokalpolitisch relevante Aspeckte (d.h., Anwendungs fall) waren der Schutz von Waldbereiche in den Wassereinzugsgebieten, landwirtschaftliche Beratung bzw. agrochemische Subventionen, die durch die Befragung von örtlichen ‚key informants' bzw. Organisationen ermittelt wurden. Das Model kann potenziell als konsistentes Instrument zur schnellen Ermittelung von Rückkoppelungen dienen, die es den Beteiligten ermöglicht, ihre Vorstellungen von Politikmaßnahmen zu überdenken und diese erneut zu testen. Die Ergebnisse der Simulation sind räumlich-zeitlich explizit und beinhalten multitemporale Landnutzungs-/Landbedeckungskarten bzw. grundlegende sozioökonomische Indizes der Gemeinschaft auf

verschiedenen aggregierten Ebenen der Mensch-/Landschaft-Agenten. Dies erlaubt eine effiziente Kommunikation mit den Beteiligten der Landnutzungsplanung bzw. des Landnutzungsmanagements.

Die ersten Simulationsergebnisse für 10 verschiedene Politikoptionen deuten darauf hin, dass die Reduzierung der geschützten Landflächen von 90 % auf 50 % bei gleichzeitiger strengerer Umsetzung der Schutzmaßnahmen zusammen mit der Bereitstellung von Beratungsleistungen für 30 % der Gesamtbevölkerung sowie die Subventionierung von 5 % der Bevölkerung mit Agrochemikalien (16 US$ Haushalt^{-1} Jahr^{-1}) das pro Kopf Bruttoeinkommen im Durchschnitt um 15 % erhöhen und die Walddegradation im Vergleich zum Stand 2002 deutlich reduzieren würden. Die simulierten räumlich-zeitlichen Daten können in weitere Analysen mit GIS (geografisches Informationssystem) bzw. statistischer Software eingesetzt werden. Die simulierten Szenarien liefern wissenschaftliche Daten als Information für die Beteiligten hinsichtlich der politischen Optionen und ihre Konsequenzen.

TABLE OF CONTENTS

1 MULTI-AGENT SYSTEM FOR SIMULATING LAND-USE/COVER CHANGE: A NEW MINDSET FOR AN OLD ISSUE

1.1 Background

1.1.1 The issue of land-use and land-cover change

Human alteration of the Earth is substantial and rapidly increasing. Change in land cover (i.e., biophysical attributes of the Earth's surface) caused by land use is the most substantial human-induced alteration of the Earth's system (Vitousek *et al*., 1997). Because land ecosystems are important sources and sinks of most biogeochemical and energy fluxes on earth, land-use and land-cover change (LUCC), when aggregated globally, significantly affect key aspects of the Earth system's functioning (Lambin *et al*., 2001). Between one-third and one-half of the land surface on earth has been transformed by human actions (Vitousek *et al*., 1997). These massive global changes alter major biogeochemical cycles, thereby contributing substantially to local and global climate change (Chase *et al*., 1999), including global warming (Houghton *et al*., 1999). LUCC also causes irreversible losses of biodiversity worldwide (Sala *et al*., 2000), and is a primary source of soil degradation (Tolba *et al*., 1992 cf. Lambin *et al*., 2001). Through modifying structures and functions of terrestrial ecosystems, LUCC significantly affects ecosystems' goods and services for human needs (Vitousek et al., 1997), subsequently influencing sustainable development.

Although not all of these impacts are negative, as some forms of LUCC in particularly developed regions are associated with continuing increases in food production or resource-use efficiency (Lambin *et al*., 2003), the overall LUCC on earth has been a main source of global environmental degradation (Turner *et al*., 1995; Lambin *et al*., 1999; Lambin *et al*., 2001). According to estimates, through the global expansion of croplands some 6 million km^2 of forests/woodlands and 4.7 million km^2 of savannas/grassland/steppes have been converted into agricutural land since 1850. Within these categories, respectively, 1.5 and 0.6 million km^2 of cropland have been abandoned (Ramankutty and Foley, 1999). According to the latest FAO assessment, from 1990-1995 there was a dramatic loss of 61.5 million hectares of tropical moist forests (i.e., the most diverse ecosystem in the world) in developing regions, while, at the same time, in developed countries the increase of forested areas was only 8.8 million hectares (Dolman and Verhagen, 2004). Modifications of land cover (i.e., changes in the structure over a short period), such as forest degradation caused by overexploitation, are also widespread (Archard *et al*., 2002).

1.1.2 The need to model LUCC processes for supporting proactive land management

Relevant understanding of LUCC phenomena and underlying processes are crucial in identifying successful strategies for mitigating the adverse impacts of LUCC and adapting to the changing environment (Vlek *et al*., 2003; Dolman and Verhagen, 2004). Rates and patterns of land-use change need to be understood to design appropriate biodiversity management. Areas of rapid LUCC need to be identified to focus land-use planning in the considered regions (Verburg *et al*., 2003). However, although the understanding of the rates and patterns of LUCC, based on the measurements of past phenomena, is important for monitoring land cover and land use, it is still merely an *ex post* evaluation of the land-use management, reflecting a reactive attitude to environmental degradation.

Our view about environmental management has shifted fundamentally from a reactive to a more proactive management strategy. "*Life affects its environment*" and "*environment constrains life*", two statements of Gaia theory (Lovelock and Margulis, 1974 cf. Lenton and van Oijen, 2002: 265) mean that environmental change and feedback are inevitable (Lenton and van Oijen, 2002), and that environmental damage, once done, is very difficult to undo. This implies that maintaining ecosystems in the face of changes requires active management for a foreseeable future (Vitousek, 1997). Accordingly, the understanding of LUCC has shifted from a reactive and condemning view, which often criticizes human impacts on the environment, to a *proactive* view, which focuses on proactive management of land resources to avoid irreversible mistakes (Victor and Ausubel, 2000; Lambin *et al*., 2003). Along with this viewpoint shift, the need for *ex ante* evaluation of policy options for proactive management of land resources becomes more urgent (Vlek *et al*., 2003; Costanza and Gottlieb, 1998).

Ex ante evaluations of policy interventions in the uses and management of land resources require a more robust understanding of processes constituting LUCC, in order to anticipate the changes under different intervention scenarios (Vlek *et al*., 2003).

Better data obtained from intensive monitoring alone are not enough for anticipation of future LUCC and its consequences unless causal mechanisms of the changes are better understood and modeled (Lambin *et al*., 1999). Improved understanding of controlling factors and feedback mechanisms in land-use systems is important for more reliable projections and more realistic scenarios of future changes (Veldkamp and Lambin, 2001; Lambin *et al*., 2003). These scenario studies provide a scientific knowledge that enables stakeholders, including policy makers, to proactively explore, discuss and examine potential outcomes (both benefits and costs) of different alternatives for intervention, thereby supporting policy-making processes for sustainable livelihoods and protecting the environment.

LUCC models are *reproducible* and *scientific reasoning tools* that can support the human's limited mental capacities in assessing land transformation and making more informed decisions about land resources management (Costanza and Ruth, 1998; Sterman, 2002). A model can be considered an abstraction of the real world, it should, however, be easy to understand and analytically manageable (Briassoulis, 2000). Because experimental manipulations or long-term studies for evaluating the performance of the complex human-environment systems are not possible or too costly, abstractive system models can help to fill the existing knowledge gaps (Costanza and Gottlieb, 1998; Sterman, 2002). LUCC models can offer a consistent and rigorous framework for identifying the scope of the problems, and highlight main causal loops within the system, thus enhancing our capacities in scientific reasoning about the likely outcomes in the future (Sterman, 2002). By clarifying and highlighting the main processes of land transformation, LUCC models can help to define environmental policy levers, i.e., points in the system where we should intervene to yield improved livelihoods and environmental qualities (Stave, 2002).

Most importantly, LUCC models can be used as *feedback tools* to facilitate learning and policy design. When rigorous LUCC simulation models are built and verified, they can serve as consistent tools to provide quick and relevant feedbacks in a form that allows stakeholders to revise and retest their ideas of interventions (Sterman, 1994). When stakeholders try the model and receive feedbacks about the likely effects of their tested interventions, their environmental learning (e.g., understanding and awareness of environmental consequences of actions) is also taking place. When the considered systems are complex, the discussions about how to solve a problem can bog down in disagreements about the likely effects of a given intervention. In this case, simulation models can act as a consistent feedback tool for scientific reasoning to enforce internal consensuses of actions (Forrester, 1987). In general, LUCC models can support policy decision-making processes by showing how our choices can affect the direction the future takes. Reflecting the overall importance of LUCC modeling in sustainable development studies, various LUCC models have been developed over the last few decades. Reviews of existing LUCC models are provided by Kaimowitz and Angelsen (1998), Briassoulis (2000), Veldkamp and Lambin (2001), and Agarwal *et al.* (2000).

Spatially explicit modeling is gaining awareness in LUCC studies. A model is called spatially explicit if a location is included in the representation of the system being modeled, and the model modifies the landscape on which it operates, i.e., spatial forms (e.g., maps) of a model's outputs are different to those of the model's inputs (Goodchild, 2001). Many reasons make spatially explicit modeling attractive in LUCC studies. A scientific reason is that many processes underlying land-use change are spatially dependent (Park *et al.*, 2003; Parker *et al.*, 2002). For example, land-use choices are constrained by biophysical factors that often vary across space. Furthermore, land-use capabilities often vary highly across space.

The most important reason for the increasing interest in spatially explicit LUCC modeling lies in the power of using spatial outputs for efficient communicating with stakeholders in land-use management and planning (Goodchild, 2001; Verburg *et al.*, 2003). This can help to improve participatory processes in research and development of land use and management. Spatially explicit representations of LUCC processes, e.g., the visual aids of Geographic Information System (GIS), are of very significant interest to the stakeholders, as most of them are not in a position to read technical papers/reports (Verburg *et al.*, 2003). At the community level, spatially explicit presentations of LUCC have also proven an appropriate means to support discussions with farmers about the distribution of resource bases, spatial interconnectivities between areas, and the consequences of local actions (Castella *et al.*, 2002a; Gonzalez, 2000; Rambaldi and Callosa 2000; Mather *et al.*, 1998; Smith *et al.*, 1999; Rambaldi *et al.*, 1998; Fox, 1995). At the policy decision-making level, spatially explicit presentations of LUCC modeling are suitable for communicating the results to policy makers (Verburg *et al*., 2003).

1.2 Problem analyses in LUCC modeling

As an old proverb states, “a problem stated is a problem half solved”. A rigorous analysis of the problems that earlier LUCC modeling has been confronted with is necessary before undertaking any modeling. Moreover, as many modeling methodologies and techniques exist, problem analyses will help us to select relevant modeling approaches, methodologies and techniques.

1.2.1 Complex nature of LUCC processes

The major challenge for achieving a better understanding of LUCC processes through modeling is the complex nature of the changes. Because land use is defined by the purposes for which humans exploit land cover, LUCC is obviously driven by complex interactions between biophysical and human factors over a range of scales in space and time (Parker *et al*., 2002; Verburg *et al*., 2003; Dolman and Verhagen, 2004). The *intrinsic complexity* of the coupled human-environment system underlying LUCC is characterized by the following aspects: (i) *nested hierarchies* of system components, (ii) *interdependencies* among system components, and (iii) *heterogeneities* of humans and their environment across time and space (Parker *et al*., 2002; Lenton and van Oijen, 2002; Eoyang and Berkas, 2002; Manson, 2001; Kohler, 2000). The following sections analyze these three aspects and subsequent problems in LUCC modeling.

Nested hierarchical structures and the problem of scale dependencies

The coupled human-environment system underlying LUCC is characterized by the *nested hierarchical structures* among the system components in space and time (Turner *et al*., 1995; Dumanski and Craswell, 1998; Verburg *et al*., 2003; Reynolds *et al*., 2003) (see Figure 1.1). A hierarchy is a partially ordered set of objects ranked according to asymmetric relations among these objects (Allen and Star, 1982; Shugart and Urban, 1988). The hierarchy theory suggests that a phenomenon at a certain level of scale (i.e., analyzed level) is explained by processes operating at the immediate lower level and constrained by processes operating at the immediate higher level, thus forming a "*constraint envelope*" in which the phenomenon or the analyzed process must remain (O'Neil *et al*., 1989: 195; Gibson *et al*., 2000: 225; Easterling and Kok, 2003: 269).

This means that a phenomenon such as LUCC is determined by factors at least at two different levels: above and below the level analyzed. The motions of driving factors in time and space are also different according to the differences of scale. The processes at the lower level are generally faster moving (shorter temporal extent) and lesser in spatial extent than the ones at the upper levels (Easterling and Kok, 2003). In other words, the behavior of any phenomenon, its causes and effects are scale dependent.

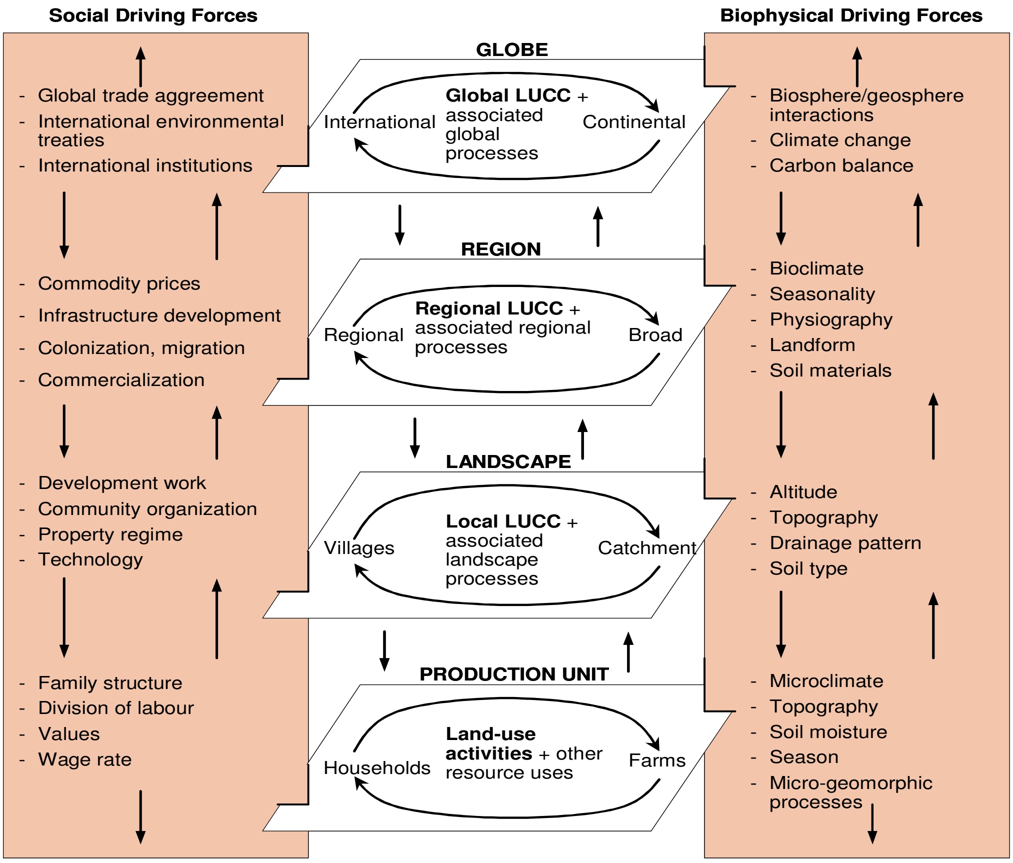


Figure 1.1 Land-use/cover change (LUCC) as the result of human-environment interactions over multiple scales in time and space. Sources: Adapted from Turner *et al.* (1995), Dumanski and Craswell (1998), and Verburg *et al.* (2003)

The reality of *scale dependencies* through the nested hierarchical structure of the human-environment system underlying LUCC suggests that straightforward aggregates of causes may not be sufficient to explain LUCC phenomena (Dumanski and Craswell, 1998; Lambin *et al*., 2003; Verburg *et al*., 2003). Unfortunately, many LUCC models are often operated at a single scale, which is usually selected arbitrarily or reasoned subjectively (Gibson *et al*., 2000) without considering the intrinsic differences in scale of the causal factors (Verburg *et al*., 2003). Some LUCC studies attempt to identify an optimal spatial scale or level of social organizations. However, because different processes underlying land-use change are important at different hierarchical levels, and the related criteria vary accordingly (Dumanski and Craswell, 1998). Land-use systems are likely never restricted to a single scale that can be regarded as optimal for measurements or predictions in the long term (Levin, 1992; Gardner, 1998; Geoghegan *et al*., 1998; Turner, 1999; Gibson *et al*., 2000; Verburg *et al*., 2003).

Another approach may be the tracing through the hierarchies to specify every causal relationship of land-use change for every scale and organizational level, as well as rules for translating information across scales (Turner *et al*., 1989). However, as the specification of causal relationships at each hierarchical level requires a specific dataset at such a scale (Dumanski and Craswell, 1998), it is very data demanding to formulate empirically all causal relationships of the complex nested hierarchical structure of the human-environment system. Furthermore, the mechanisms for transmitting cross-scale can be variable over time (Geoghegan *et al*., 1998). Therefore, even if all causal relationships are empirically grounded at a particular point in time, there is still no guarantee that such a full set of causal relations will still be maintained in the next time frame.

Functional interdependencies and feedback loops in LUCC processes: the problem of non-linear and transformative dynamics

Interdependencies always exist between all the components of the coupled human-environment system underlying LUCC, both between components within the organizational level (horizontal interplay) and between components of different levels of organization (vertical interplay), across time and space (Young, 2002 cf. Lambin *et al*., 2003). From the human side, land users may make their land-use decisions based on their land-use history and characteristics and surrounding biophysical environment. This leads to path dependencies and spatial interdependencies in land-use decision processes.

From the biophysical side, several spatially ecological interdependencies, such as slope processes, up- and down-stream effects, connectivity of natural habitats, ecological edge effects and forest gap dynamics, are crucial for the evolution of the coupled human-environment system, including LUCC (Parker *et al*., 2003).

The interdependencies among various causes of LUCC establish a causal web, i.e., one causal variable drives one or several others and vice versa (Turner, 1999; Lambin *et al*., 2003). Feedback loops carry materials, energy and information from one component to another (Bousquet and Le Page, 2004). These transforming feedback loops fuel the interdependence of the system by keeping the system components synchronized and interactive, serve to give both stability and changeability to the system, and support system evolution by providing impetus and resources for adaptation (Eoyang and Berkas, 1998; Manson, 2001).

Commonly, the landscape is taken to be in some kind of dynamic equilibrium: positive feedback loops exist and tend to amplify the land-use change (e.g., population growth often leads to rapid land-use/cover change), while some negative feedback loops co-exist and tend to counteract the change (e.g., institutional and improved land-use management may decrease the rate of adverse land-cover changes) (Lambin *et al*., 2003). Changes in driving forces can create disturbances in land ecosystems, but endogenous processes (e.g., vegetation growth/recovering) concurrently restore in part the system equilibrium (Geoghegan *et al*., 1998). The co-existence of buffering, amplification, and inversion of land transformation processes generate very non-linear dynamics in a land-use system, which have low predictability, high dimensionality, system openness, and dynamic (or far-from stable) equilibrium (Geoghegan *et al*., 1998; Eoyang and Berkas, 1998; Manson, 2001).

The reality of feedback loops among co-evolving components of the coupled human-environment system underlying LUCC challenges many assumptions of traditional LUCC models. Here, we point out the two main challenging points as follows:

First, there are *problems of multi-directional* and *endogenous causality* for statistical causal LUCC models, which follow the inductive approach. Many statistical LUCC models have the form: *LUCC = f(driving forces)*, where driving forces of LUCC (ranging from biophysical to socio-economic variables) are treated as exogenous causes of the change (see Lambin *et al*., 2003). The affecting directions of causes are assumed to be consistent across time, space and human agents. However, with the existence of feedback loops, the causality of a phenomenon becomes inconsistent or multi-directional (Eoyang and Berkas, 1998), i.e., a variable can be either exogenous (cause of the change) or endogenous (response to the change) to the land-use change (Lambin *et al*., 2003). For example, expansion of road networks can be a cause of rapid deforestation, but sometimes agricultural potential or development requirements of already deforested lands may lead to policy decisions to expand the road networks in these areas (see Lambin *et al*., 2003).

In a broader view, LUCC is a function of not only socio-economic and biophysical variables, but also of itself (Geoghehan *et al*., 1998). This actually means that, as the time scale of analysis expands, all causes of land-use change become endogenous to the human-environment system and are affected in some degree by previous land-use change (Lambin *et al*., 2003). The pathway of this effect is that temporally accumulative LUCC leads to significant impacts on the land ecosystem goods and services, consequently affecting human livelihoods and other socio-economic conditions, and thus creating new opportunities and constraints for future land use (Lambin *et al*., 2003).

Second, when interdependencies combine with the complicated nested hierarchical structure of the coupled human-environment system, feedback loops become enormous, creating the *problem of tractability* for any purely analytical LUCC model. A purely analytical/mathematical LUCC model, e.g., system dynamics models, describes the system using a causal loops diagram, which maps explicitly all possible interdependencies among possible causes and is represented by a complete set of differential equations (Forrester, 1980; Gilbert and Troitzsch, 1999). For instance, full representation of a system of 2 objects requires 4 equations: 2 to describe how each object behaves by itself ("*isolated*" behavior equation), 1 to describe the interaction between the two objects ("*interaction*" equation), and 1 to describe how the system behaves without the objects ("*field*" equation). In general, the number of required equations is defined by the "power law of computation": 2^n, where n is the number of objects in the system (Easterling and Kok, 2003: 275). If a system has 10 objects, the number of differential equations needed is $2^{10} = 1024$. The complex land-use system often consists of many times more than 10 objects, thus the number of necessary equations is enormous, and it is extremely difficult to trace and specify causes and causal relationships beforehand. It is even more unfeasible if modelers want to represent spatial relationships using differential equations (Sklar and Costanza, 1991).

Social and biophysical heterogeneity

Biophysical environment and socio-economic sub-systems underlying LUCC are often heterogeneous over time and space (Park *et al*., 2003; Parker *et al*., 2003; Lambin *et al*., 2003). Heterogeneities of both human and biophysical conditions are realized as critical drivers of LUCC outcome. Land users are usually different in their resources, values, abilities and experiences. As these factors of differences are crucial inputs in land-use decision-making processes (Parker *et al*., 2002), such social diversities potentially result in different land-use patterns. These diversities often change over time due to changes in production, demography, and learning processes. From the biophysical side, different locations have potentially different conditions of topography, soil, water availability, vegetation, accessibility to market, and so forth, and consequently have different capabilities for land use and or natural vegetation growth/recovery. The spatial heterogeneity of land-use capability often creates socio-economic incentives or opportunities for land development in particular localities, e.g., areas along roads or sub-urban areas, leading to rapid changes at such localities, so-called *hotspots* (Park *et al*., 2003).

When heterogeneity and interdependencies are combined (i.e., fine-scale processes are interconnected), sudden changes and radical flips may occur between alternate stable states in the system (Eoyang and Berkas, 1998; Geoghegan *et al*., 1998; Parker *et al*., 2002). Hence, the macroscopic properties of the coupled human-environment system, e.g., LUCC landscape patterns, become irregular and rugged, and do not follow a progressively smooth pattern as normally delineated by conventional statistic or analytic models (Eoyang and Berkas, 1998; Parker *et al*., 2003). In a system with such dynamics, many *punctuated equilibriums* (or *bifurcations*) can exist (Eoyang and Berkas, 1998), and the system is very sensitive to the initial conditions. In other words, *path dependencies* become important for system behavior.

Arthur (1989), cf. Geoghegan *et al*. (1998), notes that a path-dependent system may exhibit several properties that must be considered in LUCC modeling and assessment, such as: *variable predictability* (i.e., unpredictability followed by high predictability), *non-ergodicity* (i.e., small perturbations may significantly influence long-term development, historical events are not averaged and as important as a driving force). Sudden or irregular changes in the equilibrium in a complex system dramatically reduce the qualification of prediction (Eoyang and Berkas, 1998; Manson, 2001). Unfortunately, temporally explicit land-use dynamics have been given much less attention than spatial dynamics (Verburg *et al*., 2002).

LUCC as an emergent property of the coupled human-environment system

By definition, emergence phenomena cannot be reduced to the system's parts: the whole is more than the sum of its parts because of interactions among the parts (Figure 1.2a) (Parker *et al*., 2002). Therefore, emergence phenomena are directly related to the phenomenon of nested hierarchies and interdependencies that characterize the complex system as portrayed above.

LUCC is an emergent property that evolves from the interactions among various components of the entire human-environment system, which themselves feed back to influence the subsequent development of those interactions (Stafford-Smith and Reynolds, 2002; Reynolds *et al*., 2003). At the scale of the system's constituent units (e.g., household and land plot), many small changes in land allocation or natural vegetation growth occur, reinforce or cancel each other. These short-term and localized changes are the results of multiple decisions made by individual actors, who act under certain specific conditions, anticipate future outcomes of their decisions, and adapt their behavior to changes in their external and internal conditions (Lambin *et al*., 2003) (see Figure 1.2b). In most cases, these decisions are made without any central direction. Temporal accumulations of these short-term changes and spatial aggregations of these localized changes generate continuously emergent patterns of both LUCC at the landscape scale and socio-economic dynamics at the population scale (e.g., village). The existence of nested hierarchical structures, interdependencies, heterogeneities and co-evolutions of different system components transfer the landscape into a highly non-linear and far-from-equilibrium state (Lambin *et al*., 2003). The changes of macro phenomena, such as LUCC, socio-economic dynamics of the population, and possible policy intervention, feed back to influence the behavior of individuals that produce them (see Figure 1.2a and 1.2b).

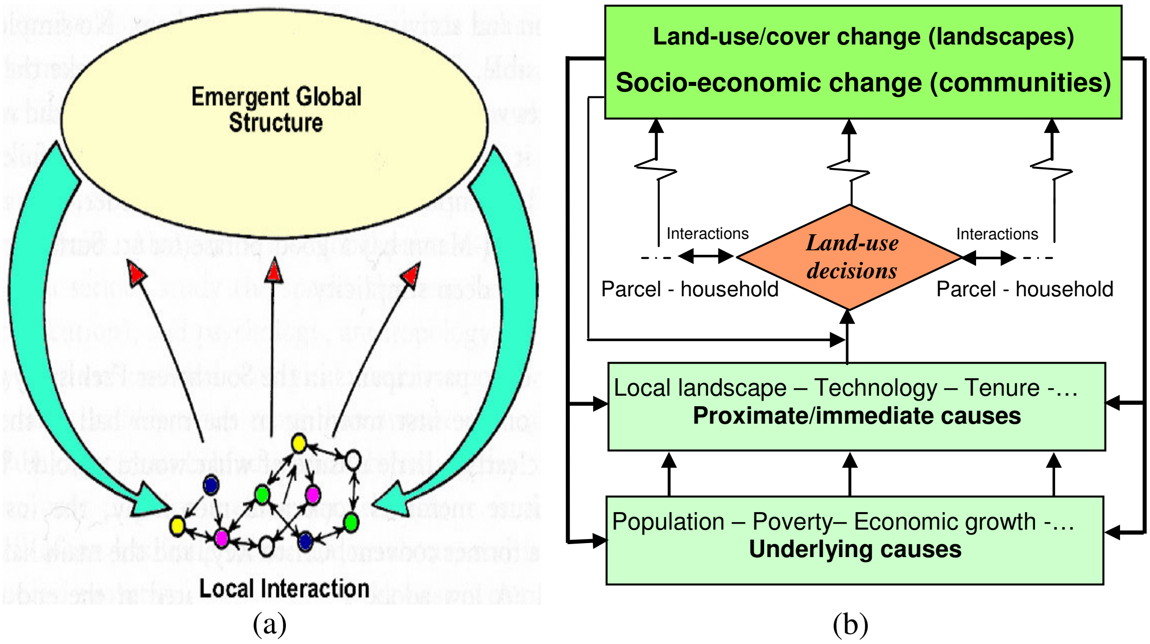


Figure 1.2 (a) Micro-macro feedback loop in a complex system as emergent phenomenon. Source: ALC (2003)
(b) Land-use/cover changes at landscape level as an emergent phenomenon generated from interactions of land-use decisions at farm/household level. Source: modified from Kaimowitz and Angelsen (1998)

If LUCC is intrinsically an emergent phenomenon of the underlying complex adaptive systems, its exact future is almost unpredictable if the system parts are examined in isolation (Lambin *et al.*, 2003; Batty, 2001). Because emergent phenomena can be decoupled from the properties of component parts (Bonabeau, 2002), the exact future of emergent phenomena is difficult, even not possible, to predict (Bonabeau, 2002; Batty and Torrens, 2001). As emergent properties arise from micro interactions, capturing them deals with cross-scale interactive processes, which are difficult to address using purely analytical or statistic methods. Therefore, models of complex ecosystems, as Peter (1991: 116) perceived, "*are no longer touted as predictive models but as heuristic devices to explore the logical implications of certain assumptions*". Similarly, land-use transitions are also intrinsically multiple and reversible dynamics, which are in neither fixed and deterministic nor predetermined patterns. The concept of land-use transition should be perceived as "*possible development paths where the direction, size and speed can be influenced through policy and specific circumstances*" (Martens and Rotmans, 2002 cf. Lambin *et al*., 2003).

1.2.2 The need for an integrated framework for modeling LUCC

Although understanding of the complex nature of LUCC processes has been conceptually achieved, this improved understanding has not yet been adequately reflected in LUCC models (Lambin *et al*., 2003). LUCC modeling faces the challenge of producing a modeling framework that enables integration of social and biophysical systems across time and space, as well as meeting the diversity of stakeholders in policy formulating processes.

Discrepancies in level and disciplines in previous LUCC modeling: the problem of integration

In spite of the improved understanding of the complex and connected nature of LUCC processes, the discrepancies between LUCC modeling studies by human and biophysical disciplines are obvious (Veldkamp and Verburg, 2004; Lambin *et al*., 2003; Huigen, 2004). Researchers in social/economic sciences traditionally study individual human behavior at the micro-level (i.e., households and farms) using qualitative or quantitative models of microeconomics and social physiology (Veldkamp and Verburg, 2004). These studies emphasize the micro-structures of land-use actors and interactions among them (Huigen, 2004), thus often yielding explicit understanding about causal processes of land-use change at the farm level. However, difficulties arise when scaling these models up to higher aggregation levels (Jansen and Stoorvogel, 1998; Verburg *et al*., 2002).

Natural scientists, e.g., geographers and ecologists, often focus more on the examination of LUCC patterns at the landscape and regional scales, which are measured in spatially explicit ways (e.g., remote sensing and GIS), in correlation with macro-properties of socio-economic and biophysical driving factors using multivariate statistics (Veldkamp and Verburg, 2004). Although the selection of driving variables may be based on regional economic theories, statistical relationships themselves are not necessarily understood as causal relationships (Verburg *et al*., 2003; Huigen, 2004).

Thus, causal processes are not captured in an explicit way using these spatial statistical LUCC models.

Due to the poor connections between spatially explicit and socio-economic approaches in land studies, there is a "*general poverty*" in real integrated human-environment approaches in LUCC research (Nagendra *et al*., 2004 cf. Veldkamp and Verburg, 2004: 1). Our understanding of the integrative LUCC processes has significantly improved over the last few decades, but this conceptual understanding has not been adequately integrated into the modeling of the processes yet (Lambin *et al*., 2003). Thus, there is an increasing demand to develop reproducible and rigorous integrated frameworks for modeling LUCC (Vlek *et al*., 2003). Such integration has added value compared to disciplinary approaches when feedbacks and interactions between subcomponents of the coupled human-environment system are explicitly addressed (Verburg *et al*., 2002). Again, the complex nature of the human and environmental systems is a challenge to do so.

Diversity of stakeholders in land management and policy formulation: the problem of flexibility required for LUCC models

In concert with the complex dynamics of land-use change processes, the diversity of stakeholders and their changing values cause great difficulties in formulating effective and relevant land management policy. There is a range of stakeholders in land-use and management, who have different perspectives, goals and interests (Haggith *et al*., 2003; Stave, 2002). For instance, governmental bodies of nature conservation are mainly concerned about deforestation and biodiversity decline, whereas agricultural or rural developers and local communities often pay attention to the improvement of local livelihoods by promoting agricultural production. Farmers may be most interested in income generating activities and what actually happens on their farms, and do not care much about land-use change at landscape scales, while regional planners may be interested in overall trends of landscape patterns. Due to the conflicts among stakeholder values, in part with the diversity of the targeted land-use systems, it is really difficult to formulate land management policies that are relevant to all stakeholders (Korfmarcher, 2001).

Nevertheless, people can change the way they take decisions based on their *learning processes* (Dietz and Stern, 1998), which is the premise for multi-stakeholder negotiations to reach consensus about environmental decisions and actions (Stave, 2002; Sterman, 2002). Furthermore, there is no learning without feedback or without knowledge of the results of our actions (Meadows, 2000; Sterman, 2002). This implies that effective processes of multi-stakeholders negotiations critically require certain tools/models that enable them to quickly generate feedbacks from the environment as the consequences of their supposed interventions. Very often, stakeholders like to explore likely environmental and livelihood outcomes of different scenarios of inputs, i.e., to answer numerous *what-if* questions (Sterman, 2002; Korfmarcher, 2001; Stave, 2002).

Unfortunately, neither traditional models of LUCC nor participatory exercises (alone) are adequately flexible for supporting these learning processes. Empirical (statistical) LUCC models are valid only within the data range of the land-use change on which they are based, and are thus not suitable for scenario studies (Verburg *et al.*, 2002). System dynamics models may have more flexibility than empirical statistical models; however, the fixed and strong links/coupling between system components may make the model *highly fragile* with respect to structural modifications. Participatory approaches (alone) rely heavily on information campaigns, facilitated discussions, stories recording, and public hearings for conveying information and capturing stakeholder inputs. Although these participatory exercises have their own merits, they are too vague for anticipating explicit landscape and community outcomes for policy considerations.

1.2.3 Problem statements

As analyzed above, there are two main gaps in the knowledge on LUCC modeling that need to be filled:

- An integrated framework for representing LUCC processes as emergent phenomena of the underlying human-environment system.
- The translation of that integrated framework into a spatio-temporally explicit modeling prototype, which understandably represents the complexity of the land-use transition processes and scientifically supports

stakeholders to make more informed decisions about land resources management.

There are a number of schools of thought and many modeling paradigms for addressing these problems. The analyses below will identify the modeling approach that will help to sharpen our modeling objectives.

1.3 The Multi-Agent System (MAS) for simulating LUCC: the paradigm shift

A promising novel approach to modeling the complex LUCC processes is the *multi-agent systems* (MAS) for simulating LUCC (MAS-LUCC). MAS has been recognized as a useful tool for building a sound theoretical framework to deal with the complexity of LUCC (van der Veen and Otter, 2001; Bousque and Le Page, 2004) and to more efficiently support environmental decision-making processes (Ligtenberg *et al*., 2004; Barreteau *et al*., 2001). The development and application of MAS in different fields, including LUCC, is associated with the progress of the complexity theory and is a rich breeding ground for the interdisciplinary movement (Bousquet and Le Page, 2004). Rather than a technology, MAS is a mindset for viewing and representing the complex system. Because MAS is conceptually complex, understanding this concept as a paradigm shift of system research and management is necessary to avoid its improper use (Bonabeau, 2002).

1.3.1 Traditional approaches in LUCC modeling

It is useful to begin with the logic of the standard modeling approaches based on system typology. There are three types of systems (Weaver, 1948 cf. O'Neil *et al*., 1989; Weinberg, 1975 cf. O'Neil *et al*., 1989; Lenton and van Oijen, 2002; Easterling and Kok, 2003) (see Figure 1.3):

Small-number systems (cf. O'Neil *et al*., 1989), i.e., *organized simple systems* (cf. Easterling and Kok, 2003) or *ordered systems* (cf. Lenton and van Oijen, 2002), comprise only few components (Figure 1.3, domain I), whose interaction mechanisms can be easily tracked and analytically proven by a complete set of mathematic equations. Thus, system behavior is adequately represented by *analytic models* (i.e., *equation-based* or *mathematic models*) in which the targeted patterns/conclusions are *deductively* inferred from proved assumptions. In deduction, assumptions contain all possible elements of the model (e.g., premises, axioms, definitions and proved causal relationships); thus, the validity of deductive inference is totally contained in the set of assumptions (Werker and Brenner, 2004).

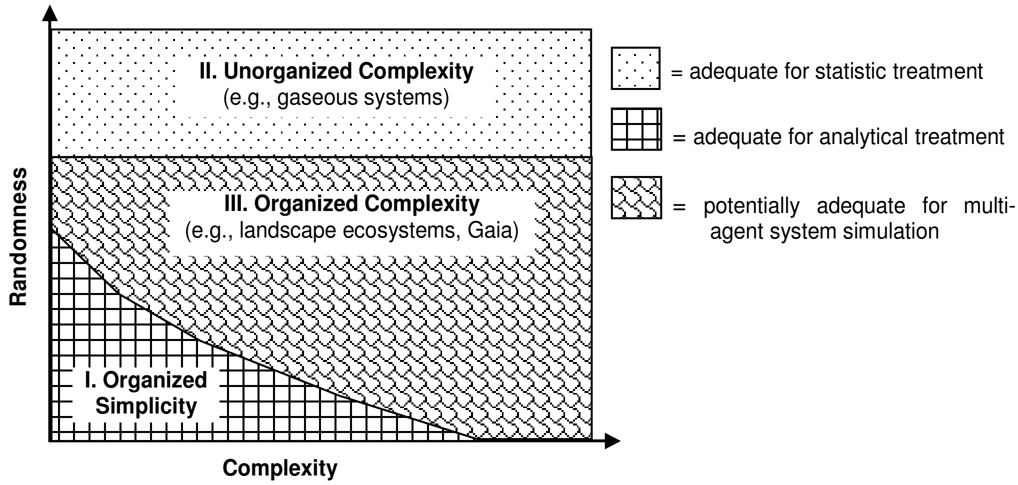


Figure 1.3 Complexity versus randomness and three complexity domains. Sources: after Weinberg (1975), O'Neil *et al.* (1989), Lenton and van Oijen (2002), Parker *et al.* (2003), Easterling and Kok (2003). Used terminologies are after Easterling and Kok (2002)

Large-number systems (cf. O'Neil *et al.*, 1989), i.e., *unorganized complex systems* (cf. Easterling and Kok, 2003) or *chaotic systems* (cf. Lenton and van Oijen, 2002), contain an extremely large number of identical components (Figure 1.3, domain II), which interact randomly (i.e., lack of structure) such that the system's probabilistic properties appear to be deterministic, obeying the *law of large numbers*. Thus, system behavior is adequately represented by *statistical models* (e.g., regression models) in which the targeted patterns/conclusions are *inductively* inferred from *data* and thus sometimes called *data-driven models*. The validity of inductive inference is contained in the data with which the model is built, not in the assumptions[1] (Werker and Brenner, 2004). To satisfy the law of large numbers in statistics, the size of the used dataset is often expected to be as large as possible.

Medium-number systems (cf. O'Neil *et al.*, 1989), i.e., *organized complex systems* (cf. Easterling and Kok, 2003) or *critical systems* (cf. Lenton and van Oijen,

[1] Assumptions in inductive models do exist, but are rather purely statistical premises than causal assumptions.

2002), lie between the domains of organized/structural simplicity and unorganized/chaotic complexity (Figure 1.3, domain III), and likely reside on "*the edge of chaos*" (Lenton and van Oijen, 2002; Waldrop, 1992). These systems are too complex and intractable for analytical solutions, but still too structured and organized for purely statistical treatments. Unfortunately, land-use systems at landscape level fall into this organized complexity domain (O'Neil *et al*., 1989; Easterling and Kok, 2002). Our analysis of the complex nature of LUCC processes as above (Section 1.2.1) has clearly illustrated this problem. It is widely recognized that Multi-Agent System Simulation (MASS) is a new paradigm to study this organized complex system (Axelrod, 2003; Parker *et al*., 2002; van der Veen and Otter, 2001).

1.3.2 Multi-Agent System Simulation (MASS) for studying complex adaptive systems

The philosophy that MASS appears as a new modeling paradigm lies in the two fundamental concepts contained within its name: i) *Simulation* as the third way of doing science, and ii) the *Multi-Agent System* (MAS) as an alternative for representing complex adaptive systems. We discuss these two concepts and their relevance to the study of complex adaptive system in the following.

Simulation as the third way of doing science

Simulation is a third way of doing science, besides the two standard methods of deduction and induction. "*Simulation is the process of designing a model of a real system and conducting experiments with this model for the purpose of understanding the behavior of the system and/or evaluating different strategies for the operation of the system*" (Shannon, 1998: 7). From this definition, there are three key aspects reflecting the logic of simulations as a way doing science: i) an imitation of the real system (i.e., system representation), ii) an artificial/virtual experimental approach to study a problem, and iii) decision support as the overriding purpose.

First, simulation is an *abstractive imitation of the real system* (of interest) *as it progresses through time* (Robinson, 2003). This means the simulation model should be designed and operated in such a way that it *mimicks* the structure and motion of the real system (Shannon, 1998). This implies that simulation firstly focuses on the *explicit* *representation of the real system* in an abstractive degree (i.e., system representation). This is not necessary in deduction (e.g., analytic/mathematical models) or induction (e.g., statistical models). Because they mimick the real system's structures and behavior, simulation models are usually easier to comprehend for management or customers than many analytical/mathematical models (Shannon, 1998).

Second, simulation is an *artificial/virtual experimental approach* to study a problem. Simulation involves g*enerating an artificial history* of the real system that is represented, and *observing that artificial history* to draw inferences concerning the changes in patterns of the system characteristics (Banks, 1999). The generation of artificial histories is similar to the analytical (deduction) approach in that simulation starts with a few assumptions, but unlike deduction, in that simulation aims neither to prove any theorem (e.g., mathematic proofs) (Axelrod, 2003), nor to provide optimum answers (e.g., mathematical programming methods) or nearly optimum answers (e.g., heuristic methods) (Robinson, 2003). A simulation simply projects the temporal performance of an operation system under a specific set of supposed inputs, i.e., the logical answer of the "*what-if*" question (Robinson, 2003). The evaluation of generated artificial histories is similar to induction methods that find patterns from data, but the data here are generated through simulation, not actually measured (Axelrod, 2003). Unlike analytical and statistical models that seek an answer for the problem at once, a simulation model is often used iteratively, as an "*experimental vehicle*", in answering a certain question concerning the considered system (Page, 1994: 15; Robinson, 2003: 4). With simulation models, users repeatedly enter *alternative scenarios*, then observe and evaluate the simulated outcome until he/she has obtained sufficient understanding or identified the proper answers to the questions (Robinson, 2003; Axelrod, 2003). As a consequence, a simulation model should be seen as a *virtual lab* that helps humans to arrive at the direct answer of the research question, rather than providing a direct answer to the research question on behalf of the model (Robinson, 2003).

Third, *supporting human decisions in system operation and management* is the *intrinsic objective* of simulation (Page, 1994; Robinson, 2003). As stated in the definition, simulation cannot be separated from its purpose of obtaining a better understanding of the system performance and identifying the most efficient strategies in system operating and management. Thus, simulation intrinsically appears to be for *decision-support purposes* (Page, 1994; Robinson, 2003). One exclusive strength of simulation is that simulation is the only way to test or explore new policies, designs, catastrophic shocks, and so forth, without committing resources to their implementation (i.e., no cost), without disrupting the ongoing functionality of the real system (i.e., no damage), without the existence in any empirical dataset (i.e., control experiment conditions), while being less time consuming (i.e., control times) (Robinson, 2003). Comprehensive justification of the advantages of simulation as a scientific method can be found in Robinson (2003), Shannon (1998), and Pegden *et al*. (1995). Therefore, simulation modeling is highly suitable for quantitative *ex ante* evaluation in environmental management, including LUCC studies.

Given its merits, simulation is widely advocated as a suitable way of studying complex ecological or social systems, especially in the context of policy/management scenario research (Axelrod, 2003; Gilbert and Troitzch, 1999; Bousquet and Le Page, 2004). If simulation is to be used for modeling LUCC, then the next question is which approach should be used for representing the complex human-environment system underlying LUCC.

Multi-Agent System (MAS) for representation of the complex human-environment system: a paradigm shift from system dynamics to organizational thinking

There are two main paradigms for simulating interrelationships between the natural system and the human system, namely: *system dynamics* and *organizational* paradigms (Bousquet and Le Page, 2004; Van Dyke Parunak *et al*., 1998; Vila, 1992).

The *system dynamics paradigm*, which originated from the work of Jay Forrester in the mid 1950s (Forrester, 1995), has been proposed as an alternative to a reductionism approach in ecosystem studies for a long time (Bousquet and Le Page, 2004). The system dynamics approach describes the human-environment system as a fixed structure of *observables*, which are the measurable characteristics of interest (i.e., *state variables*) (Van Dyke Parunak *et al*., 1998). Observables are interlinked by the *flow* of matter, energy and information. Through feedback loops among the observables, the dynamics of one particular observable is controlled by the dynamics of others. Therefore, the ecosystem is metaphorically viewed as a *cybernetic system* in which observable states are subject to the global flow of control (Villa, 1992). Functionally, system dynamics models are represented by a complete set of *differential equations* that explicitly describes *beforehand* an enumeration of *fixed* causes (Gilbert and Troitzsch, 1999). Structurally, system dynamics thinking views the ecosystems as a *fixed structure* in form of a fixed causal loops diagram where the positions of observables and functional relations between them are predefined at the beginning and fixed during the simulation processes. Through parameterization of the inter-flows in both human and environment systems, modellers establish strong links among parts of the entire system (Jorgensen *et al*., 2000). Due to these fixed and strong links among system elements, the modelled system always exists in an *equilibrium state* (Bousquet and Le Page, 2004; Parker *et al*., 2002). The system dynamics approach tends to use system level variables, since it is often easier to formulate parsimonious closed-form equations using such quantities (Van Dyke Parunak, *et al*., 1998). In general, system dynamics thinking allows us to understand and explicitly control the ecosystem and social system in a way that engineers understand and control a mechanical system (Aronson, 1998), i.e., the system of type *I* as shown in Figure 1.3.

Although the system dynamics approach can help to build interlinks between human and environmental systems (Bousquet and Le Page, 2004; Parker *et al*., 2002), and to capture parts of the dynamic complexity (Vila, 1992), this approach has many limitations with respect to representation of the human-environment system underlying LUCC with its complex properties as portrayed earlier (Section 1.2.1). The first limitation is the mathematical intractability of causal relationships with the complex human-environment system, which we have analysed above (Section 1.2.1). The problem of intractability leads to the fact that most system dynamics models have difficulty in accommodating spatial linkages (Sklar and Costanza, 1991). The second limitation is that hierarchical structures and heterogeneities of humans and the natural environment, i.e., the important aspects of the complex nature of the coupled human-environment system, are not explicitly captured with system dynamics models (see Villa, 1992). The third limitation is that the system dynamics approach does not allow modeling of the changes of system organisations (Villa, 1992), i.e., adaptations.

The appearance of the *organisational paradigm* in late 1980s and early 1990s – which originated from research progress in non-linear dynamics (e.g., cellular automata), distributed artificial intelligent (DAI), complexity theory, and others –

caused a radical change in the research landscape concerning human-environment interrelationships (Bousquet and Le Page, 2004; Gilbert and Troitzsch, 1999). According to the organisational viewpoint, the human-environment system is described as a *multi-agent system* (MAS), which is *self-organised*[2] from autonomous and decision-making *entities* called *agents* (Woodridge, 2002; Bonabeau, 2002; Zambonelli *et al*., 2003). Each agent has its *specific roles*[2] that are designed as the mimicking roles of the real entities they represent (Zambonelli *et al*., 2001; Woodridge, 2002). Each agent has its designed internal structure and mechanisms for autonomously undertaking its assigned roles, thus becoming a *separate locus of control*, without any central control (Woodridge, 2000). Due to the autonomous control of agents, instant interactions in the system are no longer mathematically traceable in time and space as in the system dynamics approach (Epstein and Axtell, 1996; Axelrod, 1997). In other words, linkages among agents, between agents and their environments, are highly *loose* and *flexible*, and established by agents based on emerged situations rather than predefined as inputs as in system dynamics models (Van Dyke Parunak *et al*., 1998). Due to these extremely flexible interactions, the MAS needs not to be solved by any closed-form analytical equilibrium solutions (Parker *et al*., 2002). MAS simulation models can perform both micro and macro properties of the considered system at the same time. During simulation, agents' behavioral structure, even agents' roles, can change, making agents *adaptive* to a newly generated situation (Villa, 1992; Epstein and Axtell, 1996). Therefore, the organisational approach appears a natural alternative to overcome the limitation of the system dynamics approach in the representation and interpretation of complex adaptive systems (Bousquet and Le Page, 2004).

The difference between the system dynamics and organisational (MAS) paradigms in interpreting the complex system is summarised in Table 1.1. It implies that MAS is rather an alternative viewpoint (mindset) in the study of complex systems than just a technology (Bonabeau, 2002; Villa, 1992; van der Veen and Otter, 2001; Bousquet and Le Page, 2004).

[2] Comprehensive conceptualization of agents' roles in self-organized societies can be found in the Gaia methodology of MAS design (Woodridge *et al*., 2000; Woodridge, 2002; Zambonelli *et al*., 2003).

Table 1.1 Differences between system dynamics and organizational (MAS) mindset in studying complex systems. Source: adapted and modified from Villa (1992)

	System dynamics viewpoint	*Organisational viewpoint*
System conceptualisation	Observables (state variables)	Agent (low-level organisation)
Suitable metaphor	Cybernetic system	Parallel computer
Specification of mechanism	Centralised	Distributed
Means of analysis	Differential equations	Rules set
		Computer simulations
Key behavior	Equilibrium, dynamic complexity	Self-organising (emergent), dynamic and structural complexity
System organisation	Fixed, single level	Variable, multi-level (micro-macro)
Ecological significance [a]	Balance of nature	Nature resilient and evolving

[a] After Holling (1987), cf. Bousquet and Le Page (2004)

The advantages of MAS over other modeling approaches can be seen in at least the three following aspects. The first advantage is that MAS represents complexity and captures emergent phenomena of a considered system, as most important aspects of structural complexity (i.e., hierarchy, interdependency and heterogeneity) can be adequately represented using MAS architecture (Parker *et al*., 2002; Bonabeau, 2002). The second advantage is that MAS provides a natural description of the human-environment system because its architecture and behavior mimic the organisational structure and behavior of the real system (Bonabeau, 2002; Woodridge, 2002). The third advantage is that there is flexibility in the designation and development of MAS (Van Dyke Parunak *et al*., 1998; Bonabeau, 2002). When a MAS framework is established, it is possible to add more agents to the MAS model, or to change levels of descriptions and aggregations. With MAS, modellers can play with different sets of agents and report organisational patterns of different levels at the same time (Bonabeau, 2002; Wilenski, 1999). These promising advantages lead to the fact that the use of MAS for simulating LUCC (MAS-LUCC) has attracted the increasing attention of the LUCC research community (Parker *et al*., 2002; Bousquet and Le Page, 2004). Building and development of MAS-LUCC models has been identified as a major focus in the Implementation Strategy of the Land-Use and Land-Cover Change project, developed under the auspices of the International Geosphere-Biosphere Programme (IGBP) and the International Human Dimensions Programme on Global Environmental Change (IHDP) (Lambin *et al*., 1999).

However, MAS-LUCC is still a *young scientific field*, and many methodological issues still need to be addressed to achieve an operational MAS-LUCC. These methodological challenges can be found in Van Dyke Parunak *et al*. (1998), Parker *et al*. (2002) and Bonabeau (2002). The first challenge is to build a MAS modeling framework that reflects the organisation of the coupled human-enviroment system in an understandable manner. This relates to identification of a right level for specification, which remains an art more than a science (Bonabeau, 2002; van Noordwijk *et al*., 2001), and obviously requires interdisciplinary knowledge (Parker *et al*., 2002).

The second challenge is the specification of decision-making models for human agents and ecological models for landscape (environmental) agents. A key challenge for MAS modellers is to decide which approach to adopt among the sheer number of competing theories and techniques for designing and parameterizing these sub-models. The third challenge, probably the greatest technological difficulty, is to implement the designed models in computer platforms. This relates to computer programming and building spatial links. Although a number of MAS computer platforms have been developed, these platforms are not drag-and-drop tools for end users like other system dynamics packages (e.g., Stella) and need high volumes of programming work. The last challenge, but not least, is to calibrate, verify and validate the MAS models to make them empirically grounded for reliable operation. Although many MAS models have been created, it is fair to say that in many cases MAS is just merely a game with hypothesized humans and landscapes, because the model is not rigorously calibrated and verified against data (Bonabeau, 2002; Kanaroglou and Scott, 2002). Therefore, at present, a serious MAS-LUCC study even has to consider methodological development as a research objective.

1.4 Research objectives

Given the problems stated and the modeling approach advocated, the goal of this thesis is:

to build an operational multi-agent system simulation model of land-use and land-cover change (MAS-LUCC model) in a spatially and temporally explicit manner, which can be potentially useful for exploring alternative scenarios to improve rural livelihoods and the environment, thereby providing stakeholders with support for making better-informed decisions about land resource management.

To achieve the goal, the thesis has the four following specific objectives:

1. To build a parameterized MAS-LUCC framework for modeling the evolutions of the coupled human-environment system at a landscape level in time and space, where landscape land-use/cover and community socio-economic dynamics are self-organized from interactions among farming households (as human agents) and land patches (as landscape agents), under the influence of certain policies and other external circumstances;
2. To calibrate and verify land-use decision-making sub-models of the human agents (households) based on empirical data collected at a study site in the Central Coast of Vietnam;
3. To calibrate and verify ecological dynamics models of land patches based on pixel-based biophysical data collected at the study site; and
4. To develop an operational MAS-LUCC model through implementing (programming) such parameterized/calibrated framework on suitable computer platform(s), for initially exploring the potential outcome of selected policy alternatives in land management at the study site.

1.5 Outline of the thesis

This thesis consists of seven chapters. Chapter 1 (this chapter) analyses the main problems and alternatives in previous LUCC modeling and provides a basis for the formulation of research objectives. Through this first chapter, multi-agent system simulation (MASS) has been advocated for modeling LUCC based on the perspectives of recent paradigm shifts in environmental management strategies and ecosystem sciences, rather than on just purely technological issues.

Chapter 2 aims at clarifying technological concepts and methods of MAS and establishing a conceptual framework for detailed technical work in later chapters. It firstly provides basic concepts of MAS and agents, main architecture of agents, and reviews current MAS computer platforms that are potentially applicable for modeling LUCC. Secondly, in the light of the MAS mindset, the chapter lays out a conceptual framework of the coupled human-environment system underlying LUCC, which is the basis for the application of MAS. Third, the chapter briefly justifies the selection of the study area for empirical specifications. The chapter ends with a layout of the modeling steps that this thesis work has followed.

Chapter 3 aims at obtaining the first specific objective. It formulates the first principles and architecture of a MAS-LUCC framework, named VietNam – Land-Use DynAmics Simulator (VN-LUDAS). The chapter has two main steps. The first step is the design of a fully parameterised MAS architecture, including the design of the organisational framework for the human-environment system, and the construction of the agent structure and behavioral rules. Both households and land patches are treated as autonomous agents, which are built in by sub-models and behavioral protocol (i.e., internal programmes). The second step is the development of a simulation protocol, which co-ordinates (does not control) the working of autonomous human and landscape agents and monitors the self-organising phenomena of these interactions (i.e., LUCC and socio-economic dynamics). The architecture of VN-LUDAS and the simulation protocol are represented explicitly using textual, graphic, and algebraic languages prior to any empirical calibration and verification. The chapter is therefore expected to provide transparency with respect to the proposed MAS framework. The output of this chapter will serve as the core of VN-LUDAS, which can be implemented in certain computer platforms. When this model core has been calibrated and verified using data or knowledge of a particular area, it will be an *empirical analogue* of the core model.

Chapter 4 aims at achieving the second specific objective. It calibrates and verifies parameters and rules/sub-models built into the human agents (i.e., farming households). The chapter focuses on two main parts. The first part is the categorization of human agents (households) into typical groups according to livelihood structure and strategy using data condensation (Principle Component Analysis – PCA) and classification (K-Mean) techniques, based on household data. The second part is the deriving of the land-use decision-making sub-model for each human agent group using spatial regression analysis (M-logit regression), based on spatial biophysical and household data. The findings can be used for two purposes. Estimated parameters of human agent types and their land-use decisions will be used as inputs for the operation of the VN-LUDAS model at the study site. The empirical findings themselves also provide a better understanding of land-use adoptions, as well as of the practice and policy of land-use management in the study area.

Chapter 5 aims at obtaining the third specific objective. As in chapter 4 that deals with empirical calibrations for human agents, this chapter calibrates and verifies sub-models for the landscape agents (land patches). The chapter has four main parts. The first is the landscape characterization using GIS-based analysis (i.e., terrain analysis, physical accessibility analysis, and remote sensing analysis). The second is the empirical estimation of agricultural yields with sub-models accounting for the dynamics of cultivated land patches, using multiple log-linear regression analysis, based on plot-specific data. The third is the development and justification of a forest growth sub-model for the dynamics of forested land patches, in which the vegetation growth component is developed based on the biological system theory; the human intervention component is taken from another empirical model. The fourth is the calibration of a sub-model of natural land-cover transition, which translates accumulated small changes in land cover (i.e., annual natural vegetation growth and/or human modifications) to conversion of vegetation cover type. The cover transitional sub-model also deals with ecological edge effects in vegetation growth. As in chapter 4, the estimated parameters of landscape agents will be used as inputs for the operational VN-LUDAS at the study site.

Chapter 6 aims at satisfying the fourth specific objective. The chapter begins with a brief description of particular land-use policies in the context of the study area (and Vietnam), and the identification of relevant policy scenarios to be tested for their trade-off impacts with the VN-LUDAS model. Second, an operational VN-LUDAS model, as an empirical analogue of the VN-LUDAS core framework in chapter 3, is summarized based on empirical specifications given in chapters 4, 5 and 6. Third, the chapter deals with the simulation runs, which use parameters estimated empirically in chapters 4 and 5 as input and the datasets surveyed in 2002 as the initial state of the human-environment system, according to the defined policy scenarios. Finally, based on the simulated LUCC and other socio-economic dynamics, potential effects of land-use policy options are discussed.

Chapter 7 draws conclusions with regard to the achievements of the thesis objectives, and makes recommendations about potential applications and further development of the first VN-LUDAS model.

2 MULTI-AGENT SYSTEM CONCEPTS, METHODS AND A PROPOSED CONCEPTUAL MAS-LUCC MODEL

2.1 Introduction

Autonomous agents and multi-agent systems (MASs) are a promising paradigm for modeling LUCC processes as emergent phenomena of the complex interactions between humans and the environment (Bousquet and Le Page, 2004; Parker *et al*., 2002; van der Veen and Otter, 2001). However, the building of a MAS model relates to many new concepts and design methods that do not exist in conventional modeling approaches. In the last few years, there have been landmark papers reviewing MAS applied to LUCC (e.g., Parker *et al*., 2002; Parker *at al*., 2001; Bousquet and Le Page, 2004) and a number of introductory text books about MAS (e.g., Woodridge, 2002; Ferber, 1999). However, most of the review papers about MAS for modeling LUCC focus on the analysis of the potentials of MAS for modeling complex systems and have not given enough clarified the underlying concepts and methods. MAS text books, even at an introductory level, are still largely computer-science based. Thus, clarification of MAS concepts and design methods, especially in connection with geo-simulation of LUCC, will be useful.

Because MAS is an organisational modeling paradigm, which differs from a conventional system dynamics (i.e., Forrester style) approach (see Chapter 1), conceptualisation of the system subjected to MAS modeling is very different from those of system dynamics modeling. Also, building a MAS model requires a computer platform for implementation. Since a variety of MAS simulation platforms has emerged in the last few years, some reflection about what kind of computer platforms are potentially suitable for MAS-LUCC is called for.

This chapter will clarify the technological concepts and methods of MAS, and establish a conceptual framework as the basis for the detailed model specifications in later chapters. It firstly provides basic technological concepts and methods of MAS, which are potentially applicable for modeling LUCC. Secondly, the chapter briefly reviews existing MAS computer platforms to provide a basis for selection of a suitable one for our work. Thirdly, the chapter lays out a conceptual framework of the coupled human-environment system underlying LUCC that is a prerequisite for the application of MAS. The final part is a layout of the modeling process that this thesis work has followed.

2.2 Basic concepts of Multi-Agent System (MAS)

2.2.1 Definition and interpretation of Multi-Agent System

There are various definitions of the Multi-Agent System (MAS). A concise definition of MAS is given by Benenson and Torrens (2004: 154) as:

"*A multi-agent system is a community of agents, situated in an environment*".

"*Agent*" refers to autonomous decision-making entities within the system (Bonabeau, 2002: 1). "*Environment*" is the space that houses agents and supports their activities (Benenson and Torrens, 2004: 154). "*Community*" refers to an "organized society" of agents in which each agent plays specific *roles* and *interacts* with other agents according to protocols determined by the roles of the involved agents (Zambonelli *et al*., 2003: 3). These concepts are defined and interpreted in detail as follows.

2.2.2 The concept of agent

Definition of agent

Benenson and Torrens (2004: 154) give a comprehensive definition of agent, which is as follows:

"*An autonomous agent (1) is a system situated within and a part of an environment; (2) that senses that environment and acts on it, over time; (3) in pursuit of its own agenda, and (4) as to effect what it senses in the future*".

An agent that shares the four characteristics as in the definition above is called a *"minimal" agent* (Benenson and Torrens, 2004). This "minimal" agent concept can be applied for representing a wide range of entities in the real world - from physical particles (e.g., atoms), biological systems (e.g., animals and plants), and human organisations (e.g., households, villages and nations) (Gilbert and Troitzsch, 1999; ALC, 2003). Here, we refer to *human agents* as agents representing human entities (e.g., persons, households, village, etc.) and *biophysical agents* as agents representing biophysical entities (e.g., plants, animals, land patches, etc.).

Characteristics of agents

The interpretation of agency has generated a wealth of literature and many characteristics of agents have been discussed (Benenson and Torens, 2004). Detailed characteristics of the agents depend on particular applications and the intentions of modellers when specifying their MAS models (Franklin and Grasesser, 1996). However, we can identify common characteristics of autonomous agents (representing biophysical or human entities) as follows.

Agents are *autonomous*. Autonomy is the most important property of an agent (Woodridge, 2002; Zambonelli *et al*., 2003). An agent has its own internal thread of execution, typically oriented to the achievement of a specific task (on behalf of a user), and it decides for itself what actions it should perform at what time (and location). Thus, an agent is a *locus of control* within the system, not passively subject to a global, external flow of control in its actions. Autonomy is the first characteristic that makes agents different from standard objects in object-based systems (Woodridge, 2002) and automata in cellular automata systems (Benenson and Torrens, 2004).

Agents are *reactive*. This means agents are able to perceive the environment (including other agents) and respond in a timely fashion to changes that occur in it, in order to satisfy their design objectives (Woodridge, 2002; Franklin and Grasesser, 1996). The reactive behavior of agents makes interactions within MAS highly flexible and unpredictable; it also enables agents to adapt to changes in the environment (Woodridge, 2002).

Agents are *proactive*. A proactive agent means the agent is able to exhibit *goal-directed* behavior by taking the initiative in order to satisfy its design objectives (Woodridge, 2002). The "goal" here, first and foremost, is what modellers want agents to achieve or solve themselves (Woodridge, 2002), not necessarily human goals. Thus, pro-activeness can be attributed to both human and biophysical agents[3]. Pro-activeness of biophysical agents can be simply that their behaviors are generated through internally calculative routines, which comparatively identify a solution satisfying the assigned objectives (e.g., Box, 2002; Reynolds, 1987). Proactive behavior of human agents often deals with their evaluation of options to achieve human goals, such as maximal

[3] This argument is in contrast to some thoughts that pro-activeness is only attributed to human-like agents (e.g., Beneson and Torrens, 2004)

utilities/profits and minimal risks (Bousquet and Le Page, 2004). In general, goal-directed behavior of agents can be achieved by making use of complex-procedure-like patterns of action (Woodridge, 2002).

Agents must have a *capability of interacting* with their environment, including other agents. Agents are inherently situated in an environment and cannot achieve assigned tasks without taking their environment into account. As some goals can only be achieved with the cooperation of others, agents are sometimes designed to be able to co-operate with others.

Agents are *perceptive*. With biophysical agent, the "perceptiveness" concept can be interpreted as an extension of the neighborhood concept in automata: biophysical agents themselves can sense information in their surroundings. The perceptiveness of human agents are more sophisticated, commonly endowed with a *cognitive model of their "world"* and the ability to recognize emerging spatial assembles and structures within their world. An agent's recognized world is normally defined to a certain extent (sometime called agent's *vision*), not necessarily for the whole system (Woodridge, 2002; Ferber, 1999; Willensiky, 1999). The reflection of actual information about an agent's world into his recognition model is called *knowledge representation*. Some of the environmental information agents have may be incorrect, thus the represented knowledge may be imperfect to some degree. Such possible erroneous knowledge is called agent's *belief*, distinguished from true (scientific) knowledge (Gilbert and Troitzsch, 1999).

Adaptation is an important characteristic of agents (potentially both human and biophysical agents). Adaptation of agents in MAS means the ability of agents to change their behavior rules/functions, based on experience absorbed during their model life span (Benenson and Torrens, 2004). Adaptations can occur within different scopes. At the first order of adaptation, parameters of behavioral rules/functions change, while the frame/structure of the rules/functions does not changes, e.g., when a human agent quantitatively changes parameters in its objective function, but the function's structure still remains. At a higher order of adaptation, both parameters and structures of rules/functions change, e.g., when a human agent changes its objective functions (see Lenton and van Oijen, 2002).

Agents are *heterogeneous*. Because each agent is designed firstly to represent a *specific individual* situated in a population/society, agents often differ from one another in terms of state, history, and behavior.

2.2.3 Environment in multi-agent system

In any MAS, agents are situated in an environment. Agents play in this environment, searching for information and are often able to modify it. Environment is a source of richness, but also a source of complexity, since the access it provides to the resources depends on the structure of the objects it contains (Ginot *et al.*, 2002). What constitutes an environment depends on what is being modeled. The environment of an agent is often considered its *spatial context* (Gilbert and Troitzsch, 1999; Ferber, 1999). At a given moment, an agent (representing human or biophysical entities) will have a position in the system and be associated in a simulated space.

Environment properties

Russell and Norvig (1995) suggest that the following environment properties should be considered in MAS designation:

- *Accessible* versus *Inaccessible*. An environment is accessible to an agent if the agent can get perfect information about the environment's state, and is inaccessible otherwise. The real environment is often inaccessible to some degrees.
- *Deterministic* versus *Non-deterministic*. An environment is deterministic if the outcome of any action performed is uniquely defined, and non-deterministic if otherwise. The real environment may be non-deterministic, i.e., uncertain, with some attributes to some degree.
- *Static* versus *Dynamic*. A static environment is one that is assumed to remain unchanged except by the performance of actions by the agent. In contrast, a dynamic environment is one that has other processes operating in it, and hence changes in ways beyond the agent's control. The biophysical world is a highly dynamic environment. When applying MAS for geo-simulation, representing dynamic environmental properties is an increasing awareness, but also a challenge (S. J. Park, personal communication).

- *Discrete* versus *Continuous*. A discrete environment is one which can be garanteed to only be in a finite number of discrete states, whereas a continuous environment may be in an uncountable number of states. Although discrete environments inevitably have some mismatches with the continuous environments as in the real world, using discrete representation of the environments has many advantages, mainly i) being simpler to design and operate with computers (Woodridge, 2002), and ii) being able to apply the agency concept to the environment (Box, 2002). The idea of treating the environment as agents is discussed in later sections.

In a MAS, efforts to represent the inaccessibility and non-determinism of the environment lead to the fact that each agent has its limited "*sphere of influence*" (i.e., "*vision*" as referred to in above section), within which the agent has at best partial control over its environment (Woodridge, 2002; Zambonelli *et al*., 2003) (see Figure 2.1).

Agent-based environment and "all agent" system

Conventionally, in a MAS, all elements that are not agents go under the heading *non-agent environment*. Access to a non-agent environment is difficult for primitives[4], since it multiplies the number of objects and data structures they have to handle (Ginot *et al*., 2002). To simplify the primitives' conception and use, the environment should be described with agents, i.e., agent-based environments, so that primitives treat and exchange only agents, and data structures that contain agent states. A MAS with an agent-based environment is called an "*all agent*" system (Ginot *et al*., 2002). Some newly developed MAS packages, such as NetLogo (Wilensky, 1999), work with "*all agents*" systems, i.e., the system consists of agents only. The choice of "*all agent*" systems forces us to better allocate active roles to the environment in the system. This greatly simplifies their conception and enhances their flexibility.

In MAS-LUCC, applying the viewpoint of "*all agent*" causes all landscape units to be treated as autonomous agents. This is a relatively new approach for modeling the dynamic environment within MAS-LUCC models (Box, 2002). The scientific basis

[4] A primitive is a low-level operation from which higher-level operations can be constructed. In programming, primitives are the basic operations supported by the programming language. A programmer combines these primitives to create more complex operations, which are packaged as functions, procedures, and methods (http://www.webopedia.com/term/p/primitive.html).

for this designation is that landscape units themselves host natural processes that change their nature, as noted by Box (2002): "*any given portion of the Earth's surface is a unique location, which reacts to localized processes with no notion of the global behavior of the landscape*". This notion implies that landscape units are reactive, heterogeneous and autonomous (as loci of control) with local ecological processes operating in them in ways beyond the human agent's control. In reality, for a given portion of land, the processes of hydrology, pedology, and vegetation growth are largely autonomous and responsive to localized conditions on that portion of land and its immediate neighbourhood (Box, 2002). It is obvious that ecological mechanisms of an annual cropping parcel differ from those of a forest stand or of a residential site. It is also safe to say that local ecological processes constantly occur by themselves, without any intervention from human agents. Therefore, a landscape unit potentially meets the characteristics of a "minimal" agent as defined above (see Section 2.2.2).

The application of the agent concept to represent the dynamic environment is found only in the most recent literature on MAS-LUCC. Few LUCC modelers, among them are Box (2002) and Huigen (2003) have recently treated landscape units as agents when designing the environment within MAS-LUCC. This study adds to the current meagre body of literature.

2.2.4 Interactions

Interactions are the key characteristic of any MAS (Woodridge, 2002; Ferber, 1999). Interactions in MAS occur in many ways. It is useful to categorize these interactions according to their *means*. There can be three main types of interaction: i) *biophysical interactions*, ii) *interactions mediated by the environment*, and iii) *interactions by communication* among agents. The definition and brief interpretations of these interactions are as follows.

Biophysical interactions

Biophysical interactions exert *actions* on the environment or other agents. It is useful to distinguish two sub-types of biophysical interactions: interactions between agents and their environment, and interactions between environmental agents.

In biophysical interactions, human agents act and change their environments, and the environments constrain agent activities. This implies that agents exist through interactions with their environment and their interactions are, therefore, inevitable and occurring constantly (Zambonelli *et al*., 2003) (see Figure 2.1). For example, human agents change land covers (i.e., biophysical properties of the Earth's surface), and they benefit from the land but also suffer from environmental externalities (i.e., unexpected environmental feedbacks) due to their production activities. Therefore, biophysical agent-environment interactions play a decisive role in the evolution of the coupled human-environment system.

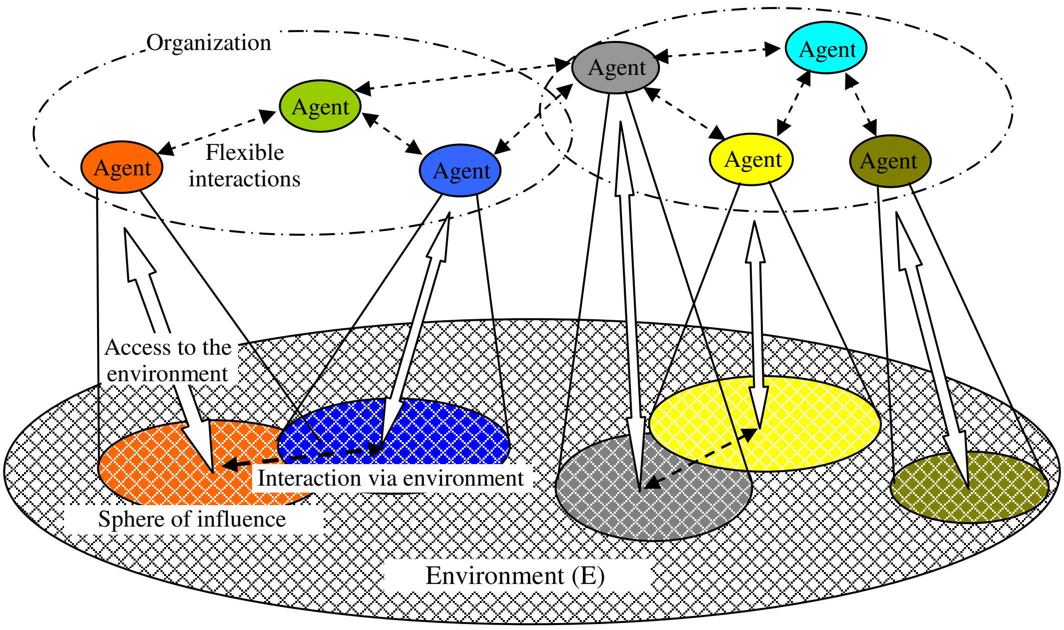


Figure 2.1 Multi-agent system and interactions. Sources: adapted and modified from Woodridge (2002), Woodridge *et al.* (2000), and Zambonelli *et al.* (2003)

Biophysical interactions among environmental agents are the basis of several ecological spatial interdependencies, such as the landscape processes (topographical sequences) of soil and hydrological phenomena, ecological edge effects, vegetation gap dynamics, habitat connectivity, food-webs, and so forth (Parker *et al*., 2002). The importance for consideration of this type of interactions in MAS designs depends on concrete applications, modeling purpose, and knowledge/data capability. Many MAS models for urban dynamics do not take spatial biophysical interdependencies into account (e.g., Benenson and Torrens, 2004; Ligtenberg *et al*., 2001; Loibl and Toetzer, 2003), perhaps because biophysical conditions of the land are not critically important for individual choices of urban activities. In the case of modeling of land-use changes in rural areas, especially in heterogeneous environments such as watershed ecosystems, biophysical interactions among landscape units become important for individual decision-making on land use, because the biophysical conditions of the land (e.g., soil, water, etc.) are vital for agricultural and forestry productions. However, considering these biophysical interactions in the MAS-LUCC model is still at a modest stage due to either methodological challenges (e.g., agent-based design and model coupling) or lack of knowledge about interaction rules (S. J. Park, personal communication).

Interactions through communication

Interactions through *communication* (or communicative interactions) are the direct exchange of messages among agents or between agents and their environments (Figure 2.1). Communicative interactions between agents and their environment exist thus that agents always infer knowledge about the spatial organization of their "*sphere of influence*" (i.e., agent's vision) to arrive at their decisions and actions (Ligtenberg *et al*., 2001; Woodrige, 2002). This type of interaction is inherent in any MAS design, as agents must have a particular accessibility to the environmental information.

Communicative interactions among agents often concerns the exchanges of information about contracts, goods, or services among human agents (Bousquet and Le Page, 2004), which reflect social relations among human agents such as purely information exchange, negotiations or co-operations. However, the inclusion of this type of interaction in the MAS to a suitably abstractive degree is flexible, depending on concrete applications and modeling objectives. Moreover, during simulation processes, the existence of communicative interactions of this type is also flexible, depending on specific decisions of each agent rather than being indispensable, as is the case in the agent-environment interactions.

Interactions mediated through environment

Interactions *mediated through environment* mean that the results of an agent's action transforms the common environment, thus causing retroactive effects on other agents

(Figure 2.1). Because the agent's activities always cause environmental changes and the environment is always the resource for the agent's activities, interactions among agents mediated through the environment are *inevitable* and *constantly occur* in the system. Due to these interactions, the dynamics of the environment and its heterogeneity can be seen as a medium for collective adaptation (co-adaptation) between agent's organizations and the environment structure (Bousquet and Le Page, 2004). The notion "*the structure of the environment and the organization of the group of agents are mutually co-determining*" (Theraulaz, 1994, cf. Bousquet and Le Page, 2004) is one of the key concepts used for "*SWARM intelligence*", which have been used for many MAS models (Bonabeau *et al*., 1999; Kenedy *et al*., 2001). This type of interaction is also close to the concept of externality in economics (Bousquet and Le Page, 2004).

The complexity of complex adaptive systems, e.g., the coupled human-environment system, does not necessarily rely on the numerous complicated rules of interaction (Axelrod, 2003; Camazine *et al*., 2001). The system complexity is shown in the diverse nature of the system's global response, which often sensitively depends on the *initial state of the system* and *nonlinear interactions* among system components (Axelrod, 2003; Bonabeau, 2002; Sanchez and Lucas, 2002; Camazine *et al*., 2001). Since these nonlinear interactions involve amplification or cooperativity, global complex behaviors may emerge from interactions among even similar components governed by simple rules (Camazine *et al*., 2001; Bonabeau, 2002). Therefore, this specification often uses nonlinear forms of behavioral rules, random variables/parameters and functions. Later, in the next chapters, parameter calibrations and database development for the model will be done through case studies at a site in Vietnam, to illustrate how to adapt the model to a concrete site and use it with real data.

2.3 Agent architecture

An agent's architecture is essentially a map of the internals of an agent showing data structures, the operation that may be performed on these data structures, and the control flow between these data structures (Woodridge, 2002). Agent architecture is one of the main research focuses in MAS science, since it determines decision-making and interaction protocols of agents (Bousquet and Le Page, 2004). Several types of agent architecture have been proposed. Brief reviews of agent architectures can be found in

Russell and Norvig (1995), Woodridge (2002), Bousquet and Le Page (2004), Benenson and Torrens (2004). Here, instead of repeating such reviews, we select the two most common approaches to agent architectures in geo-simulation for more detailed clarification. The first architecture type is *production rules system*, which is very often used for representing *reactive (replex) behavior* of agents (Russell and Norvig, 1995; Gilbert and Troitzsch, 1999; Bousquet and Le Page, 2004). The second popular architecture type is parameterized functions, which is widely used for representing rational or bounded rational behaviors (Russell and Norvig, 1995; Benenson and Torrens, 2004). The main sub-types of parameterized function architecture are also discussed.

2.3.1 Production rules system and reflex decision-making mechanism

Most (reactive) agents are built using some kind of rule system, of which the simplest is a *production rules system* (Gilbert and Troitzsch, 1999; Bousquet and Le Page, 2004). A typical production rules system has three components: i) a working memory (agent state), ii) a set of interaction rules, and iii) a rule interpreter (Gilbert and Troitzsch, 1999). These components are described in the following.

Working memory (agent state)

This is a set of internal state variables (i.e., slots), which stores data/information about the agent (internal information) and the state of the environment (external information) (Ligtenberg *et al.*, 2001; Woodridge, 2002; Gilbert and Troitzsch, 1999). The stored external information is linked to the spatial structure (organisation) within the agent's "*sphere of influence*" (i.e., agent's vision), enabling agents to recognise up-to-date its emerging environment. Depending on the detail of the application, state variables can record historical sequential events of agent-own data or spatial environmental data.

Set of rules

Almost all agent architectures use rule systems. Each behavioral rule consists of two part: the *condition* part (i.e., detector) specifying when the rule is to fire, and the *action* part (i.e., effector) stating what is to happen when the rule fires. This rule is called condition-action rule, or stimulus-response rule, if-then rule, or production rule (see

Russell and Norvig, 1995). Such condition-action rules are often used in MAS systems (Bonabeau, 2002; ALC, 2003). A collection of condition-action rules with appropriate organising can describe how agents interact and behave, as for instance noted by ALC (2003):

- Some rules act on the detector-originated message, processing information from the environment (perceiving personal and environmental information),
- Some rules send messages that activate other rules (mechanism of messages passing through rule-control-rule), and
- Some rules send messages that act on the environment (e.g., final land-use decision rule)

Using the rule set as a basis for designation of autonomous agents has many advantages. First, the rules themselves are expressions of reactive behavior of agents to the situation they find themselves in (Gilbert and Troitzsch, 1999; Bousquet and Le Page, 2004). Which rules fire and when they do so depends on the contents of the working memory and thus on the past experiences of the agent and on the state of the environment as the agent perceives it; thus, modellers do not have to design beforehand exactly what an agent has to do. Second, the condition part of certain rules can be extended to a more intensive model of performance measure, such as optimisation procedures or an ecological model, to enable agents to behave more proactively (Polhill *et al*., 2001; Becu *et al*., 2004; Bousquet and Le Page, 2004).

Rule interpreter

Given a specific dataset and a rule set held by an agent, a *rule interpreter* (cf. Gilbert and Troitzsch, 1999) is a logical program specifying how agents make *decisions* in response to a perceived situation at each moment in time and each location in space to achieve its designed objectives. The rule interpreter considers each rule in turn, and decides *if the conditions of the rule are met*, and if necessary carries out the action (Gilbert and Troitzsch, 1999). This is simply a look-up (i.e., matching) procedure for identifying a rule whose condition matches the current state stored in internal state variables, then performing the action associated with that rule. Because this decision-making mechanism totally relies on condition-action rules and is without any performance measure and alternative evaluation, it is called *reflex decision-making mechanism* (Russell and Norvig, 1995).

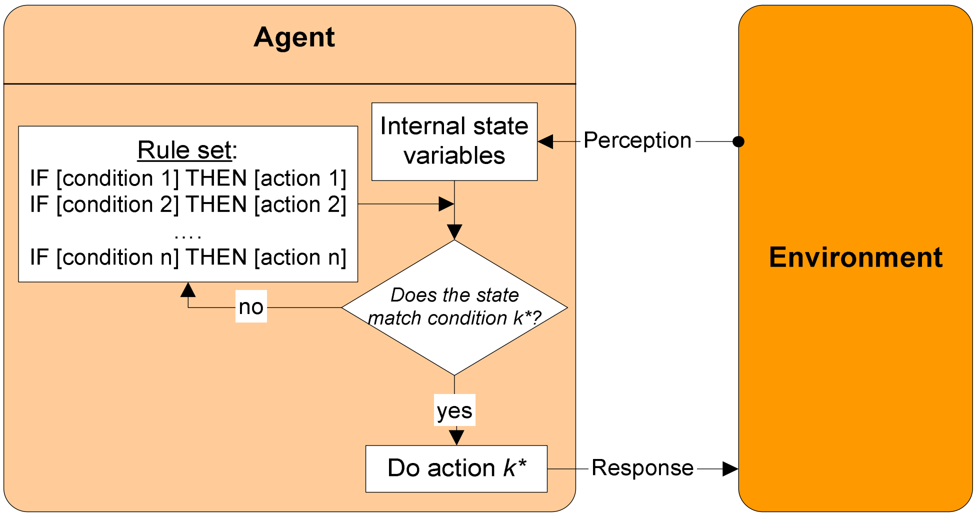


Figure 2.2 Production rules system with reflex decision-making mechanism. Sources: synthesized from Russell and Norvig (1995) and Gilbert and Troitzsch (1999)

Figure 2.2 illustrates the rule interpreter cycle using the reflex decision mechanism. When a rule has been fired by carrying out its action part, the rule interpreter cycles round and looks again at all the rules to find which to fire next. The action that the agent carried out might have changed the contents of its state variables, so the rules that fire on the second cycle may not be the same as the ones that fired first time round. It is usual for rules to specify either the action that directly affects the agent's memory or that affects the environment in a way that the agent can perceive (Gilbert and Troitzsch, 1999). If a rule does not have either of these consequences, it will be fired on every cycle until the effect of some other rule makes its condition part no longer true.

The reflex decision-making mechanism is suitable to represent *reactive behavior* of both human and biophysical agents. With human agents, the application of reflex decision-making assumes that people do not (or cannot) calculate any anticipated values of alternatives, but rather react in a timely fashion according to their daily routines to select directly options based on current conditions (Cioffi-Revilla and Gotts, 2003; Haggith, 2002). As it works totally based on rules, the reflex decision-making mechanism is also highly suitable for modeling state transitions of biophysical agents. An example for this application is the transition of land cover of land patches as illustrated in Chapters 3 and 5.

2.3.2 Parameterized functions and goal-directed decision-making

Knowing about the current state of the environment is not always enough to decide what to do. In reality, besides a current state description, the agent needs some sort of *goal* information, which describes situations that are desirable. This requires some kind of *performance measure* of states of the agent and associated environment with respect to the achievement of the agent's goals, to provide an evaluation basis for making decisions about actions (Russell and Norvig, 1995; Woodridge, 2002; Eastman, 2001). With this approach, the decision process is understood as the procedure that allows agents to select decision criteria and combine them to arrive at a particular evaluation, and by which evaluations are compared and acted upon. Notice that *rules are still needed* for agents to make use of these performance measures when making decisions about actions (Eastman, 2001). *Decision theory*, pioneered by John von Neumann and Oskar Morgenstern (1944), cf. Russell and Norvig (1995), combines *probability theory* with *utility theory* to give the first general theory that can evaluate good actions from bad ones from the viewpoint of each human agent.

Utility function, choice probability and rational decision-making

The utility theory provides a formal and complete framework for measuring the preferences of an human agent associated with the performance (Russell and Norvig, 1995; Malzcewski, 1999). *Utility* is a function that maps states of the agent and its environment into a real number, which describes the associated degree of the agent's "*happiness*" (or "*preference*") regarding a possible action (i.e., alternative) with respect to the agent's goals (Woodridge, 2002; Russell and Norvig, 1995; Benenon and Torrens, 2004). The utility function of a human agent is often defined by both *deterministic* and *probabilistic (stochastic)* components:

$$U_i = U(A_i \mid Ag, Env) = f(V_i, \varepsilon_i) \tag{2.1}$$

where $U(A_i|Ag, Env)$ (or neatly U_i) is the utility function of alternative A_i anticipated by agent Ag situated in environment Env. V_i is the deterministic component of the utility function (often called *value function*) representing known/measurable values of the state of agent Ag and environment Env. ε_i is the stochastic component, referring to the uncertain nature of the perceived state (Malzcewski, 1999; Briassoulis, 2000).

According to multi-attribute utility theory, the value function (V_i) perceived by an agent involves the identification of two elements: i) a vector/list of the agent's state variables (decision variables); and ii) a vector/list of relative importance of the decision variables (*preference coefficients*) representing personal preference to the decision variables (Malzcewski, 1999). By multiplying the vector of decision variables with the vector of the preference coefficients, trade-offs among decision variables are taken into account in the multi-attribute value function (V_i) (Malzcewski, 1999; Eastman, 2001):

$$V_i = V_i(A_i|Ag, Env) = [X_{ij}] \times [\beta_{ij}] = \sum_j \beta_{ij} X_{ij} \tag{2.2}$$

where $V_i = V(A_i|Ag, Env)$ is the value function expected for option A_i by agent Ag situated in environment Env, $[X_{ij}]$ is the vector (list) of decision variables, and $[\beta_{ij}]$ is the vector (list) of preference coefficients.

Due to the existence of the stochastic component (ε_i), the utility function is normally standardised in a *choice probability* (Dale *et al*., 1993; Beneson and Torrens, 2004). Given a complete specification of the utility function and that human agents are rational, the values of utility can be transformed to choice probabilities. A *rational behavior* assumes that human agents have *perfect knowledge* in recognizing all available options and choose the best among them, objectively (Benenson and Torrens, 2004). Thus, given a set of possible actions and their associated utilities, the rational human agents necessarily tries to choose an action that provide the *maximal utility*, i.e., expressing *optimizing behavior* (Kitamura *et al*., 1997; Wu, 1998; Woodridge, 2002; Beneson and Torrens, 2004). The probabilistic expression of utility-maximising behavior is expressed as:

$$P(A_k|Ag, Env) = Prob(U_k \geq \max\{U_i | i=1,2,\ldots,N\}) \tag{2.3}$$

where $P(A_k|Ag, Env)$ (or neatly P_k) is the probability that agent Ag situated in environment Env decides to choose alternative A_k. U_k is the utility function expected for alternative A_k, and $\{U_i|i=1,2,\ldots,N\}$ is the set of expected utilities of all possible alternatives.

Various models can be developed for calculating the choice probability $P(A_k|Ag, Env)$ by assuming different distribution functions for the stochastic component (ε_i) of the utility function (Kitamura *et al.*, 1997). When ε_i adopts an extreme probability distribution (e.g., Gumbel distribution) and all alternatives A_i are assumed to be independent from each other, it is possible to calculate the probability of agent's choice among alternatives using *multi-nominal logistic* form (McFadden, 1973, cf. Nelson *et al*. 2004):

$$P_k = P(A_k|Ag, Env) = \frac{e^{V(A_k|Ag,Env)}}{\sum_i e^{V(A_i|Ag,Env)}} \qquad (2.4)$$

where e = 2.718 ... is the base of natural logarithm, k indexes the alternative considered, and i indexes all possible alternatives (i=1,2,…,N).

The use of choice probabilities P_k instead of raw utily functions has many advantages. First, P_k values are normalised transformations of utility functions (i.e., $P_k \in [0,1]$ and $\sum P_k = 1$), thus are easier to be compared and interpreted. Second, choice probability is a convenient tool for modeling bounded-rational behavior of human agents (which will be discussed in the next section) (Benenson and Torrens, 2004).

Assuming that the agent is purely rational (i.e., having purely optimal behavior), the agent will choose *exactly* action A_{k*} that provides the maximal expected utility (see Figure 2.3):

$$A_{k*} = arg\ max\{P_i\ ; i= 1, 2, \ldots, \text{n}\} \qquad (2.5)$$

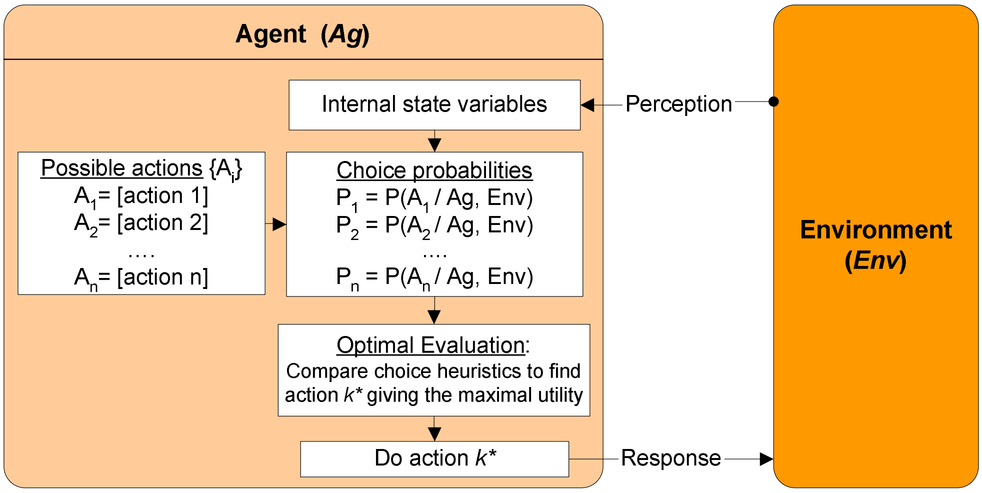


Figure 2.3 Utility-based architecture showing rational (i.e., utility-maximizing) decision-making mechanism of agent. Sources: synthesized from Russell and Norvig (1995), Woodridge (2002), Benenson and Torrens (2004)

Rational, in particular, optimizing behavior is widespread in economics-oriented modeling of human decision-making (Weibush *et al.*, 1997; Balman, 1997; Berger, 2001). Given a system of rational human agents built on parameterised utility functions and choice heuristic, we can do analytic descriptions of the outcome of an agent's selected action and then use calculus to explain trade-off dynamics in that system (Benenson and Torrens, 2004; Chapters 4 and 5 of this thesis). Moreover, it is convenient to calibrate these parameterised functions using statistical methods with empirical data, thus making the agent's decision models more empirically grounded (see Chapters 4 and 5 of this thesis).

Formulisation of bounded rational decision-making

In reality, humans can behave in a variety of ways, not just utility-maximising. The bounded-rationality approach assumes that humans make choice decisions on the basis of information that is partial in any possible respect (Benenson and Torrens, 2004). Agents are *bounded in knowledge* and differ in reaction in respect to the set of available options, characteristics of each option, the agent's ability to compare these options, and so forth (Simon, 1982 cf. Benenson and Torrens, 2004). Socio-psychological research is definitely in favour of the paradigm of bounded decision making, and different theories and concepts of choice behavior have led to implementation in different ways. However, MAS for geo-simulation tends to avoid these complicated notions, and parameterised utility functions and choice probabilities still remain a convenient basis to be extended for modeling bounded rationality.

The use of parameterised utility functions and choice probabilities for formulising bounded rationality differs from the rational case in terms of, at least, the two following aspects. First, the *order* in which options are taken for consideration may be important. Regarding this point, Benenson and Torrens (2004) review the three common routines by which agents pick an option in the set of available options. The *random choice* routine is that human agents just pick randomly, with equal probability, an option in the option set for further considerations (see Figure 2.4a). The *satisfier choice* routine is that human agents pick an option for further examinations based on evaluation of a "*satisfier threshold*" Th_{Ag} of utility (see Figure 2.4b). As far as we know, this design is not popular, because it is difficult to set a "*satisfier threshold*" of people. The *ordered choice* routine is that human agents consider firstly the top ranked option in the list of descending ordered options (see Figure 2.4c).

Second, given an option preliminarily picked, the final decision to choose this option is still the result of testing and rejecting with the choice probability (see Figures 2.4a, b, and c). If the ordered choice routine identifies preliminarily the action A_{k*} with the maximal choice probability $P(A_{k*}|Ag, Env)$, there must be still some chance that the agent *Ag* does not select this optimizing option. This is natural. For instance, individual households behave differently each time in choice situations. Farmers rarely have complete information at their disposal with which to select a site for a crop or for collecting forest products. A land plot with the lowest utility score may still have some chance of being selected in the early colonization process (Dale *et al*., 1993).

The probabilistic trying of A_{k*} with the likelihood $P(A_{k*}|Ag, Env)$ can be practically performed using *random-proportional* rule (Dorigo *et al*., 1999). Let *q* be a random variable uniformly distributed over [0,1], then the *random-proportional* rule - used by the agent *Ag* situated in environment *Env* to actually choose the action A_{k*} with the choice probability $P(A_{k*}|Ag, Env)$ - is the following:

$$Acceptance\text{-}of\text{-}A_{k*} = \begin{cases} true & if\ q \le P(A_k|Ag, Env) \\ false & otherwise \end{cases} \qquad (2.6)$$

a) Random choice agorithm

b) Satisfier choice algorithm

c) Ordered choice algorithm

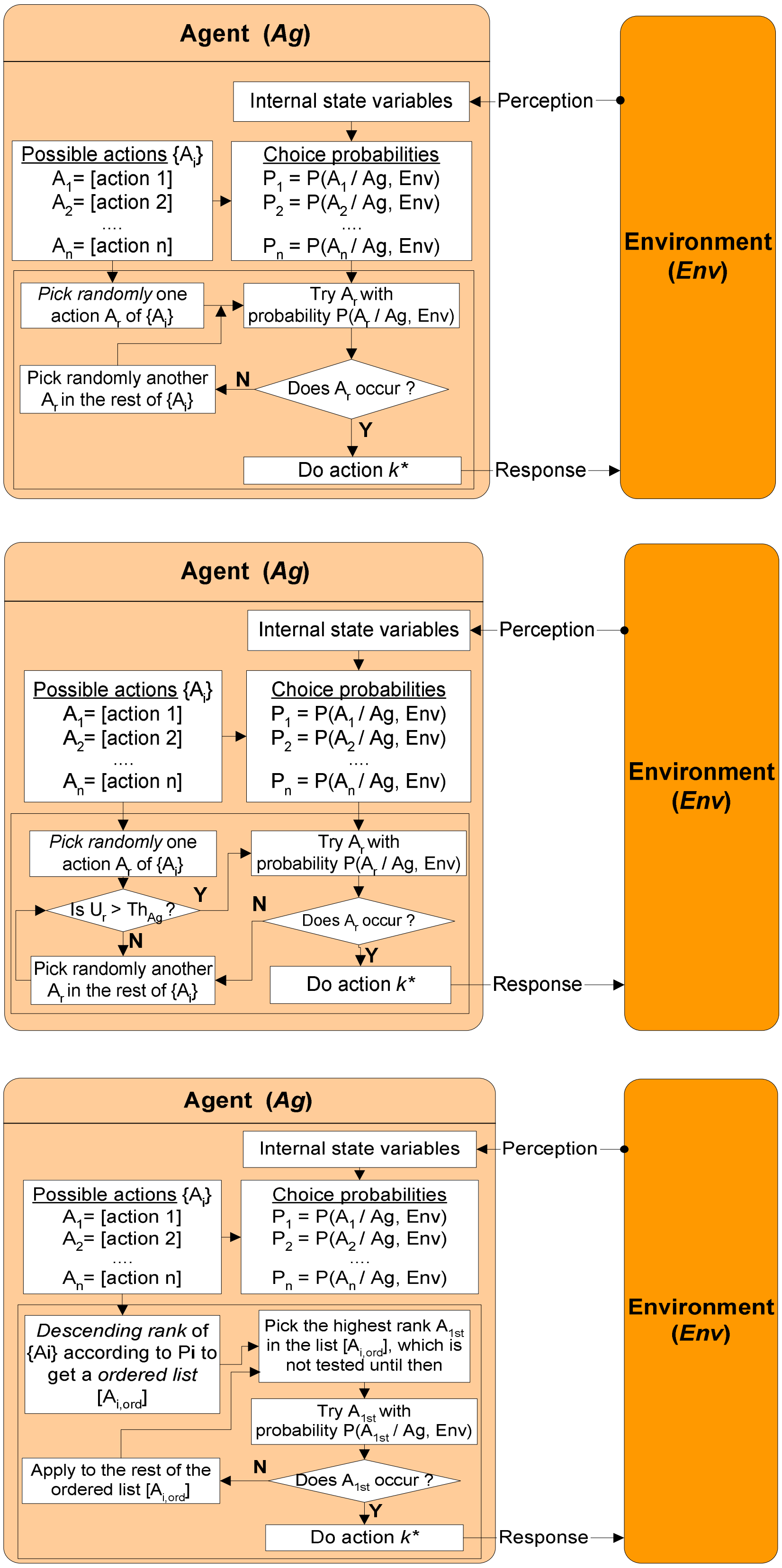


Figure 2.4 Three common formulisations of bounded rationality in MAS for geo-simulation. Source: modified from Benenson and Torrens (2004)

where *Acceptance-of-*A_{k*} = *true* means A_{k*} is finally chosen, and *Acceptance-of-*A_{k*} = *false* means A_{k*} is rejected. The Equation 2.6 means, the likelihood that the action A_{k*} is not chosen is (1 - $P(A_{k*}|Ag,Env)$).

Also, the calculation for reaching an optimum solution may be conditioned by production rules (Polhill *et al.*, 2001; Chapter 3 of this thesis). Some production rules can be used as filters in picking an option from the option set preliminarily, and/or other rules may be used as conditions for testing or rejecting the considered option later.

Parameterised ecological functions for landscape agents

The idea is that the whole landscape environment still be considered as a regular grid of congruent cells, which are rather autonomously *landscape agents* than automata. Following this approach, each landscape cell/patch is built as a programmed unit; including i) state variables (corresponding to GIS-raster data layers), ii) *internal parameterized ecological sub-model*(s), and iii) *behavior protocol* (i.e., internal program of the land patch) enabling each patch to behave autonomously and responsively to local spatial conditions (at its location and possibly neighborhood) and interventions of human agents (Box, 2002).

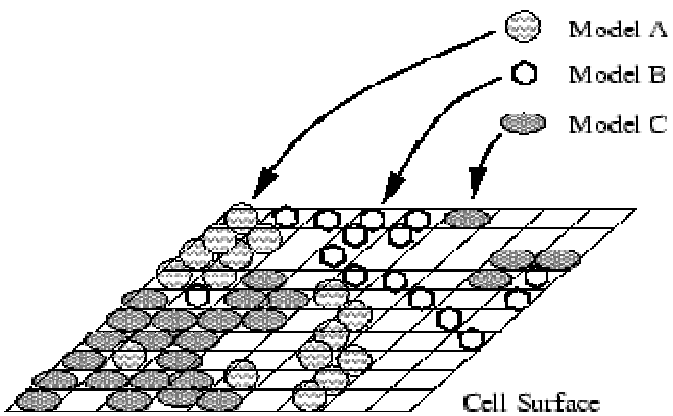


Figure 2.5 Landscape units as intelligent agents: landscape units (patches/cells) of particular typologies (i.e., patch type A, or B, or C) are built-in with specific ecological models (i.e., model A, B, and C, respectively), thus are able to autonomously decide and act. Source: Box (2002)

A key challenge for modelers designing an agent-based environment is to decide among the sheer number of competing biophysical models to harmonize with data availability. Landscape agents are expected to behave in obeying physical forces and/or ecological principles, thus they need to be specified with rigorous biophysical models proved by either empirical data or biophysical laws. However, pixel-based biophysical data are not often available or are expensive to gather, and biophysical interaction rules are not yet pair of our scientific knowledge. Thus, which biophysical models should be included in which approach and extent very much depends on the modeling objective, and are rather art than purely science (Van Noordwijk, 2001; Bonabeau, 2002). Internal biophysical models can vary from a set of empirically ecological functions (e.g., empirical agricultural yield functions as in Chapter 5) or more system dynamics models (e.g., the forest dynamics model in Chapter 5). Different typologies of landscape agent can have different ecological built-in sub-models (e.g., Figure 2.5).

2.4 Computer platform for MAS

MAS modeling has a symbiotic relationship with computing technology. As the technology has grown in power, the scale and sophistication of the computer platforms available for MAS modelers has increased. It is useful to distinguish three types of MAS platforms based on the state of development: i) generic object-oriented programming languages, ii) libraries of standardized routines, and iii) MAS packages.

2.4.1 Generic object-oriented programming (OOP) languages

Since the early 1990s, most agent-based models have been developed though implementing directly conventional OOP languages such as *Turbo Pascal*, *C*$^{++}$, *Java*, *SmallTalk*, *Lisp* and some others (Gilbert and Bankes, 2002). Although conventional languages such as *C*$^{++}$ and *Turbo Pascal* can be used for programming MAS models, high-level languages such as *Lisp* – a common language for artificial intelligence programming (Russell and Norvig, 1995; Gilbert and Troitzsch, 1999) – potentially give comparable advantages in programming MAS. The first advantage of high-level languages is that a great deal of the routine housekeeping are done behind the scenes and does not have to be written afresh for each program. The others advantages of the particular Lisp system are: it includes a rich set of functions to display graphs and diagrams, it is interactive, and it is in the public domain, available free over the Internet (Gilbert and Troitzsch, 1999).

However, in general there are several disadvantages of using a general-purpose language for MAS. First, with these languages alone, every modeller has to re-implement basic algorithms. Second, the graphics libraries of such languages are often ill-suited to visualization, especially for spatial dynamic modeling. Finally, the programming code is easily accessible only to those familiar with the language and the compiler needed to run it (Gilbert and Bankes, 2002).

2.4.2 MAS libraries/toolkits

The next development in MAS computer platforms was the distribution of libraries of standardized routines that could be included in one's own purpose-built program. One of the first system of libraries for programming MAS was SWARM (Minar *et al.*, 1996), which has been the favored tool of many MAS modelers in the last decade. SWARM offers researchers a small number of *general purpose toolkits*, which can be assembled to create MAS simulations more easily than starting from first principles. The toolkit is written in a language called *Objective C* and the simulations themselves need to be written in the same language (Gilbert and Troitzsch, 1999). Thus, using SWARM normally requires a good programming skill (Najlis *et al.*, 2002). However, SWARM's users have more freedom, as the system is generic and not necessarily tied by any specific structure of applications (Bousquet and Le Page, 2004).

Some other library systems that are much more user-friendly were developed by the end of 1990s. Typical systems of this type are REPAST (Recursive Porous Agent Simulation Toolkit) and ASCAPE. REPAST is a MAS simulation kit created by Social Science Research Computing at the University of Chicago. It is an integrated library of Java classes that allows programmers to build simulation environments (e.g., regular lattices), create agents in social networks, collect data from simulations automatically, and build user-interfaces easily. REPAST borrows many concepts from the SWARM library. However, REPAST is different from SWARM since it has multiple pure implementations in several languages and built-in adaptive features such as genetic algorithms and regression (Gilbert and Bankes, 2002). ASCAPE is another Java-based library system created by Miles Parker at Brooking Insitute (Parker, 2001). ASCAPE is derived from the programs developed for Epstein and Axtell's *Growing Artificial Societies* (Epstain and Axtell, 1997), generalized to allow a wider range of models and refined to provide more powerful features (Gilbert and Bankes, 2002).

These library systems have great advantages, but also certain limitations. As more complex algorithms, toolkits, and libraries have been developed, more sophisticated models have become feasible for researchers working on their own or in small teams. Both ASCAPE and REPAST are excellent for simulations involving agents located on a rectilinear grid. However, these toolkit systems may pose particular limitations. Although some models are much easier to program than conventional OOP languages, they still require users to have a good working knowledge of the programming language that they are aimed at (e.g., Java in the case of ASCAPE and REPAST). Moreover, the construction of specific models using other approaches can be hindered by the need to find ways of working around the built-in assumptions, i.e., modelers have limited freedom in developing models. They are also less useful for models that require a GIS to simulate an actual terrain (Gilbert and Bankes, 2002).

2.4.3 MAS packages

The most recent development of MAS platforms is the appearance of MAS packages. Differing from the MAS library/toolkit system, the MAS package is a collection of routines assembled with a common standardized user interface, thus really providing an environment for MAS modeling. However, it is still at an early development stage, current MAS packages still have nothing in common with the scale or sophistication of drag-drop best-known packages (Gilbert and Bankes, 2002). The primary supports of current MAS packages for model use are *visualizations* of model state (especially the ubiquitous displays of two-dimensional grids of agent positions) and some modest facilities such as collecting statistics in a single run. One of the first MAS packages is *StartLogo*, which is an extension of the *Logo* programming language, belonging to the language family of *Lisp*. *StarLogo* is a programmable modeling environment that is well-suited for exploring the behaviors of complex decentralized systems. However, since it is designed especially for educational and demonstrational purposes in schools, some sacrifices in functionality have had to be made (Gilbert and Bankes, 2002).

Some packages based on *SmallTalk* language have been developed, including SDML, CORMAS and DESIRE. These are more complex and powerful than *StartLogo*, but may require longer time to learn. Unlike the *Java* libraries such as ASCAPE and REPAST, they do not demand that users are fluent in the underlying programming language (i.e., *SmallTalk*), but they do require users to learn a complex interface that can be as difficult to master as a full programming language (Gilbert and Bankes, 2002).

NetLogo (Wilensky, 1999), a really multi-agent programming language and modeling environment, is particularly well-suited for modeling complex systems evolving over time. This makes it possible to explore connections between micro-level behavior of individuals and macro-level patterns that emerge from their micro interactions. Historically, *NetLogo* is the next generation of the series of multi-agent modeling languages including *StarLogo* (Resnick & Wilensky, 1993; Resnick, 1994). *NetLogo* is a stand-alone application written in Java so it can run on all major computing platforms. As a language, *NetLogo* is a member of the *Lisp* family that supports agents and concurrency. The package is comprehensive enough for users to easily run simulations and "*play*" with models to explore natural and social phenomena under various conditions, and it is advanced enough to serve as a powerful programming tool for researchers in many fields (Tissue and Wilensky, 2004). NetLogo also comes with a rich library of example models, i.e., a large collection of simply pre-written simulation models that can be used and potentially modified to serve the modeler's own purpose. Given these advantages, this package was chosen for this research.

In general, computer platforms for MAS are still at an early development stage. The facilities for other phases of a model's life cycle, model evaluation, model maintenance, and many types of model use are rather limited at present (Gilbert and Bankes, 2002). Tool developers have not yet confronted issues of comparing multiple model runs, loading or calibrating models from data, automatically generating large numbers of cases from experimental designs, or collecting and statistically analyzing the results of large numbers of experiments. "*Minimal*" simulation models - which incorporate some aspects of an actual system without incorporating a fine level of detail about the entire system (ALC, 2003) - may be easily extended from similar prototypes available in library systems. However, for the development of a MAS simulation model for really complex systems, such as the human-environment system associated with LUCC, a great deal of interdisciplinary knowledge about the studied system and programming skills are required (ALC, 2003).

2.5 VN-LUDAS: A proposed conceptual MAS framework for modeling LUCC

Following the mindset of MAS and the synthesis of generic framework for the coupled human-environment system proposed by Haggith *et al.* (2003) and Freudenberger (1995), a conceptual framework (Figure 2.6) is proposed in this study as a basis for further specification of a MAS-LUCC model, called VietNam - Land-Use DynAmics Simulator (VN-LUDAS), in the next chapter. The key features of the conceptual model are briefly described as follows.

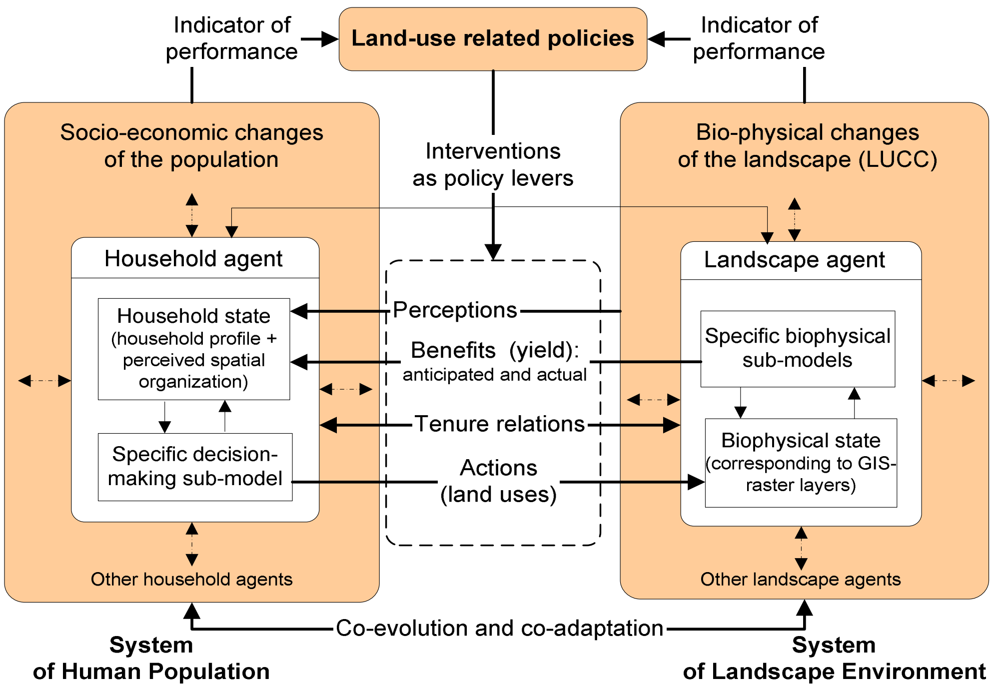


Figure 2.6 The conceptual VN-LUDAS framework

2.5.1 Formulising the system of landscape environment

The system of landscape environment is considered at the level of *landscape agent*, i.e., congruent land patches (30 m x 30m) with their own attributes and ecological response mechanisms to environmental changes and human interventions. Landscape agent is the minimal spatial unit for measuring spatial variables, i.e., spatial divisions within a landscape agent are not represented. State variables of landscape agents are corresponding to *GIS-raster layers* of biophysically spatial variables (e.g., terrain condition, land cover, accessibility to rivers/streams), economically spatial variables (e.g., proximate distance to roads), institutionally spatial variables (e.g., owner, village territory, protection zoning class), and histories of particular patch properties.

Ecological response mechanisms of landscape agents are represented by *internal sub-models of agricultural and forest productivity dynamics*, which work in response to the current state, history, and (sometimes) spatial neighbourhood of the landscape agents. Due to the regularly autonomous working of such ecological sub-models, many attributes of landscape agents change over time, even without any intervention of humans. A cellular automata sub-model of land-cover transition built into every landscape agent enables it to transform through small and gradual changes (i.e., vegetation growth and/or modification) into categorical changes (i.e., conversions) in land cover. As a consequence, annual landscape patterns are self-organised from micro-changes at pixel level.

2.5.2 Formulising the system of human population

The human system is considered in terms of *household agents*, i.e., heterogeneous farming households with their own state and decision-making mechanisms about land uses. Household agents are the minimal unit for measuring human variables, i.e., human divisions within households (i.e., household members) are not represented. The state of house agent includes *household profile* and *spatial organisation* perceived within the agent's vision (i.e., "*the sphere of influence*" in Figure 2.1). The household profile is represented by a long list of household socio-economic variables (e.g., educational status, household size, labor, land endowment, income, etc.) and variables measuring accessibilities of the household to certain policies (policy-related variables). Generally, variables of the household profile change over time, but with different motions. Some variables, e.g., annual income and land endowment, change as a results of annual land-use activities. Policy-related variables change in response to the change of policy factors. Demographic variables, such as household's ethnicity and size, are stable with small stochastic variance, but household age advances regularly over time.

The decision-making mechanism is represented by a *decision-making program*, which works by taking inputs from household profile, perceived spatial organisation, and sometimes information from other household agents (e.g., when checking land ownership of a given patch). The decision program is a logical procedure that includes both reflex and bounded-rational decision-making mechanisms, especially in location choice. It assumes that household agents behave reactively according to production rules when deciding where to collect forest products, while they are assumed to *likely* select options returning optimal utility when looking for a location for cultivation. The decision program is universal with all household agents in terms of its logical sequence. However, as the agent's state, parameters and event structure of utility functions are individual specific, decision outcomes are extremely diverse.

2.5.3 Means of human-environment interactions

Human-environment linkages are mainly characterised by i) *tenure relations* and ii) the *perception-response* loop. *Tenure relations* between agents and patches are explicit *rules* regulating the househols access to and usage of land resources, possibly *de facto* and/or *de jure*. Although there may be many forms of *de facto* tenure relationships, the model focuses on *ownership* and *village territory,* as they are often strongest tenure rules shaping land use. Ownership is a tenure relation applied specifically for an individual household agent, i.e., the holder of the land. Village territory is a tenure relation applied specifically to a group of household agents who share the same village membership. *De jure* tenure regulations, which are defined by policies, are applied similarly for all household agents in the system.

The *perception-response loop* involves information/physical flows between household agents and their landscape environments. *Perception* corresponds to the perceived spatial status of biophysical condition around them and *anticipated benefits* that the household can make use of in arriving at decisions. When household agents use lands, they receive some *actual benefits* (e.g., agricultural products) that can lead to changes in certain attributes of their profile, thus the interaction means now become physical. *Response* reflects the physical effects of household agents on the environment through their land-use actions. Through practicing land-use activities, the household agents modify the structure of spatial organisation in their environments, which may happen within the vision of other households, thus indirectly affecting the behavior of the others.

2.5.4 Land-use related policies as external drivers

Land-use related policies are treated as external drivers of the system (i.e., not changes as consequences of system processes), rather modified by model users for exploring the likely policy outcomes. At a first glance, the pathway of policy influence on system behavior is through modifying directly the functional relationships between the human and environment system. Since interactions within the MAS are determined by internal mechanisms of agents, policy factors affect the system behavior throught three pathways: i) affect policy-related variables of household agents, ii) affect institutional variables of landscape agents, and iii) add/cancel/modify directly interaction rules. Some policy factors, such as subsidy of agricultural inputs or agricultural extension, potentially affect policy-related variables in the household profile, thus leading to changes in the outcomes of household decisions. Some other policy factors, such as zoning regulation, can modify spatial patterns of ownership and protection zones and thus can lead to changes in the spatial organisation perceived by the household agents. Some policy factors, such as forest-use regulations, may add new rules, or cancel old rules, or modify existing rules for harvesting forest products.

2.6 Selection of the study site

Covering more than 70 % of the national territory and being home for at least one-third of the population, the uplands are understandably a major concern of the Vietnam government (Jamieson *et al.*, 1998; Rambo, 1995; Chu Huu Quy, 1995). The Vietnam uplands are home to the poorest communities, whose welfare and livelihoods depend directly on the natural resource base. Sustainable management of lands, forests and other natural resources of the uplands are critical not only to benefit local people but also for the national interest. Therefore, achieving the successful development of these vast areas of mountains and forests is a matter of high priority (Nguyen, 1998; Rambo, 1995).

Vietnamese tropical forests have been rapidly declining over the last 50 years. Between 1943 and 1993, the national forest coverage had declined from at least 43 % to

20% (Vo Quy, 1996) or even to as low as 16 % (WCMC, 1996). This impressive forest loss has been accompanied by the rapid spread of denuded or barren lands. Although a substantial proportion of forested lands has been replaced with permanent agriculture, another proportion, at least equally extensive, has been left barren (De Koninck, 1998). This is generally the result of a few years of excessive cropping, followed by the land being abandoned and left prone to erosion, with the result that rapid leaching of the topsoil occurs and the land becomes literally barren (De Koninck, 1998). Apparently, by the early 1990s, as much as 40 % of Vietnam's land was in such a state (Vo Quy and Le Thac Can, 1994).

The Central Coast of Vietnam consists of 13 provinces from 20° to 11° North latitudes, covering about 97,000 square kilometers with 15 million inhabitants (Bui Dung The, 2001). The whole region can be described as having a very narrow and long strip of flat coastal land in the east and a much larger hilly and mountainous area in the west. Roughly, more than two thirds of the Central Coast consists of hills and mountains. Compared with other regions of the country, this area is the poorest and least developed agriculturally mainly due to a hostile climate and a poor natural resource base. Thus, management of land resources, especially upland resources, is of crucial importance to the region.

The Hong Ha watershed, which has been chosen for implementing this MAS-LUCC study, lies in mountain area in the west of the Thua Thien – Hue province, Vietnam Central Coast. The watershed is within the upstream area of Bo River, which is the main water body in the agricultural plain in the north of the Thua Thien - Hue province. The watershed is a good representative for a mountainous upland zone of the Central Coast of Vietnam (Le Van An, 1999; Le Quang Bao, 1999). The fundamental reasons why this study area was selected for MAS-LUCC study are: i) the heterogeneous nature of biophysical conditions, ii) the diverse livelihood patterns of local farming households living in the forest margins, and iii) the need to formulate policies balancing nature conservation and economic development purposes in the area. The complex terrain of the watershed creates a patchy landscape of land-use suitability, which in part results in complex spatial patterns of land cover (see Chapter 5).

Like many rural communities living in the forest margins, farming households in Hong Ha have diverse livelihoods reflecting a complex behavior in land and forest use. A household often holds many plots of different cultivation types (i.e., swidden, paddy, garden and even forest plantations). At the same time, it may also raise livestock and gather forest products (see Chapter 4). These diverse livelihood patterns suggest a complex behavior in land use in which farmers often face trade-offs among various options in a diverse environment. Understanding how farmers make decisions about land use in this complex situation is crucial for improved land management and planning.

It has been debatable how to formulate effective land management policies balancing different demands and interests of the main stakeholders in land and forest management. As the area is located in critical watershed areas, in 1991 the government declared the area as a protected watershed, according to which all natural forests are principally protected and slope lands are designated for planting protection forests. However, the local communities, who are forest-dependants, are creating high pressures on land and forest resources for maintaining their livelihoods (Le Quang Bao, 2002). Along with emerging trends of national policy innovation, and under the facilitation of some participatory development projects, stakeholders in the area have generally agreed to move towards a better-integrated management of the watershed to improve both local livelihoods and ecological functions of the watershed. However, negotiation processes for a viable improved management of the watershed are progressing slowly due to a lack of mutual and insightful understanding of trade-offs of the land-use system, as well as ambiguous understanding about potential outcomes of proposed alternatives. It needs feedback tools to foster these effective multi-stakeholder processes.

Given the research objectives, the VN-LUDAS should be developed in conjunction with the reality of the Hong Ha watershed, since the site is a good representative for the uplands in the Central Coast of Vietnam. Moreover, as in many parts in central Vietnam, the Hong Ha watershed has received many interventions from the government for both resource protection and rural development. Therefore, findings of this research can be applied widely.

2.7 Modeling steps

The simulation-sound process for obtaining the operational VN-LUDAS consists of 10 steps as illustrated in Figure 2.7. According to Banks (1999), a simulation-sound modeling process is rather more iterative than strictly a series of steps of a strict order. Step 1 and 2 were done in chapter 1 and this chapter, respectively. Step 3 is totally solved in Chapter 3, including the construction of a system structure (organisation), specification of algorithms of sub-models built into household and landscape agents, algebraic parameterization, and pseudo-coding.

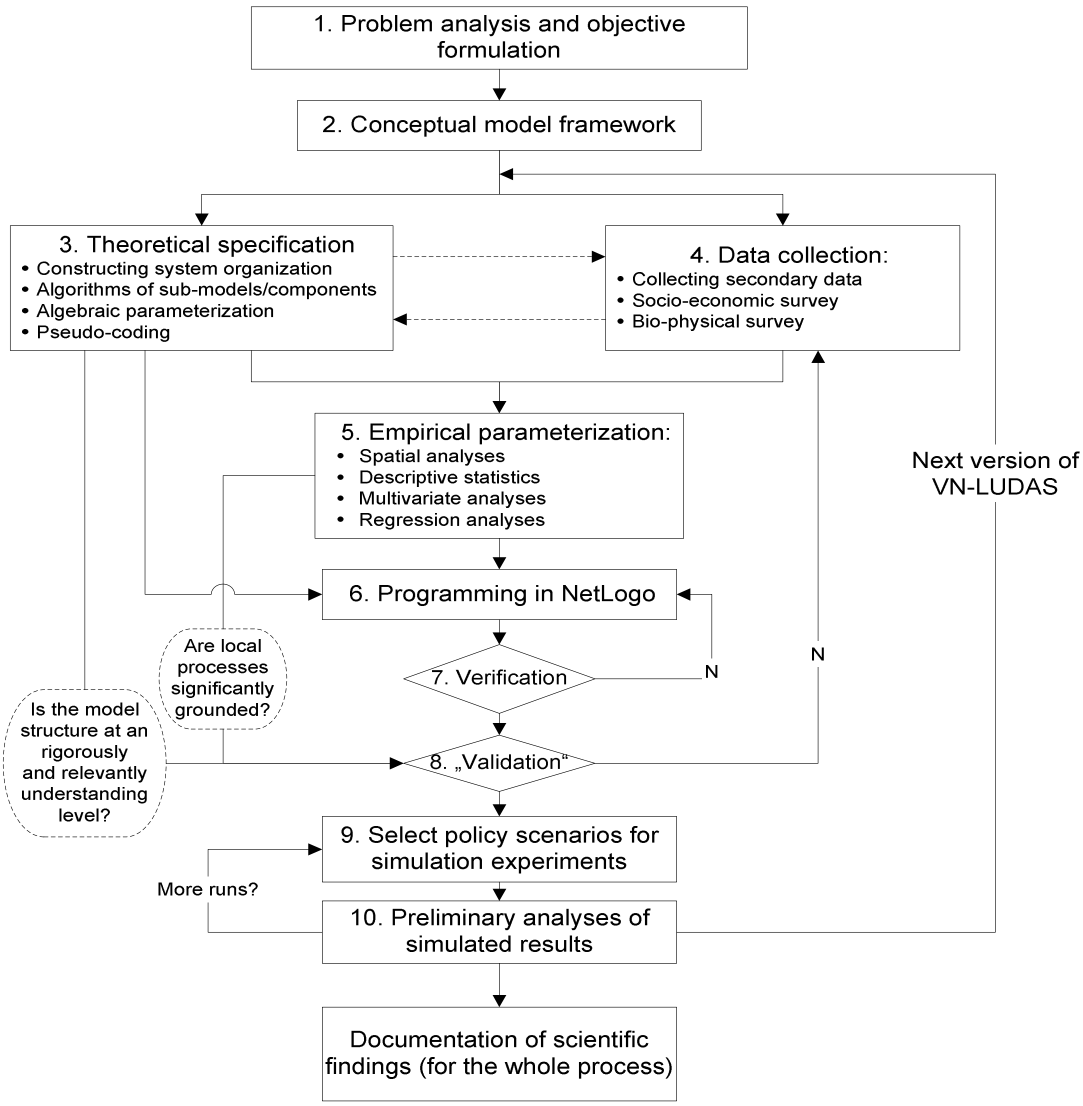


Figure 2.7 A simulation-sound process for building VN-LUDAS model

Step 4 consists of a themetic survey in the Hong Ha watershed to collect data. The socio-economic and biophysical surveys are decribed in the methodological parts of Chapters 4 and 5, respectively. As shown in Figure 2.7, this step was conducted together with Step 3. Step 5 comprises the statistic and spatial analyses for empirical parameterization of sub-models of human agents (Chapter 4) and landscape agents (Chapter 5).

Step 6 is the coding (programming) of the model theoretically specified in Step 3, taking the empirical values estimated in Step 5, into the NetLogo platform to produce the first operational VN-LUDAS model. Step 7 is the verification of the operational VN-LUDAS model. Verification is to technologically check the performance of the operational model, detecting bugs/mistakes and improving the codes. Thus, verification (Step 7) and programming (Step 6) are two iterative processes (Figure 2.7).

Step 8 is model validation. This is the most questionable issue that MAS modeling often faces. In fact, validation in a simulation study is, first and foremost, the determination of how rigorously the modelled structure and process *represent* the real system (Banks, 1999; Bousquet and Le Page, 2004). Thus, the creditability of a simulation model firstly lies in: i) how the real system (structure and interactions) is adequately conceptualized (mapped) into the MAS model, i.e., the question of *system representation* (Step 3), and ii) how well the parameters of the agent's decision processes are *empirically grounded*, i.e., the question of data availability at the agent level (Step 4) (Verburg *et al*., 2002). Of course, one can use the classical procedure for validation, which is to compare simulated outputs to observed data, which is widespread for validating statistic/analytic models. However, Verburg *et al*. (2003) note that observed LUCC outcomes are not sufficient to validate the MAS-LUCC model. Another classical approach for validation is sensitivity analysis. However, Bousquet and Le Page (2004) remark that sensitivity analysis for complex models is not very often used. In short, to explain why these two classical validation procedures are not suitable for MAS-LUCC, Batty and Torrens (2000: 6) note "*we persist in developing models that are intrinsically complex, but which we attempt to validate against some reality which we present as intrinsically simple*" [simulated data *vs.* observed data].

Steps 9 and 10 are policy experiment design and scenario analyses, respectively, which are dealt with in Chapter 6. The two steps are iterative. For each scenario that is to be simulated, decisions need to be made concerning the length of simulation runs, the number of runs (replications), and the manner of initialization. Simulation runs and their subsequent analysis are used to estimate measures of performance for the scenarios that are being simulated. Based on the analysis of runs that have been completed, the tester determines if additional runs are needed and if any additional scenarios need to be simulated.

3 THEORETICAL SPECIFICATION OF VN-LUDAS: A MULTI-AGENT SYSTEM FOR SIMULATING LAND-USE AND LAND-COVER CHANGE

3.1 Introduction

There is an increasing interest in using multi-agent system (MAS) tools for modeling land-use/cover change (LUCC). MAS applied for LUCC (MAS-LUCC) has been recognized to be highly appropriate for representing the complex nature of both spatial interactions and decentralized human decision-making on land use (Batty, 2001; Ligtenberg *et al*., 2001; Parker *et al*., 2001; and Parker *et al*., 2002), where LUCC and associated population dynamics are self-organizing processes emerging from interactions of autonomous agents (Deadman, 1999).

From the aspect of system representation, the MAS provides a *natural description* of a human-environment system. Firstly, MAS architecture make it possible to map the concepts and structures of the real world onto the model in ways that preserve natural objects and connections (Parker *et al*., 2003; Sawhney *et al.,* 2003; Bonabeau, 2002; Batty, 2001). Secondly, the MAS is useful for formalizing the "natural" behavior of human agents, which are normally complex, activity-based, and stochastic to a particular degree and not uniformly transitional (Bonabeau, 2002). Natural structure of MAS and formulised natural behavior of agents allow validation and calibration of the model through expert judgment (Bonabeau, 2002), as well as participation of stakeholders in model-building processes.

MAS models give users a great advantage in terms of high *flexibility* (Parker *et al*., 2003; Gilbert and Troitzsch, 1999). The flexibility of MAS can be recognised along multiple dimensions, such as: i) ease of adding more agents, ii) providing a natural framework for tuning the complexity of the agents (behavior, degree of rationality, ability to learn and evolve, and rules of interactions), and iii) higher ability to change levels of description and aggregation (e.g., aggregate agents, subgroups of agents, and single agents, with different levels of description coexisting in a given model) (Bonabeau, 2002).

Moreover, recently developed MAS computer platforms support visualising simulated LUCC outcomes in a *spatially explicit* way. Recent platform availability supports stronger development of MAS models for decision support purposes. Spatial MAS models with user-friendly graphic user interface (GUI) can serve as virtual social laboratories to enable stakeholders to foresee potential LUCC outcomes under different land management/planning alternatives, thus to foster effective multi-stakeholder negotiation processes.

In order to achieve credability and transparency for a MAS model, specifications should focus on two different aspects: i) *system architecture* and 2) *system implementation* (see Cioffi-Revilla and Gotts, 2003; Bousquet and Le Page, 2004). Specification of system architecture refers to the rigorous representation of the structure of the MAS-LUCC model in terms of agents and their internal structure. System implementation refers to the specification of the computational procedure applied for the specified architecture (simulation). Accordingly, the specific objectives of this chapter are:

i) to construct a fully parameterised architecture of VN-LUDAS based on the conceptual model described in the previous chapter, and
ii) to develop a simulation protocol for the VN-LUDAS architecture.

3.2 Specification of VN-LUDAS architecture

Following the conceptual model laid out in Chapter 2, a framework representing the coupled human-environment underlying land-use and land-cover changes are elaborated in more detail. The framework (named VN-LUDAS) consists of four modules (also see Figure 3.1):

i) The HOUSEHOLD-POPULATION module represents the system of human population in which farming households are treated as human agents,
ii) The PATCH-LANDSCAPE module represents the system of landscape environment in which congruent land patches are considered as environment agents,
iii) The DECISION module (program), that is indeed a DECISION program built into household agents for their autonomous behavior; it is however temporarily considered as a separate module for easier designation. This

module plays the core role in determining the performance of the whole model, and

iv) The GLOBAL-POLICY module includes selected policy factors and some other external factors, which are supposed to be modified or set by model users for exploring potential impacts of policy alternatives.

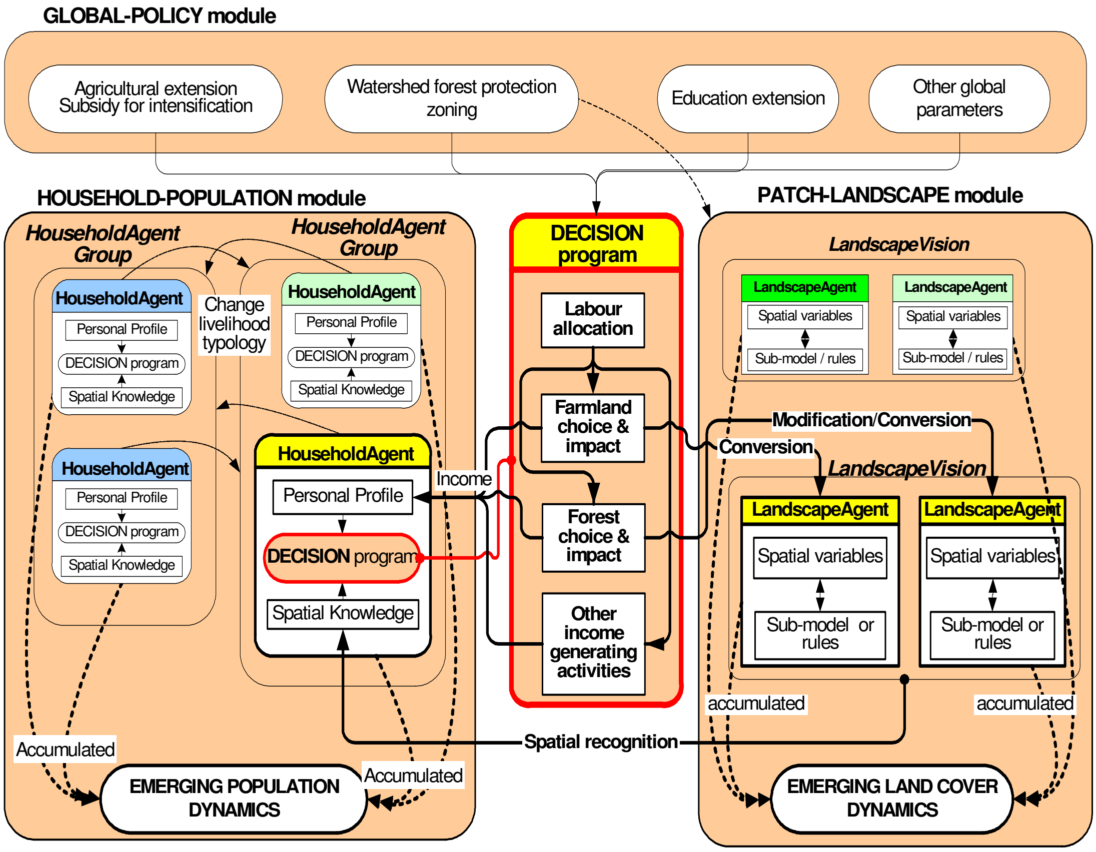


Figure 3.1 Block diagram illustrating the overall structure of VN-LUDAS model as a MAS

3.2.1 System of human population: The HOUSEHOLD-POPULATION module

Module overview

The module represents the dynamics of human population emerging from spatial local interactions among household agents and their environment (including other household agents). The human system is self-organised in a hierarchy of three organisation levels: *household agent*, *group of household agents*, and *population* (see Figure 3.2).

Household agent (named *HouseholdAgent*) represents individual farming households, consisting of two basic components: a spatial knowledge about the environment and a model of himself. The agent's spatial knowledge is maintained through a link with his *landscape vision* (named *LandscapeVision*) (see Figure 2.1, Chapter 2). In other words, *HouseholdAgent* has access to his own *LandscapeVision* environment. The agent's model of himself consists of his dynamic *personal profile* and program of instructions for generating his behavior under different circumstances, named DECISION program. Although the DECISION program applies to all agents, since the program employs agent personal data to run, the generated agent's behavior is definitely individual-specific. Household agents can interact directly or indirectly.

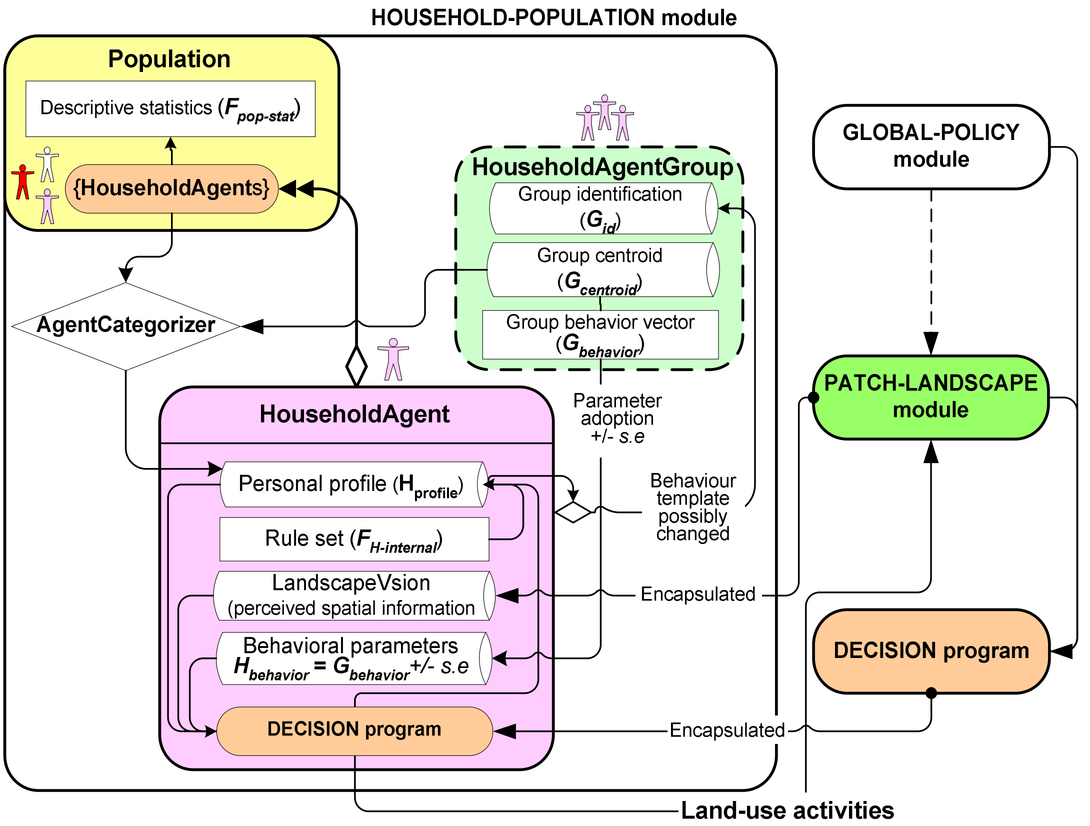


Figure 3.2 Block diagram showing agent-based architecture of HOUSEHOLD-POPULATION module

Group of household agents (named *HouseholdAgentGroup*) is a collection of household agents having similar production typology, thus assumed to have the same template of land-use behavior. Behavior parameters are outside the model based on empirical group data, later entered and stored in *HouseholdAgentGroup*, then adopted by agents within the group with random errors. Moreover, household agents have a mechanism to change their production strategies and subsequently adopt new behavior structures based on comparing their emerging characteristics to all production typological groups in the population. Although these behavior parameters are static at the group level, when accumulative changes of agent characteristics go far enough, the agent will join another group. The choice of a new group is determined by the greatest similarity between him and other groups. Once a household agent joins a new group, he will adopt the behavior template of the new group. All agents are re-categorized every year based on a *HouseholdCategorizer* routine working regularly.

Population (named *Population*) is a collection of all renewed and interactive agents, thus its patterns are results emerged from agent-based processes at the bottom of the hierarchical system. Simple statistic programming procedures are used for calculating some socio-economic indicators of population dynamics.

Structure of household agent

The structure of a *HouseholdAgent* is formally expressed as follows:

$$HouseholdAgent = \{H_{profile}, LandscapeVision, H_{behavior}, F_{H\text{-}internal}, \mathrm{DECISION}\} \quad (3.1)$$

where $H_{profile}$ is a set of variables in the personal profile, *LandscapeVision* is spatial information the household agent perceives from the landscape; $H_{behavior}$ is a set of behavior parameters the household agent uses in his decision process, $F_{H\text{-}internal}$ is a set of rules performing the dynamics of variables in $H_{profile}$, and DECISION is a high-order logical procedure to perform sequential processes of decision-making and actions about land use. Since the DECISION is quite a complicated procedure, we specify it as a separate module as will be seen in later sections. However, the DECISION works as an internal model of the household agent.

Personal profile ($H_{profile}$)

In general, agent profile ($H_{profile}$) includes four sub-types of variables: social identity (H_{social}), human resources (H_{human}), land resources (H_{land}), economic resources (H_{income}), and policy-related attributes (H_{policy}):

$$H_{profile} = \{H_{social}, H_{human}, H_{land}, H_{income}, H_{policy}\} \tag{3.2}$$

Social identity of the agent (H_{social}) includes identification code (H_{id}), age (H_{age}), village ($H_{village}$), group membership (H_g), ethnicity (H_{ethnic}) and leadership (H_{leader}):

$$H_{social} = \{H_{id}, H_{age}, H_{village}, H_g, H_{ethnic}, H_{leader}\} \tag{3.3}$$

Notice that village ($H_{village}$) and group membership (H_g) are variables accounting for the *positions* of the household agents in the population: which village do they belong to? Which group are they similar with? Thus, these typological variables are important for calling specific sets of household agents in the simulation program.

Agent's human resources (H_{human}) consists of household size (H_{size}), labor pool (H_{labor}), dependency ratio (H_{depend}), and education status of the household head (H_{edu}):

$$H_{human} = \{H_{size}, H_{labor}, H_{depend}, H_{edu}\} \tag{3.4}$$

Household land resources (H_{land}) are indicated by total area of land holdings ($H_{holding}$) and its land-use composition vector ($[H_{\%i}]_{i=(1,M)}$), i.e.,a vector of percentage area of each land use of total holding area:

$$H_{land} = \{H_{holding}, [H_{\%ui}]_{\mathrm{i=(1,M)}}\} \tag{3.5}$$

where i indexes land-use types.

The component household income (H_{income}) comprises gross income per capita ($H_{grossincome/pers}$) and an income composition vector ($[H_{\%ins}]_{\mathrm{s=(1,S)}}$), i.e., a vector of percentage income components of annual gross income:

$$H_{income} = \{H_{grossincome/pers}, [H_{\%ins}]_{s=(1,S)}\} \quad (3.6)$$

where s indexes sources of income (i.e., production components).

The variables in an agent's personal profile are shown in Figure 3.3.

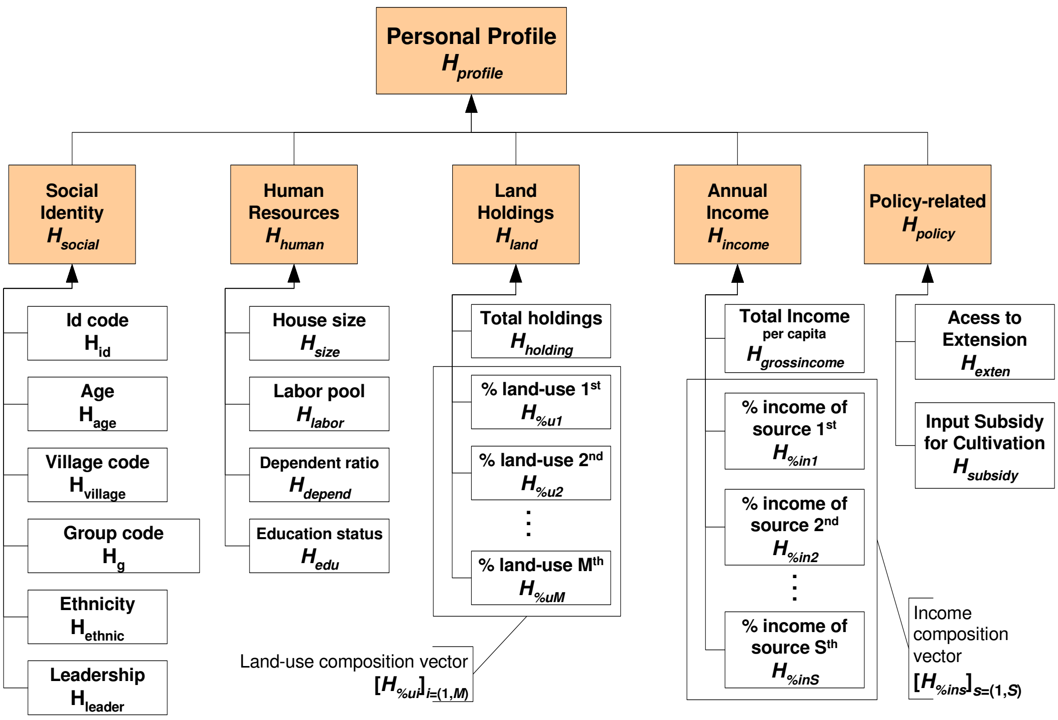


Figure 3.3 Schema shows components and attributes of agent profile ($H_{profile}$).

Agent's policy-related attributes include agent's access to agricultural extension services (H_{exten}) and material subsidy for intensifying cultivation ($H_{subsidy}$):

$$H_{policy} = \{H_{exten}, H_{subsidy}\} \quad (3.7)$$

Rule set ($F_{H\text{-}internal}$)

There are three kinds of changes in variables in $H_{profile}$. First, *natural changes*, which occur regularly without any effects of agent actions, e.g., the increase of the age of the household head and the small stochastic variation of the household labor. Second, *external-driven changes*, which occur due to interventions from outside the system, e.g., change of education status due to education extension program. Third, *impacts*, which occurs due to the effects of household agent activities during simulation. The first and second kind of changes are controlled by the rule set $F_{H\text{-}internal}$ that we are going to describe below. The later changes are specified in the DECISION module in later sections.

Formally, the rule set $F_{H\text{-}internal}$ is expressed as follows:

$$F_{H\text{-}internal} = \{F_{H\text{-}age}, F_{H\text{-}edu}, F_{H\text{-}size}, F_{H\text{-}labor}, F_{H\text{-}depend}, F_{H\text{-}exten}, F_{H\text{-}subsidy}\} \quad (3.8)$$

where $F_{H\text{-}age}$, $F_{H\text{-}village}$, $F_{H\text{-}edu}$, $F_{H\text{-}size}$, $F_{H\text{-}labor}$, $F_{H\text{-}depend}$, $F_{H\text{-}exten}$ and $F_{H\text{-}subsidy}$ are rules performing the dynamics of the variables H_{age}, $H_{village}$, H_{edu}, H_{size}, H_{labor}, H_{depend}, H_{exten} and $H_{subsidy}$, respectively. The mathematic details of these rules are as below.

- *Rule* $F_{H\text{-}age}$: ${}^{t+1}H_{age} = F({}^{t}H_{age})$

Age of the household head (H_{age}) will increase 1 year after each time step, until reaching the empirical upper bound max(H_{age}). When the age of the household head reaches the upper bound, it is reasonable to assume that another household member will replace him/her to take responsibility. The age of this new household head will be approximated at about the mean age of the agent group, being stochastically within the standard error range ($\pm\,\sigma$) of the mean age (of the population):

$${}^{t+1}H_{age} = \begin{cases} {}^{t}H_{age} + 1 \quad , & if \quad {}^{t}H_{age} < \max(H_{age}) \\ round(\overline{H}_{age} - \sigma_{age} + random(2\sigma_{age}), & if \quad {}^{t}H_{age} \geq \max(H_{age}) \end{cases} \quad (3.9)$$

where $\overline{H}_{age}$ and $max(H_{age})$ are respectively the mean age and the maximal age, respectively, of the agent group, *round*(a) rounds the number a to the integer, and *random*(a) ($a > 0$) randomly generates a floating number within [0, a].

- *Rule* $F_{H\text{-}edu}$: ${}^{t+1}H_{edu} = F({}^{t}H_{edu})$

The education status of household agent (H_{edu}) may change over time. If the agent was educated by the previous time step (${}^{t}H_{edu}=1$), he will be still educated in the next time step (${}^{t+1}H_{edu}=1$); otherwise, he will have a chance θ_{edu} of becoming educated in the next time step. If θ_{edu} is given by educational extension policy, education status of the agent next step can be formulised as follows:

$$^{t+1}H_{edu} = \begin{cases} 1, & if\left({}^{t}H_{edu} = 0 \quad and \quad q \leq \theta_{edu}\right) \\ {}^{t}H_{edu}, & otherwise \end{cases} \tag{3.10}$$

where q is a random number distributed evenly over [0,1], θ_{edu} is a tuneable point within [0,1] representing the chance for a uneducated agent to access education, approximated by the annual proportion of uneducated agents receiving education as expected by the educational extension policy.

- *Rules* $F_{H\text{-}size}$: ${}^{t+1}H_{size} = F({}^{t}H_{size})$, $F_{H\text{-}labor}$: ${}^{t+1}H_{labor} = F({}^{t}H_{labor})$, *and* $F_{H\text{-}depend}$: ${}^{t+1}H_{depend} = F({}^{t}H_{depend})$:

The dynamics of household size (H_{size}), labor pool (H_{labor}), and dependency ratio (H_{depend}) are indeed event-driven phenomena, potentially affected by many causes beyond the consideration of this study. It is, therefore, quite difficult to extract deterministic rules of these dynamics for specific households based on common demographic datasets of communities. Here, we proximate stochastically the values of these three household attributes within uncertainty ranges of the value by the previous time steps. The lower and upper bounds of the uncertainty range are defined by standard error (σ) calculated from empirical household datasets. The bounded-random rule applied for these three variables are expressed as following:

$$^{t+1}H_{size} = round({}^{t}H_{size} - \sigma_{size} + random(2\sigma_{size})) \tag{3.11}$$

$$^{t+1}H_{labor} = round({}^{t}H_{labor} - \sigma_{labor} + random(2\sigma_{labor})) \tag{3.12}$$

$$^{t+1}H_{depend} = {}^{t}H_{depend} - \sigma_{depend} + random(2\sigma_{depend}) \tag{3.13}$$

where *random(a)* gives a random point floating evenly within [0, *a*]. Equations 3.11, 3.12 and 3.13 express *bounded-random behaviors* of the related attributes. If the probability distributions of these attributes are specified (possibly learned from the sampled dataset), random numbers within the uncertainty range can be generated according to these probability distribution functions using advanced random generators in *NetLogo* 2.1 package (Wilenski, 1999).

- *Rules* $F_{H\text{-}exten}$: ${}^{t+1}H_{exten} = F({}^{t}H_{exten})$ *and* $F_{H\text{-}subsidy}$ ${}^{t+1}H_{subsidy} = F({}^{t}H_{subsidy})$:

Similar to education status, the dynamics of a household agent's access to extension services (H_{exten}) and agrochemical subsidies ($H_{subsidy}$) are controlled by the GLOBAL-POLICY module:

$$ {}^{t+1}H_{exten} = \begin{cases} 1, & if \quad q \le \theta_{exten} \\ 0, & if \quad q > \theta_{exten} \end{cases} \tag{3.14}$$

$$ {}^{t+1}H_{subsidy} = \begin{cases} W_{subsidy} & if \quad q \le \theta_{\%subsidy} \\ 0, & if \quad q > \theta_{\%subsidy} \end{cases} \tag{3.15}$$

where q is a random number distributed evenly over [0,1]. θ_{exten} and $\theta_{\%subsidy}$ are tuneable points within [0,1], representing the chance of a household agent accessing extension services and subsidy programs, respectively (i.e., $Prob({}^{t+1}H_{exten}=1) = \theta_{exten}$, $Prob({}^{t+1}H_{subsidy} = W_{subsidy}) = \theta_{subsidy}$). θ_{exten} and $\theta_{subsidy}$ can be approximated by the annual proportion of household agents who have received extension services as expected by agricultural extension and subsidy policies. $W_{subsidy}$ is the annual average subsidy amount the household agent received, set by the subsidy policy.

Behavior parameters ($H_{behavior}$)

The behavior parameters ($H_{behavior}$) include a *vector of preference coefficients* ($[H_{\beta}]$) used for computing utility functions and land-use choice probabilities, and a vector of strategic labor allocation percentage ($[H_{labor\text{-}strategy}]$) used to partition the agent's labor pool into labor budgets for each production component:

$$H_{behavior} = \{[H_{\beta}], [H_{labor\text{-}stategy}]\} \tag{3.16}$$

The vectors of preference [H_β] and labor allocation coefficients [$H_{labor\text{-}stategy}$] are adopted from the agent group random errors (i.e., $\pm\ \sigma_\beta$). Thus, the behavior parameters of a household agent are formally expressed as:

$$[H_\beta] = [\beta_{ga} \pm \sigma_{ga}]_{\ a\in A} \tag{3.17}$$

$$[H_{labor\text{-}stategy}] = [\%L_{gs} \pm \sigma_{gs}]_{s\in S} \tag{3.18}$$

where a ($a\in(1,2,\ldots,A)$) indexes decision variables (X_a) in the utility function. β_{ga} is the preference coefficient of X_a in the utility function of group g, σ_{ga} is standard error of β_{ga}, s ($s\in(1,2,\ldots,S)$) indexes the production component of group g, $\%L_{gs}$ is percentage labor allocated for production component s of group g, σ_{gs} is the standard error of $\%L_{gs}$, and g ($g\in(1, 2,\ldots, K)$) indexes the agent group.

Structure of *HouseholdAgentGroup*

The *HouseholdAgentGroup* is a collection of household agents having the same structure/template of decision-making behavior, evaluated in terms of particular criteria (so-called *grouping criteria*). Thus, each agent group is considered a *behavior template* storing deterministic behavior parameters (called *group variables*), which were empirically identified and estimated through quantitative case studies outside the model (see chapter 4). The formal expression of the *HouseholdAgentGroup* is:

$$HouseholdAgentGroup = \{G_{id}, G_{centroid}, \mathrm{G}_{\mathrm{behavior}}\} \tag{3.19}$$

where each component is described below.

G_{id} is the group identification code that matches with H_g stored in the household profile (see Figure 3.3). $G_{id} = H_g$, $H_g \in K$

$G_{centroid}$ is the group centroid, which is a vector of mean values of grouping criteria, reflecting the livelihood/production typology of household agents:

$$G_{centroid} = [\overline{H}_{g1}, \overline{H}_{g2}, \overline{H}_{g3}, \ldots, \overline{H}_{gc}] \tag{3.20}$$

where group centroid vector $[\overline{H}_{g1}\ \overline{H}_{g2} \ldots \overline{H}_{gc}]$ consist of the average value of grouping criteria H_1, H_2, …, H_c of all agents within group g.

$G_{behavior}$ is a vector storing deterministic behavior parameters that are identical for all group members. The formal expression of $G_{behavior}$ is:

$$G_{behavior} = \{[\beta_{ga}]_{a\in A}, [\sigma_{ga}]_{a\in A}, [\%L_{gs}]_{s\in S}, [\sigma_{gs}]_{s\in S}\} \quad (3.21)$$

where $[\beta_{ga}]_{a\in A}$ is a vector of deterministic preference coefficients used for computing utility functions and land-use choice probabilities. $[\%L_{gs}]_{s\in S}$ is the vector of labor allocation percentage. $[\sigma_{ga}]_{a\in A}$ is the vector of standard errors of preference parameters, and $[\sigma_{gs}]_{s\in S}$ is the vector of standard error of allocated labor percentage.

All group behavior parameters are deterministic and empirically defined outside the model, while behavior parameters of member agents are created, receiving random values around the fixed parameters of the group, bounded by the related standard error. The vectors of preference $[\beta_{ga}]_{a\in A}$ and percentage allocated labor $[\%L_{gs}]_{s\in S}$ are group-specific in terms of not only concrete values of β_{ga} and $\%L_{gs}$, but also of behavior structure. For instance, a vector of preference coefficients [0.2 -0.5 0.1 0.4 0.8] shows a behavior structure that is totally different from the behavior structure with the preference vector [0.5 -0.1 0.6 0 0]. In the later case, the two last decision variables are not taken into account.

Dynamics of household agent behavior

The application of the same behavior template for all agents of a group raises the issue of how well this design presents the heterogeneous and dynamic nature of household behavior within a real population. For instance, throught time, household agents may change either their spatial preferences or livelihood strategy. Here, agent behavior dynamics are partly represented through types of changes in agent behaviors, described as follows:

Stochastic micro change of agent behavior parameters

In accordance with the *uncertainty* degree of the behavioral rules specified at the group level, there is a stochastic standard error occurring when group parameters are inherited down to instant agents (see equations 3.17, 3.18 and Figure 3.4). These random errors make the behavior parameters list ($H_{behavior}$) of an instanthousehold agent "*slightly*" different from the parameter list of other agents within the same group.

Although behavior parameters of the household agents within a group only differ slightly, the decision-making outcomes (or household responses) differ considerably because the execution of behavioral rules employs household private data. This private data combined with micro variants of behavior parameters will result in heterogeneous decision-making responses in every time step.

Structural change of agent behavior: AgentCategorizer procedure

It is assumed that household agents tend to *copy* the production strategy of the more favored group with the most similar livelihood structure, e.g., a poor household agent with some production success tends to adopt the behavior of a medium group rather than that of a better-off group. This adapting mechanism is handled through the categorising routine *AgentCategorizer*.

AgentCategorizer is an automatic classification procedure that helps the household agent to regularly (every time step) *categorize* himself into the most similar agent group, based on comparing and ranking dissimilarities between himself and all agent groups in the population. The algorithm *of AgentCategorizer* is basically similar to the *K*-mean clustering procedure, except that the group centroids here were predefined outside the simulation model by descriptive statistics of household groups, and thus fixed during the simulation runs. The categorising process consists of three following steps.

First, a given household agent *h* measures dissimilarities about livelihood typology, in term of grouping criteria, between himself and all defined household groups in the population. As measuring units may be different among grouping criteria, *Squared Chi-squared Distance*, a relative form of the standard Euclidean Distance, is used:

$$D_{hg} = \sum_{c=1}^{C} w_c \left[\frac{(H_{h,c} - \overline{H}_{g,c})^2}{\left| H_{h,c} + \overline{H}_{g,c} \right|} \right] \tag{3.22}$$

where D_{hg} is the Squared Chi-squared Distance from household agent h to the centroid (i.e., mean center) of the group g, $H_{h,c}$ is the instant value of criterion c (c = 1, 2,..,C) of agent h, $\overline{H}_{g,c}$ is the mean value of criterion c of the group g, w_c is the weight coefficient of the criteria explaining the discriminant of agent groups. The default value of w_c is $1/C$.

Second, household agent h assigns himself into the most similar/nearest group:

$$\text{Set } H_g = g^* \text{ with } g^* = \arg\min\{D_{h1}, D_{h2}, D_{h3}, \ldots, D_{hK}\} \tag{3.23}$$

where D_{h1}, D_{h2}, D_{h3}, …, D_{hK} are distances from household agent h to all groups 1, 2, …, K; and g^* is the most similar (nearest) group to household agent h.

Third, once the group identification H_g of agent h has changed, the agent will be asked to delete the old behavior parameter list and to adopt the new behavior template of the new group. When adopting a new behavior template, there are changes in not only parameter values, but also possibly the behavior structure: some decision variables and production components are added or deleted. In this way, it is reasonable to say that the household agent already "*adopted*" the strategy and behavior of the closest advantage group.

Population

The *Population* class is a collection of all household agents. Additionally, the class is equipped with a routine $F_{Pop\text{-}stat}$ for computing descriptive statistic parameters of the whole population (equation 3.24). By comparing the simulated temporal population data sets, emergent phenomena of particular socio-economic patterns may be discovered and observed.

$$Population = \{\{Agents\}, F_{pop\text{-}stat}\} \tag{3.24}$$

where $\{Agents\}$ represents the simulated data panel of all household agents within the system at a time step, where each instant household agent listed by row and the agent's attributes are listed by columns, and $F_{pop\text{-}stat}$ is a procedure for computing basic descriptive statistics of the population evolving over time. The common need is that people want to see the changes in the community's socio-economic structures associated with land-use/cover changes. Thus, we designed the $F_{pop\text{-}stat}$ procedure for calculating the following socio-economic performance indicators of the community: i) overall average of annual income per capita, and ii) income distribution and equity level. Moreover, given the simulated data panel over time, principally, statistic patterns of other performance indicators can be calculated and plotted as needed.

The equity level of income is visualized by plotting the *Lorenz* curve, and is indicated through the *Gini* coefficient (see Figure 3.4). The Lorenz curve of personal income is a cumulative frequency curve showing the distribution of a population against income. If the distribution of the income is equal, the plot will be shown as a straight diagonal (45^0). Unequal distributions will yield a curve (see Figure 3.4). The gap between this curve and the diagonal is the inequality gap (fraction A in Figure 3.4). The *Gini* coefficient measures the degree of inequality in the income frequency distribution, and is calculated as the ratio of the area between the diagonal and the *Lorenz* curve (fraction A) to the total area beneath the diagonal (fraction A+B) (Rodrigue, 2003). The model used the algorithm for reporting the Gini coefficient and plotting the Lorenz curve from the Wealth Distribution model of Wilensky (1998), written in *NetLogo*.

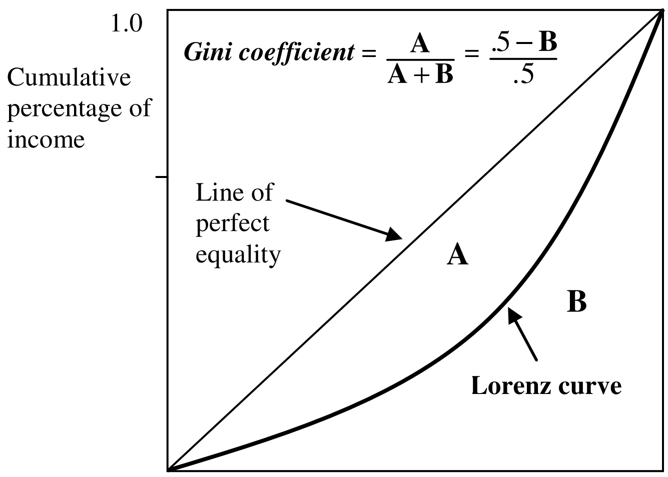


Figure 3.4 Lorenz curve and *Gini* coefficient of household income. Source: adapted from Rodrigue (2003)

3.2.2 System of landscape environment: The PATCH-LANDSCAPE module

Module overview

This module presents the landscape environment following the agent-based design: each landscape unit is treated as an autonomous agent. The landscape system is also represented in the form of a hierarchy of spatial scales referring to three levels of organisation: *landscape agent*, *landscape vision,* and *entire landscape* (see Figure 3.5).

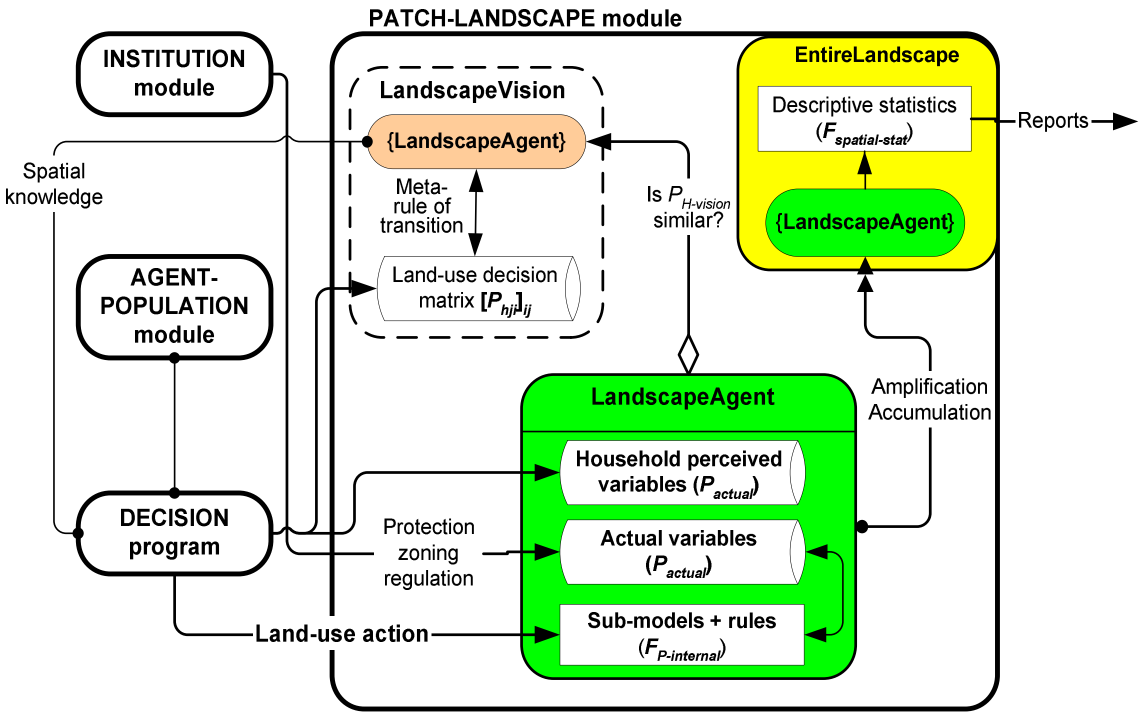


Figure 3.5 Block diagram of PATCH-LANDSCAPE module

Landscape agent (named *LandscapeAgent*) represents congruent patches (i.e., 30 m x 30 m), consisting of two main components: the patch's internal state variables and internal ecological sub-models. State variables correspond to GIS-raster layers, consisting not only of biophysical variables, but also of socio-economic properties of the land. Ecological sub-models built into landscape agents perform ecological processes hosted by the landscape unit, in ways beyond the household agent's control and responding to event-based interventions by household agents. The *AgriculturalYieldDynamics* is a patch sub-model to anticipate the agricultural yields of the patch in response to its natural characteristics, household inputs and history of land use. The *ForestYieldDynamics* is another patch sub-model for calculating the yield (stand basal area) of the forest growing on the patch, in response to the physiologic condition of the patch (i.e., previous stand basal area) and the harvesting activities of human agents (i.e., selective logging). The *NaturalTransition* is a patch sub-model that transforms small changes in vegetation (i.e., cover modification by humans or annual growth) into categorical changes of vegetation covers (i.e., cover conversion).

Landscape vision (named *LandscapeVision*) is a collection of landscape agents within the vision (i.e., the "*sphere of influence*" – see chapter 2, Figure 2.1) of a household agent. Given a household agent holds a number of patches and this vision is expressed as specific radiuses from the holdings, then his *LandscapeVision* is a collection of many circular neighbourhoods generated around his holding patches. All spatial information within *LandscapeVision* is reflected exactly into the household agent's spatial knowledge that has been mentioned in agent structure. The patch within *LandscapeVision* has added some more attributes that are specific for the corresponding human agent, while patches outside do not have these attributes. Each household agent has his "*personal*" *LandscapeVision*, upon which he infers knowledge, makes land-use decisions, claims, and acts and creates impacts. As household agents may change the positions of their holdings, their *LandscapeVision* will also move accordingly, and then the spatial information perceived by the human agent is renewed. At any time, household agents modify or convert the state of a patch, such impacts (can be small and insignificant at one time) are accumulated over time and aggregated over space, gradually resulting in spatio-temporal dynamics of the overall landscape.

The entire landscape (named *EntireLandscape*) is a collection of all individual patches, thus overall patterns of the entire landscape are the result of the aggregated and accumulated impacts. As the spatial system is self-organised, emergence properties may be observed. Simple statistic programming procedures are used for calculating some socio-economic indicators of population dynamics.

Structure of landscape agent

The structure of a landscape agent is formally expressed as follows:

$$LandscapeAgent = \{P_{actual}, P_{H\text{-}specific}, F_{P\text{-}internal}\} \tag{3.25}$$

where P_{actual} is a set of *actual* variables of the patch that are dependent of the household agent, $P_{H\text{-}specific}$ is a set of patch variables, where the values are specifically given by a human agent and have meanings with respect to that household agent only, and $F_{P\text{-}internal}$ is a set of internal ecological sub-models that perform patch dynamics (see Figure 3.5). The elaborations of these components are given below.

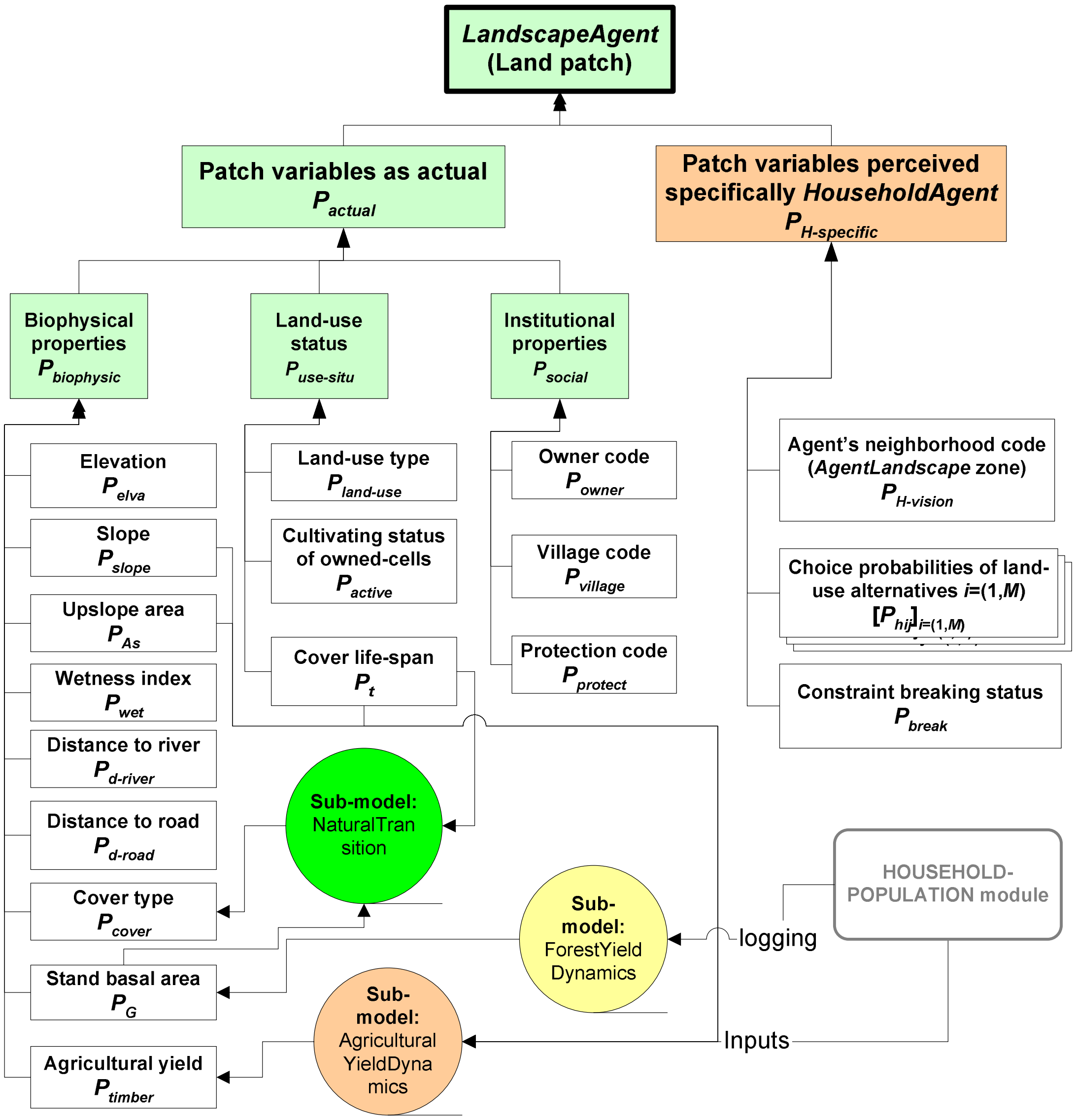


Figure 3.6 Ecological variables and sub-models built into landscape agents

Actual variables of landscape agent (P_{actual})

The set of actual variables of a patch (P_{actual}) consists of three components: biophysical conditions ($P_{biophysic}$), land-use status (P_{use}), and institutional conditions (P_{social}):

$$P_{actual} = \{P_{biophysic}, P_{use\text{-}situ}, P_{institution}\} \tag{3.26}$$

Biophysical conditions ($P_{biophysic}$) include the following variables:

$$P_{biophysic} = \{P_{elev}, P_{slope}, P_{As}, P_{wet}, P_{d\text{-}river}, P_{d\text{-}road}, P_{cover}, P_{a\text{-}yield}, P_G\} \tag{3.27}$$

where P_{elva}, P_{slope}, P_{As}, P_{wet} are elevation, slope angle, upslope contributing area and topographic wetness index, respectively. $P_{d\text{-}river}$ and $P_{d\text{-}road}$ are approximate distances from the patch to the nearest river/stream and road, respectively. P_{cover} is the code of cover type of the patch. $P_{a\text{-}yield}$ is agricultural yield of the current agricultural land-use type of the patch, and and P_G is forest stand basal area of the patch. The ecological meanings of upslope contributing area (P_{As}), topographical wetness (P_{wet}), and basal area of forest stand (P_G) are given in Chapter 5 (Sections 5.3.1 and 5.3.3).

The land-use situation ($P_{use\text{-}situ}$) is expressed by the following variables:

$$P_{use\text{-}situ} = \{P_{landuse}, P_{active}, P_t\} \tag{3.28}$$

where $P_{land\text{-}use}$ is the code of the land-use type currently applied to the patch. P_{active} is a dummy (Boolean) code that shows whether the patch is being used (P_{active} = 1) or fallowed/abandoned (P_{active} = 0). P_t is the life-span of the existing cover type of the patch (in case cover type is crop land, then P_t is also known as cropping time-length).

Institutional properties of a patch ($P_{institution}$) include the following variables:

$$P_{institution} = \{P_{owner}, P_{village}, P_{zoning}, P_{protect}\} \tag{3.29}$$

where P_{owner} indicates the owner of the patch (i.e., $P_{owner} = {}^{h}H_{id}$ if the patch is a holding, $P_{owner} = 0$ if otherwise). $P_{village}$ is a village code of the patch (i.e., $P_{village}$ = [village_code] if the patch is located within a village territory. $P_{village}$ = 0 if otherwise), P_{zoning} is zoning index (see Section 3.2.4). and $P_{protect}$ is protection code of the patch (i.e., $P_{protect}$ = 1 if the patch is located within the protection zone, $P_{protect}$ = 0 if otherwise). The institutional patch variables P_{owner}, $P_{village}$ and $P_{protect}$ help to create the *tenure relationship* between landscape agents and household agents.

Household-specific variables of landscape agent ($P_{H\text{-}specific}$)

As mentioned above, $P_{H\text{-}specific}$ is the set of patch variables, where the values are given specifically by a household agent, and which have meanings with respect to that house agent only:

$$LandscapeAgent\ j \xleftrightarrow{P_{H-specific}} HouseholdAgent\ h$$

$P_{H\text{-}specific}$ includes four components: vision code ($P_{H\text{-}vison}$), set of variables storing land-use choice probabilities anticipated by the household ($[P_{hij}]_{i\in M}$), and constraint-breaking status (P_{break}):

$$P_{H\text{-}specific} = \{P_{H\text{-}vision}, [P_{hij}]_{i\in M}, P_{break}\} \tag{3.30}$$

$P_{H\text{-}vision}$ is a categorical patch variable indicating whether the patch is within the vision R_h of the household agent h: $P_{H\text{-}vision} = {}^hH_{id}$ if the patch is within the vision of the household agent h^{th} (R_h), and $P_{H\text{-}vision}$ = 0 if otherwise (default).

$$P_{H\text{-}vision} = {}^hH_{id}\ \ (default = 0) \xleftarrow{\text{VisionR}_\text{h}} HouseholdAgent\ h \tag{3.31}$$

$[P_{hij}]_{i\in M}$ is a set of variables storing choice probabilities of possible land use type i (i = 1, 2, …, M), which are anticipated by household h at location j. The values of $[P_{hij}]_{i\in M}$ are generated by the *FarmlandChoice* routine in the DECISION program of household agent h, which is explained in detail in Section 3.2.3.

$$[P_{hij}]_{i\in M}\ \ (default = 0) \xleftarrow{\text{FarmlandChoice}} HouseholdAgent\ h \tag{3.32}$$

P_{break} is a patch variable indicating whether the agents breaks particular constraints (P_{break}=1), or accepts the constraint (P_{break}= 0 as default). If P_{break}= 1, the patch will be considered as other normal patches when optimising land-use choice, otherwise the patch will be excluded from land-use optimisation space. The default value of P_{break} is 1. The value of P_{break} is generated by the *ConstraintBreaking* routine in the DECISION program of household agent *h*; which is explained in detail in Section 3.2.3.

$$P_{break} = 1 \text{ (default=0)} \xleftarrow{\text{ConstraintBreaking}} HouseholdAgent\ h \qquad (3.33)$$

Internal ecological sub-model of landscape agent ($F_{P\text{-}internal}$)

As mentioned in the module overview, there are three ecological sub-models built into landscape agents: agricultural yield dynamics (named *AgriculturalYieldDynamics*), forest yield dynamics (named *ForestYieldDynamics*), and natural transition of vegetation cover (named *NaturalTransition*).

$$F_{P\text{-}internal} = \{AgriculturalYieldDynamics,\ ForestYieldDynamics,\ NaturalTransition\} \qquad (3.34)$$

- *AgriculturalYieldDynamics sub-model:*

The *AgriculturalYieldDynamics* is a patch's sub-model for performing the dynamics of variable $P_{a\text{-}yield}$ in response to variations in natural conditions, time and management practice. Careful justifications for selecting a relevant modeling approach, predictors of agricultural yield, and function forms of this sub-model are given in chapter 5 (Section 5.3.2). Here, we just summarize the sub-model in connection with other modules of VN-LUDAS.

The mathematical expression of the sub-model is:

$$^{i}P_{a-yield} = a_i . I_{chem}^{\beta_{i1}} . I_{labor}^{\beta_{i2}} . P_{slope}^{\beta_{i3}} . P_{As}^{\beta_{i4}} . P_{t}^{\beta_{i5}} \qquad (3.35)$$

where *i* indexes the current agricultural land-use type of the considered patch, a_i is a constant; β_1, β_{i2}, β_{i3}, β_{i4}, and β_{i5} are yield elasticity coefficients to agrochemical input

(I_{chem}), labor input (I_{labor}), patch slope (P_{slope}), patch upslope contributing area (P_{As}) and cropping time (P_t).

In equation 3.35, agrochemical input (I_{chem}) and labor input (I_{labor}) are neither landscape agent's variables nor household agent's variables: they are temporal variables created by the DECISION program. Thus, this patch sub-model is connected to the DECISION program of household agents and reflects the diversity of household agents. We will see later that in the DECISION program, the decision on agrochemical inputs for agricultural production is influenced by subsidy policies. Thus, the sub-model is also connected with policy factors.

Patch slope (P_{slope}) and upslope contributing area (P_{As}) represent natural potential and risk of the location for agricultural production, and they are often highly variable across space. Thus, the agricultural yield model can capture the spatial heterogeneity of the landscape. The inclusion of cropping time (i.e., life-span of the current agricultural land-use type of the patch) means that the land-use history is taken into account by the *AgriculturalYieldDynamics* sub-model.

The empirical coefficients a and β in Equation 3.35 are given by quantitative case studies conducted outside the VN-LUDAS model. In the application of the VN-LUDAS for the Hong Ha study site, empirical estimations for coefficients of the agricultural functions are given in chapter 5 (Section 5.4.2).

- *ForestYieldDynamics sub-model*

The *ForestYieldDynamics* sub-model is a patch sub-model for performing the dynamics of the forest stand basal area (P_G) in response to the vegetative condition of the site (i.e., previous stand basal area) and human disturbance (i.e., logging activities). Differing from the agricultural yield sub-model that is based a single function, *ForestYieldDynamics* is a dynamics model, which is based on a set of equations and a computation algorithm. The mathematic development of these equations and the related computational algorithm are described in detail in chapter 5 (Section 5.3.3 and Figure 5.4). The summary of this sub-model is given as follows.

Based on the concept of forest yield and growth, the forest stand basal area at time point t (${}^tP_G$) is expressed as follows:

$$^{t}P_G = (^{t-1}P_G + ^{t-1}Z_G) - G_{removals} \tag{3.36}$$

where $^{t-1}P_G$ is stand basal area in the previous year (at the first round of the computing loop, $^{t-1}P_G$ is exactly the initial stand basal area that we need to set as required), $^{t-1}Z_G$ is the natural increment of stand basal area in the previous year, and $G_{removals}$ is the basal area logged by household agents. Notice that in forestry, the stand basal area is used for indicating both timber yield and wood density of the site. The component $G_{removals}$ is event-driven, its existence depends on the decision-making of household agents.

Vanclay (1994: 109), based on Von Bertalanffy's equation, developed a theoretical equation expressing the basal area growth of a forest stand as the whole:

$$Z_G = dP_G/dt = a(P_G)^{\varepsilon} - b(P_G) \tag{3.37}$$

where a and b are constants, ε is a very small constant ($\varepsilon \rightarrow 0$). However, Vanclay did not say how to specify the parameters a and b.

Through some simple mathematic development, based on a few acceptable assumptions, the equations for determining a and b are formulated as follows (see Section 5.3.3 for more details):

$$a = {}^{max}Z_G / [(^{equil}P_G)^{\varepsilon} (\varepsilon^{\varepsilon/(1-\varepsilon)} - \varepsilon^{1/(1-\varepsilon)})] \tag{3.38}$$

$$b = {}^{max}Z_G / [^{equil}P_G (\varepsilon^{\varepsilon/(1-\varepsilon)} - \varepsilon^{1/(1-\varepsilon)})] \tag{3.39}$$

where $^{max}Z_G$ is the maximal growth rate of stand basal area, and $^{equil}P_G$ is the stand basal area at the equilibrium state of the forest stand (also called *natural basal area*). The values $^{max}Z_G$ and $^{equil}P_G$ are often available in forest science literature, or possibly estimated by forestry experts. The constant ε can be set by a very small value (e.g., $\varepsilon = 10^{-6}$).

Assuming that the human impact on forest quality is mainly in terms of selective logging, the removed basal area ($G_{removals}$) principally includes three components: harvested amount (G_{logged}), logging damage (G_{damage}) and logging-driven mortality ($G_{mortality}$):

$$G_{removals} = G_{logged} + G_{damage} + G_{mortality}/T \tag{3.40}$$

where G_{logged} is the basal area logged by human agent(s), G_{damage} is standing basal area damaged immediately by logging operation, and $G_{mortality}$ is basal area lost retentively as tree mortality occurring over some years (T) after the logging event (see Alder, 2000).

G_{damage} and $G_{mortality}$ are calculated based on the empirical study of logging impacts by Alder and Silva (2000) in the Brazilian Amazon:

$$G_{damage} = {}^{t-1}P_{\mathrm{Gr}}\,(0.0052\ G_{logged}/g_{logged} + 0.0536) \tag{3.41}$$

$$G_{mortality} = {}^{t-1}P_{\mathrm{Gr}}\,(0.0058\ G_{logged}/g_{logged} + 0.0412) \tag{3.42}$$

where G_{logged} is the amount of basal area logged (m^2) (i.e., logging intensity), and g_{logged} is the mean basal area of logged trees (m^2).

- *NaturalTransition sub-model*

The *NaturalTransition* is a set of transition rules that performs the natural transitions among vegetative cover types. In general, the firing of these rules is based on the evaluation of the four patch variables: previous cover type (${}^{t-1}P_{cover}$), life-span of the existing cover type (P_t) (see Green, 1993; Quintero *et al*., 2004), existing stand basal area (P_{Gr}), and distance to the nearest natural forest ($P_{d\text{-}forest}$). As rule specifications need to deal with concrete land-cover types, no general specification is illustrated here. The specific algorithms and parameters of the *NaturalTransition* sub-model for the Hong Ha (Vietnam) case are described in chapter 5 (Section 5.3.4 and Figure 5.5).

Dynamic rules for other patch variables

The *cropping time* (P_t) has an increment of 1 year over time, but is reset to 0 if the patch is fallowed (when $P_{active} = 0$):

$${}^{t+1}P_t = \begin{cases} {}^{t}P_t + 1 & if\ {}^{t}P_{active} = 1 \\ 0 & if\ {}^{t}P_{active} = 0 \end{cases} \tag{3.43}$$

where the transition of P_{active} from fallowed (when P_{active}=0) to cropping (when P_{active}= 1) status or vice versa depends on the final decision-making of the household agent specified by the procedure *FarmlandChoice* in the DECISION module. The status of P_{active} is important for activating patches to be evaluated by the *FarmlandChoice* procedure (see equation 3.47).

LandscapeVision **and decision-making matrix**

The concept of *landscape vision* (*LandscapeVision*) here refers to a local and personal landscape space recognised by an agent. A *LandscapeVision* perceived by a household h is expressed as follows:

$$ {}^{h}LandscapeVison = \{\{LandscapeAgents \mid P_{H\text{-}vision} = {}^{h}H_{id}\}, [P_{hij}]_{ij}\} \qquad (3.44) $$

where $\{LandscapeAgents \mid P_{H\text{-}vision} = {}^{h}H_{id}\}$ is a collection of all landscape agents where $P_{H\text{-}vision}$ matches the identification of the household agent (${}^{h}H_{id}$), defining the spatial extent of the *LandscapeVision*; and $[P_{hij}]_{ij}$ is a matrix of anticipated choice probabilities for all possible land uses and for all locations within the landscape vision, creating a basis for the rational land-use choice of the household agent. Thus, $[P_{hij}]_{ij}$ is also called *decision-making matrix* (Dorigo *et al*., 1999), or *response surface* (April *et al*., 2003).

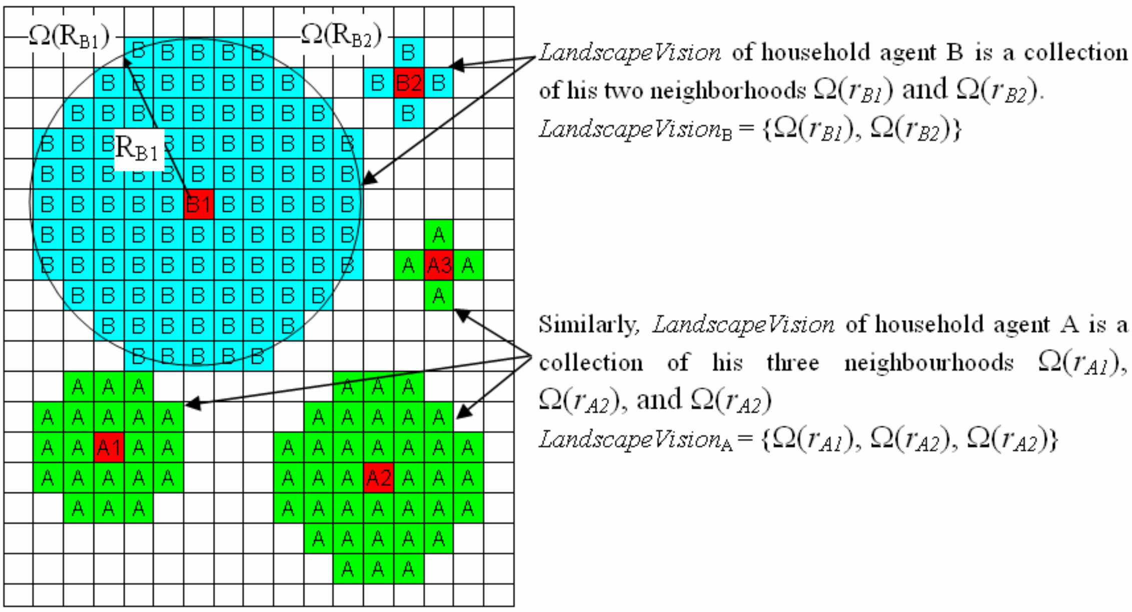


Figure 3.7 An example illustrates the spatial pattern of the *LandscapeVision* concept.

It is useful to clarify the spatial pattern of *LanscapeVision*. Let $\Omega(R_k)$ be the neighbourhood of holding patch k of household agent h, where R_k is the *vision radius*, which is an agent-specific parameter. As the farming household may have many scattered holding plots, the spatial pattern of *LandscapeVision* is normally expressed in the form of *aggregated disjoint neighbourhoods* (Figure 3.7). Every household agent has his "*private*" defined landscape. Household agents recognise spatial information, analyse trade-offs and optimise spatial land-use choices within their *LandscapeVision* only, not necessarily upon the entire landscape. After each time step, following the moving of household agents, creating and abandoning their holding patches, their *LandscapeVision*s will accordingly change.

Table 3.1 Land-use decision matrix ($[P_{hij}]_{ij}$) of household agent h is a combination of vectors of location options ($[P_{hij}]_{j\in N}$), and vectors of land-use alternatives $[P_{hij}]_{i\in M}$).

	Patch$_1$	…	*Patch$_i$*	…	*Patch$_N$*	*Vector of location options for a given land use*
Land use$_1$	P_{h11}	…	U_{h1j}	…	P_{h1N}	$[P_{h1j}]_{j\in N}$
…	…	…	…	…	…	…
Land use$_i$	P_{hi1}	…	P_{hij}	…	P_{hiN}	$[P_{hij}]_{j\in N}$
…	…	…	…	…	…	…
Land use$_M$	P_{hM1}	…	P_{hMj}	…	P_{hMN}	$[P_{hMj}]_{j\in N}$
Vector of land-use options for a given location	$[P_{h1j}]_{i\in M}$	…	$[P_{hij}]_{i\in M}$	…	$[P_{hNj}]_{i\in M}$	$\sum_{ij}P_{hij} = 1$

Note: h, i, and j index household agent, land-use type and location, respectively. M is the total number of possible land-use types. N is the total number of accessible patches within the household agent's vision.

LandscapeVision is a landscape area that household agents use as a basis for computing their land-use decision matrix[5] (see Figure 2.1, Figure 3.7 and Table 3.1). The land-use decision matrix ($[P_{hij}]_{ij}$) is a table containing land-use choice probabilities P_{hij} distributed over patches j (j = 1, 2, …, N) within the *LandscapeVision* and possible land-use types i (i = 1, 2, …, M) anticipated by the agent h (Table 3.1). The decision matrix will be used for directing the household agent's search towards the best pair

[5] A similar concept of decision matrix is "*ant decision table*" in Dorigo *et al.* (1999), or "*response surface*" in April *et al.* (2003).

comprising the best patch j^* and the best land-use type i^*, which potentially gives him the maximal utility. The land-use decision matrix is computed by the *DecisionMatrixCalculate* procedure specified in the DECISION module (see Section 3.2.3). By specifying the land-use decision matrix for every household agent, a *meta-rule of transition* has already been stated (see Wu, 1998).

Entire landscape (*EntireLandscape*)

The entire landscape (*EntireLandscape*) is a composition of all landscape agents in the system, and a routine computing descriptive statistics of particular indices of land-use and land-cover changes at landscape level:

$$EntireLandscape = \{\{LandscapeAgents\}, F_{spatial\text{-}stat}\} \tag{3.45}$$

where $\{LandscapeAgents\}$ is simulated spatial data for the whole landscape, which can be converted into popular GIS-raster formats for further spatial analysis; $F_{spatial\text{-}stat}$ is a procedure for computing basic descriptive statistics of land use/cover evolving over time, which are shown in the form of graphs on the user interface of the model for real-time observation. Moreover, given the simulated spatial data panel saved, further spatial analysis can be done as needed.

3.2.3 Structure of DECISION module (program)

Module overview

The DECISION program represents the mechanism of decision-making processes of household agents. From the agent-based architecture viewpoint, the DECISION program acts as an internal decision-making routine encapsulated into the "*body*" of household agent, together with agent-specific data, forming an exact model of himself as shown in Figures 3.1 and 3.2. The DECISION program is an *organized scheduling program* to perform the houshold agent's decision-making and action on land and forest use. As it is only a procedure, the module differs from the two previous ones because it does not store any data. However, the module plays an engine role through: i) *integrating different information flows* from other modules into the household agent's evaluation, ii) *scheduling* the household agent's decision-making processes and subsequent actions, thus iii) *governing behaviors* of household agents.

Starting the DECISION program, a *labor allocation list* divides the household labor pool into different labor budgets to allocate to main production lines, reflecting his production strategy according to the livelihood typology of the group that the household agent belongs to.

Accordingly, the land-use decision of household agents includes two parallel main processes: i) the *claims* and *uses of landholdings for cultivation* (formulised by the *FarmlandChoice* procedure), and ii) the collection of forest products on public or state lands (formulised by the *ForestChoice* procedure). The *FarmlandChoice* procedure generally follows the bounded-rational approach, based on the assumption that household agents tend to optimize their choice of location and associated agricultural land use within their capacity of labor and institutional constraint. The *FarmlandChoice* algorithm falls into the ordered choice method under the bounded rational paradigm as discussed in Chapter 2 (Section 2.3.2 and Figure 2.4c). However, during the decision process, many condition-action rules are used in association with bounded optimization processes.

The *ForestChoice* procedure follows the reflex behavior approach, based on the assumption that household agents do not, or cannot, do any rational thinking when collecting forest products, rather they react to the forest environment according to their daily routines/rules. As in the *FarmlandChoice*, the collection of forest products by household agents is constrained by labor availability and protection regulations. However, there is some chance for them to break the protection rules (i.e., institutional constraints), as introduced institutional rules are always imperfectly enforced.

Direct income from other production activities (affecting land use indirectly), such as livestock production and off-farm activities, are defined according to empirical patterns.

Production structure and labor allocation

A household agent's decision and actions are shaped by his *production strategy*[6], which relates to *production structure* and *resource allocation strategy*. Production structure is usually represented in the form of a nested hierarchy of production components/sub-components (see Figure 3.8 as an example). Resource allocation strategy here simply

[6] A similar term is *livelihood strategy.*

means how much resources of an household agent are likely to be distributed/allocated following the defined livelihood structure. The strategy can be approximately represented by relative magnitudes of resource flows to each main production line. For instance, a better-off household may plan to partition their labor pool following an agriculture-based production strategy: {*80% crop production, 10% livestock production, 5% forest collection, 5% others*}; another poor household may adopt a forest-based strategy: {*20% crop production, 0% livestock production, 70% forest collection, 10% others*}; while a medium one has a diversified strategy with more even labor distribution against components.

Allocated labor budgets are the constraints of the loops of farmland and forest choices. Given an allocated labor budget, the loop of farmland choices, for instance, repeats following the maximising utility principle until the labor budget is finished.

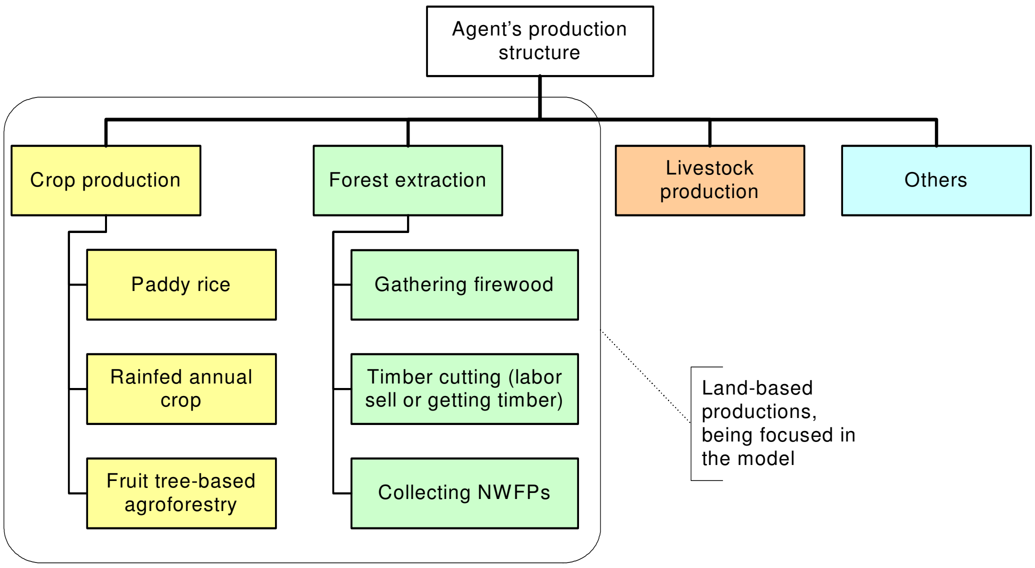


Figure 3.8 Typical production/livelihood hierarchy of villagers in Hong Ha commune, Thua Thien – Hue province, in the upland of the Central Coast of Vietnam

Given a defined production structure and strategy, the household agent will accordingly mobilise his *resources* to production activities. His resources can include labor pool (i.e., labor availability), land endowment (i.e., total landholdings), and economic reserves (i.e., consumable resources such as money, rice store, etc.). Land endowment has already been considered as an output of the decision process. In case of low input and subsidized agricultural systems in our study area, financial resources for agriculture are important, but likely an external resource rather than an internal one.

Thus, *labor pool* is here considered as the main resource of household agents and labor allocation flows reflect their livelihood strategy, specified by the vector $[H_{labor\text{-}strategy}]$ of percentage labor (Equation 3.18). In the Vietnam case study, the labor pool is partitioned into four labor budgets corresponding to four production components:

$$[H_{labor\text{-}strategy}] = [\%L_{crop}, \%L_{livestock}, \%L_{timber}, \%L_{others}] \qquad (3.46)$$

where $\%L_{crop}$, $\%L_{livestock}$, $\%L_{timber}$, $\%L_{others}$ are proportions of labor budget allocated to crop production, livestock production, timber logging, and other activities, respectively.

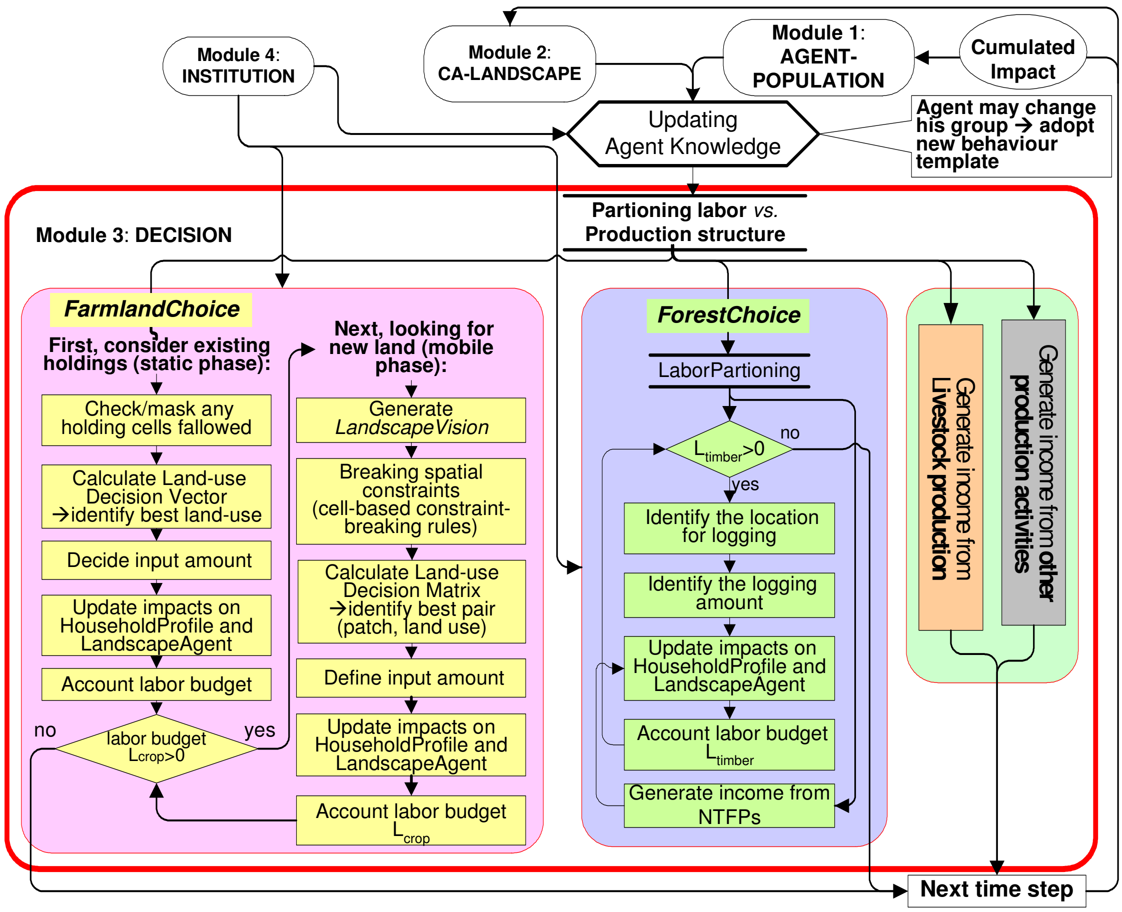


Figure 3.9 Flowchart of DECISION module shows parallel processes of decision-actions in different production lines

In the DECISION module, the processes of labor allocation include the two following steps (Figure 3.9):

i) Depending on group typology, the agent strategically *partitions labor pool into labor budgets* for each production line, based on the labor strategy list $H_{labor\text{-}strategy}$. It is noted that there may be a proportion of unused labor, which often happens in reality. As reasons accounting for unused labor are generally hard to define, it is difficult to build deterministically dynamic rules for the proportion of unused labor. We propose that this unused labor amount is predefined and then subtracted out of the strategic labor allocation.

ii) In each production line, *iterative processes* occur and are *constrained by the allocated labor budget*. Concrete production types will be chosen and carried out following the utility-maximising rule (with crop production line) or other reflex rules (with forest extraction line), until the labor budget is finished as the goal of household agent (see Figure 3.9).

The following part describes the decision-making and action processes of land-based livelihood components (i.e., agricultural land-use and forest extraction), because these processes directly lead to land-use and land-cover changes.

Agricultural land-use decision-making and actions (*FarmlandChoice* procedure)

Algorithm of agricultural land-use decision-making process

The household agent's decision-making and actions with respect to agricultural land-use here employ the bounded-rational decision making approach, represented by the *FarmlandChoice* procedure. The decision-making procedure consists of two separate phases: use of old landholding (called *static phase*) and use of new land (called *moving phase*).

Static phase: Decision-making on uses of existing landholding

To minimise costs and take advantages of his optimal location choices made in the previous time step, it is assumed that the household agent considers his available land holdings before deciding to search and open any new holding patches. Naturally, the household agent makes decision following the sequential steps:

i) *Decide to fallow existing landholding*: This decision is taken through executing a pseudo-random-proportional transition rule applying the patch attribute P_{active} (see equation 3.47). When fallowed patches are excluded from the decision space, the remaining holdings called *active holding patches*.

ii) For each active holding patch, *chose the most suitable land-use type* among available alternatives, based on his existing knowledge. This task includes the computing *land-use decision vector* $[P_{hij}]_i$, and the selection of utility-maximal land-use type $i*$ ($i* = arg\ max\ [P_{hij}]_i$). The land-use decision list is calculated using the *DecisionListCalculate* routine.

iii) *Decide input amount*, based on his capacity and external support (e.g., government subsidy for intensifying crop production). The step is done through firing a pseudo-random-proportional rule that incorporates factors of subsidy policy in determining how much fertilizer/pesticide are allocated to the selected production. Labor input is approximated at the average value of the group.

iv) *Account and check balance of labor budget*. The conclusions of the above decisions will determine how much of the labor budget would be allocated accordingly. The agent will stop allocating land if the balance of the labor budget is zero or becoming negative. Here, one of the following two cases can happen:

- If the labor budget is finished before (or barely finished when) all old landholdings are used, then stop land-use activities and wait for next time step. This happens when the household agent is rich in landholding pool, or has little labor, or has effective intensification of crop production.
- If the labor budget has not been finished yet when all old landholdings are used, then go to the moving phase. The positive balance of the labor budget gives the agent the possibility to open and use new land. This happens when the agent is rich in labor pool, or landless, or fallows certain amount of land.

The balance of the labor budget is a synthesis outcome resulting from the defined labor budget for crop production, land-use type and intensification choice, and a condition for performing the next decision-making phase.

Moving phase: Searching new land and deciding on use of new landholding

The decision-making phase is an intensive computational process, including spatial exploration, evaluation of agricultural land-use alternatives, checking physical and institutional constraints, bringing about location decisions, and making a final land-use decision. The main events included in the procedure are summarised as follows:

i) *Generate spatial knowledge*. The household agent generates his spatial knowledge through updating his past *LandscapeVision*, based on his vision R_h.

Looping over all patches on his *LandscapeVision*, the household agent performs the two following tasks:

ii) *Execute the ConstraintBreaking procedure:*

- *No constraint-breaking with physical constraints*: For which patches within *LandscapeVision* are in physical constraint zones, the household agent will certainly have to accept this physical constraint: $Prob(P_{break} = 1) = 0$.
- *Random-proportional rules for breaking institutional constraints*: $Prob(P_{break}= 1)=\theta$
    - *Breaking ownership/village constraint*: For that patch within *LandscapeVision* that has already been claimed by other agents or is in territory of another village, the agent will have the chance ($\theta_{own\text{-}break}$ and $\theta_{vill\text{-}break}$) to break this ownership/village constraint, i.e., considering the patch as normal. $\theta_{own\text{-}break}$ and $\theta_{vill\text{-}break}$ are thresholds in the ownership-breaking rule and village-breaking rule, respectively.
    - *Breaking protection zoning constraint:* For those patches both within the *LandscapeVision* and the watershed forest protection zone, the agent will also have a chance ($\theta_{protect\text{-}break}$) to include these

protected patches into his decision space. It is reasonable to approximate $\theta_{protect\text{-}break}$ at protection regulation enforcement.

iii) *Execute the DecisionMatrixCalculate to compute the land-use decision matrix*. The land-use choice probability P_{hij} integrates three different information flows: household attributes, environmental conditions within household's vision, and certain global factors; thus the function is used as a trade-off ranking index for spatial land-use choice. As a result of the calculation of the decision matrix $[P_{hij}]_{ij}$, the agent has generated his "*private*" patchular grids stack of distributed choice probabilities ($\Sigma_{ij}P_{hij}$=1) (see Table 3.4). Based on the land-use decision matrix, the agent selects the patch and coupled land-use that gives highest choice probability.

v) *Decide the input amount*: similar to the third step in the static phase.

iv) *Account and check balance of labor budget.* The loop of the moving phase will stop if the balance of labor budget is zero or becoming negative, otherwise it is repeated until the labor budget is finished.

<u>*Decision to fallow holding patches (ToFallow procedure)*</u>

We do not model the agent's annual decision of fallowing his holdings to be a deterministic function of land productivity, because generally this kind of decision is related to not only production responses, but also to many random events (e.g., marital, health problems, etc.) (see Dale *et al*., 1993) or the other implicit reasons (e.g., religion or traditional beliefs). Thus, the reasons why farmers fallow their land are difficult to model in a deterministic manner. However, it is reasonable to assume that there is an almost *linear increasing probability* of an agent fallowing his holdings, along with cultivating length (see Dale *et al*., 1993). We employ this principle to formulate a probabilistic function of an agent fallowing a patch.

Let T_{crop} be the total length of the cropping period for a particular land-use option *i*, then the probability of household agent *h* fallowing his holding patch *j* after T_{crop} cropping year is almost ≈ 1.0, and the linear increment each year in probability of fallowing the patch is $1/T_{crop}$. Thus, the probability of the agent *h* fallowing the patch *j* after P_t continuous cropping year(s) ($P_t < T_{crop}$) is: $P_t \times 1/T_{crop} = P_t/T_{crop}$. The decision rule to fallow the holding patch in proportion to this fallow probability can be formalised by the following random-proportional rule:

$$^{t+1}P_{active} = \begin{cases} 0\,(fallowed) & if\ q \le {}^{t}P_t / T_{crop} \\ 1(cropping) & if\ q > {}^{t}P_t / T_{crop} \end{cases} \tag{3.47}$$

where q is a random number floating evenly over [0,1], ${}^{t}P_t$ is *current* life-span of the existing agricultural types of the patch j (i.e., continuous time length of the patch state P_{active} = 1). T_{crop} is the length of the whole cropping period of the land-use i experienced in reality, which can be easily approximated for every land-use type by local experts or averaging from a household group data set. The determination of T_{crop} can be conditioned by nesting the land-use type with a few other site factors (e.g., slope class).

Detecting spatial constraints and constraint-breaking rules (ConstraintBreaking procedure)

A constraint is here defined as a variable that serves to *limit or exclude entirely* alternatives under consideration (Eastman, 2001). Spatial constraint is here expressed in the form of a binary (Boolean/dummy) surface, as in many other studies, (Eastman, 2001). Two main types of spatial constraint are considered: i) physical constraints, and ii) institutional constraints. The physical constraint or cultivation occurs with patches located either in non-cultivatable surfaces, i.e., water/road/residential/rocks or very steep land (P_{slope} ≥ 35%). Institutional constraints occur with patches located in protection zones (i.e., if $P_{protect}$ = 1), called *protection-zoning constraint*, or in territories owned by other household agents (i.e., if $P_{owner} \neq {}^{h}H_{id}$ and $P_{owner} \neq 0$) (called *ownership constraint*), or in territories of another village (i.e., if $P_{villager} \neq {}^{h}H_{id}$ and $P_{village} \neq 0$) (called *village constraint*).

In land-use decision-making, it is important to formulise the decision of the household agent who *accepts* or *breaks* the constraints he faces at a given patch. Constraint acceptance (set P_{break}= 0) by the household agent for a given patch means the household decides to exclude the patch of constraint from his land-use evaluation because of no feasibility for use. Constraint acceptance often occurs with physical constraints. Constraint breaking (P_{break}= 1) by the household agent for a given patch means the agent decides to include the patch of constraint into his land-use evaluation if there is some possibility to use the patch. Constraint breaking often occurs with spatial institutional constraints, called *rule breaking*. The *ConstraintBreaking* procedure is a set of three rules as follows:

Patch-based rule for detecting and breaking physical constraints:

$$P_{break} = \begin{cases} 0 & if\ P_{cover} = \{"road","river"\} \quad or\ P_{slope} > 30^0 \\ 1 & otherwise \end{cases} \tag{3.48}$$

Patch-based rule for detecting and breaking ownership constraint:

$$P_{break} = \begin{cases} 1 & if\ ((P_{owner} \neq H_{id} \quad and \quad q \leq \theta_{own-break}) \quad or \quad P_{owner} = 0) \\ 0 & otherwise \end{cases} \tag{3.49}$$

where $\theta_{own\text{-}break}$ is the possibility of the agent getting landholding from other agents. The nature of $\theta_{own\text{-}break}$ is the success chance of the agent in the negotiation process with the owner of the patch to get the patch. As the land-market is not considered at this stage of the model, we let the $\theta_{own\text{-}break}$ be defined by model users. An extended procedure to include the land-market mechanism will be done in later model development.

Patch-based rule for detecting and breaking village constraint:

$$P_{break} = \begin{cases} 1 & if\ ((P_{village} \neq H_{village} \quad and \quad q \leq \theta_{vill-break}) \quad or \quad P_{village} = 0) \\ 0 & otherwise \end{cases} \tag{3.50}$$

where $\theta_{vill\text{-}break}$ is the possibility of the agent acquiring land in other villages. This parameter is defined by local experts or estimated from plot-based survey.

Patch-based rule for detecting and breaking protection-zoning constraint:

$$P_{break} = \begin{cases} 1 & if\ ((P_{protect} = 1 \quad and \quad q \geq \theta_{enforce}) \quad or \quad P_{protect} = 0) \\ 0 & otherwise \end{cases} \tag{3.51}$$

where $\theta_{enforce}$ is the probability of expelling household agents from the watershed forest protection zone, reasonably approximated by the enforcement level of the protection policy (see Figuge 3.10). Thus, the chance of the agent breaking the protection constraint is 1-$\theta_{enforce}$. If the protection regulation is perfectly (and ideally) enforced, $\theta_{enforce}$ is 1 (i.e., 1-$\theta_{enforce}$ = 0) agents have no chance in accessing the patch; otherwise, an agent will have a small chance (1-$\theta_{enforce}$ > 0) to include the protected patch in his land-use decision.

After executing all the constraint-breaking rules above for all patches within the *LandscapeVision*, those patches that still suffer constraints ($P_{break} = 0$) are not considered in the land-use evaluation in the later step.

In equation 3.51, the protection code ($P_{protect}$) is controlled by the protection-zoning rule in the GLOBAL-POLICY module (see Section 3.2.4). The response of this variable to the protection zoning factor is expressed as follows:

$$P_{protect} = \begin{cases} 1 & if \quad P_{zoning} > \theta_{protect} \\ 0 & otherwise \end{cases} \tag{5.52}$$

where $\theta_{protect}$ is the threshold of zoning index P_{zoning} for defining the protection zone (see Figure 3.10).

Individual evaluation of land-use choice probability (P_{hij})

The household's evaluation of choice probabilities for possible agricultural land use follows the *ordered choice* methods within bounded rationality approach (Chapter 2, Section 2.3.2 and Figure 2.4c). The basis of the method is the calculation of the choice probability for an agent *h* selecting a patch *j* from his considered patch set *N* and a land-use type *i* from *M* available land-use alternatives (P_{hij}) using a M-logit form. Depending on the context of the decision-making process, the probability P_{hij} can be distributed in either a one-dimensional space of land-use alternatives only ($[P_{hij}]_{i\in M}$), or a two-dimensional space of both land-use and site alternatives ($[P_{hij}]_{ij\in MN}$). The former case happens in the static phase of *FarmlandChoice* (when the site has already been defined/fixed), and the decision vector $[P_{hij}]_{i\in M}$ represents trade-offs among land-use alternatives. The later case takes place in the moving phase of *FarmlandChoice*, and the decision matrix $[P_{hij}]_{ij\in MN}$ represents spatial trade-offs among location and land-use alternatives.

- *Calculation of land-use decision vector* $[P_{hij}]_{i\in M}$ *for a given holding patch (DecisionListCalculate procedure)*

By definition, the deterministic component of the utility function (V_{hij}) is both site- and household-specific. The deterministic utility of a particular landuse type can be a linear function of agent and patch attributes ($X_{a\in A}$), which are assumed to be important for the household's decisions:

$$V_{hij} = \sum_{a}^{A} \beta_a X_a = [\beta_a]_a \times [X_a]_a \tag{3.53}$$

where V_{hij} is a deterministic component of the utility function inferred by household agent h at patch j for land-use option i, a indexes number of decision factors X_a, β_a is *preference coefficient* of X_a perceived by household agent h.

Given that locations are fixed, at a given patch j, household agent h faces M land-use options. If the stochastic component of the utility function is assumed to follow the Gumbel distribution, the conditional probability that the household agent selects one specific land-use type i of an available set M can be mathematically stated in a *multi-nominal logistic* form:

$$P_{hij}(choice \quad i) = P(V_i \geq \max[V_k]_{k\in M, k\neq i}) = \frac{\exp(V_{hij})}{\sum_{i=1}^{M}\exp(V_{hij})}, \text{ with } \sum_{i=1}^{M} P_{hij} = 1 \tag{3.54}$$

where V_{hij} is the deterministic utility of land-use type i at the given patch j, perceived by household h. M is the total number of available land-use alternatives. The computation of land-use decision vector $[P_{hij}]_{i\in M}$ at patch j is done by the *DecisionListCalculate* procedure.

Given the calculated vector of choice probability $[P_{hij}]_{i\in M}$, the household agent will decide to finally choose a land use among an alternative set M following the ordered choice procedure. The procedure consists of the following steps:

i) Rank the vector $[P_{hij}]_{i\in M}$ in descending order, according to P_{hij}, then the household has the new vector, called $[P_{hij}]_{i\in M,ordered}$.

ii) Attend to the new vector $[P_{hij}]_{i\in M,ordered}$ and try the first choice probability $P_{hij}\big|_{i*}$ in the new vector (i.e., the highest in $[P_{hij}]_{i\in M,ordered}$) using the *random-proportion rule* (see chapter 2, equation 2.6):

$$Acceptance\text{-}of\text{-}land\text{-}use\text{-}i* = \begin{cases} true & if\ q \le P_{hij}\big|_{i*} \\ false & otherwise \end{cases} \tag{3.55}$$

where q is a random number floating evenly over [0,1].

iii) If the tried $P_{hij}\big|_{i*}$ is successful (resulting "*true*"), then the household finally choose land use $i*$ and quit the calculating loop, otherwise it will return to the vector $[P_{hij}]_{i\in M,ordered}$ to pick the second alternative and repeat equation 3.55.

This bounded optimisation allows the household agent some chance to select a land-use type that may not be the best between the set of alternatives. But the chance for success with the best land use is highest (see also Section 2.3.2).

- *Calculation of land-use decision matrix $[P_{hij}]_{ij\in MN}$ over a LandscapeVision (DecisionMatrixCalculate procedure):*

If household agent h is looking for both location $j^*\in N$ and the coupled land-use $i^*\in M$ that potentially provide him with the maximum utility within his *LandscapeVision* (conditioned by $P_{break} = 1$), he now faces $M\times N$ options of pairs ij ($i\in M$ and $j\in N$). The P_{hij} is now distributed over a 2-dimensional space, forming a virtual response surface for land-use decision (i.e., land-use decision matrix) that takes into account trade-offs not only among available land-use alternatives, but also among accessible patches (with $P_{break}= 1$) within his *LandscapeVision*. If the residue is assumed to follow an extreme value distribution (Gumbel), the conditional probability of agent h selecting a pair ij from the available set $M\times N$ can be also expressed in an extended multi-nominal logistic form:

$$P_{hij}(choice\ \ ij) = P(V_{ij} \ge \max[V_{kl}]_{k\in M, l\in N}) = \frac{\exp(V_{hij})}{\sum_{j=1}^{N}\sum_{i=1}^{M}\exp(V_{uij})}, \text{ with } \sum_{j}^{N}\sum_{i}^{M} P_{hij} = 1 \tag{3.56}$$

where $k \neq i$, $l \neq j$, and N is number of accessible patches (having $P_{break} = 1$) within the *LandscapeVision* of household agent *h.* The computation of the land-use decision matrix $[P_{hij}]_{ij \in MN}$ is done by the *DecisionMatrixCalculate* procedure.

Similar to the static phase, in this moving phase the household agent also decides to finally choose a pair of a land use type (among the land-use alternative set *M*) and a location (among the possible patch set *N*) following the ordered choice procedure:

i) Rank the vector $[P_{hij}]_{i \in M, j \in N}$ in *descending order* according to P_{hij}, then the household has the new vector, named $[P_{hij}]_{i \in M, j \in N, ordered}$.

ii) Attend to the new vector $[P_{hij}]_{i \in M, j \in N, \text{ordered}}$ and try the first best pair $P_{hij}\big|_{i^* j^*}$ in the new vector (i.e., the highest pair in $[P_{hij}]_{i \in M, j \in N, ordered}$) using the *random-proportion rule* (see equation 2.6):

$$\textit{Acceptance-of-}\text{land-use-}i^* = \begin{cases} \textit{true} & \textit{if } q \leq P_{hij}\big|_{i^* j^*} \\ \textit{false} & \textit{otherwise} \end{cases} \qquad (3.57)$$

where q is a random number floating evenly over [0,1].

iii) If the try $P_{hij}\big|_{i^* j^*}$ is successful (resulting "*true*"), then the household agent finally choose pair i^*j^* (land-use i^* and patch j^*) and quit the calculating loop, otherwise the household agent will return to the vector $[P_{hij}]_{i \in M, j \in N, ordered}$ to pick the second alternative and repeat equation 3.57.

Because it is impossible to statistically estimate the preference coefficients β_a (in equation 3.53) for every individual household agent, we approximate β_a based on the fixed preference coefficient of the household agent group using the bounded-random rule:

$$\beta_a = \beta_{ga} - \sigma_{ga} + random(2\sigma_{ga}) \quad \text{with } a \in A,\ g \in K \qquad (3.58)$$

where β_{ga} is the preference coefficient of decision variable X_a of household agent group g. σ_{ga} is standard error of β_{ga} (linked to Equation 3.17). All β_{ga} and σ_{ga} values can be statistically calibrated using the plot-based dataset of every household agent group (see

Chapter 4). Notice that structures of decision variable vector $[X_a]_{a\in A}$ and preference vector $[\beta_a]_{a\in A}$ are specific for every household group.

Decision of input amount

Labor input is bounded-random around the average labor input of the selected land use for the whole population:

$$IN_{labor(i)} = mean(IN_{labor(i)}) - \sigma_{labor} + random(2\sigma_{labor}) \quad (3.59)$$

where $IN_{labor(i)}$ is labor input for land use i^{th}, σ_{labor} is standard error of the mean of $IN_{labor(i)}$ over all agents in the group ($mean(IN_{labor(i)})$). The parameter $mean(IN_{labor(i)})$ and $\sigma_{l\text{-}input}$ are estimated from the empirical plot-based dataset.

Cost of used agrochemicals, i.e., fertilizer and/or pesticide ($IN_{chem(i)}$), takes the average value of the group ($mean(IN_{chem(i)})$) with a randomness within standard error bounds ($mean(IN_{chem(i)}) \pm \sigma_{chem(i)}$) if the household agent does not receive any agrochemical subsidy; otherwise such agrochemical input is added with a subsidy amount ($W_{subsidy}$).

$$IN_{chem(i)} = \begin{cases} mean(IN_{chem(i)}) - \sigma_{chem} + random(2\sigma_{chem(i)}) + W_{subsidy} & if\ q \le \theta_{\%subsidy} \\ mean(IN_{chem(i)}) - \sigma_{chem} + random(2\sigma_{chem(i)}) & otherwise \end{cases} \quad (3.60)$$

where i indexes land-use type, and q is a random number distributed evenly over [0,1]. $\theta_{\%subsidy}$ is a tuneable point within [0,1], representing the chance for an household agent to access to agrochemical subsidy program(s). $W_{subsidy}$ is annual average subsidy amount the household agent received.

Decision-making on forest activities (*ForestChoice* procedure)

The procedure *ForestUseChoices*, a *reflex* rule-based procedure, is a set of household-specific rules determining what, where and how much timber is logged. Logging activities are important to model because of directly causing changes in forest covers. The collection of non-timber forest products (NTFPs) is considered in the term of income generation in proportion to the labor amount allocated to this activity only, without any spatially explicit decision-making mechanism, because the activity does not cause considerable changes in forest covers. In the reflex decision-making mechanism, household agents recognise the spatial environment and respond directly in spontaneous reactions, rather than act based on an intensive computational optimisation. As reflex rules are often local specific, the following rules are relevant to the human population in our study site only.

The *ForestChoice* procedure is an iterative loop of reflex rule-based processes constrained by allocated labor budget. The iterative loop consists of three following steps (also see Figure 3.9).

Search locations for logging

Househould agent h will perform in turn the following sub-steps. First, the household identifies the set of available patches for logging (DF). The patch set (DF) is a collection of all patches covered by dense forests. Only patches of the set DF have trees with adequate size for logging.

$$DF = \{LandscapeAgents \mid P_{cover} = \text{“dense\_forest”}\} \tag{3.61}$$

Second, within the patch set DF, the household agent will identify patch j^* that is the nearest to the house of household agent h:

$$patch\ j^*|_h = arg\ min\{{}^hP_{d\text{-}house(j)} \mid j \in DF\} \tag{3.62}$$

where j indexes patches with the patch set DF. ${}^hP_{d\text{-}house(j)}$ is the distance from a patch (in the DF set) to the house of household agent h. *Patch* $j^*|_h$ is the patch of being nearest to the house of household agent h. The basis assumption of this location searching rule is that the household agent prefers to log timber on the closest forested patch to minimize the transaction cost.

Third, if patch j^* is within a forest protection zone (delineated by protection zoning regulation in GLOBAL-POLICY module), household agent h has a chance of $(1\text{-}\theta_{enforce})$ to log trees growing on patch j^*. Thus, the chance for the household agent finally select patch j^* can be defined using the following random-proportional rule:

$$\textit{Select-patch-}j^* = \begin{cases} true & if\ q \leq (1-\theta_{enforce}) \\ false & if\quad otherwise \end{cases} \tag{3.63}$$

where q is a random number floating evenly over [0,1], and $\theta_{enforce}$ is the probability of expelling a household agent from the watershed forest protection zone, reasonably approximated by the enforcement level of the protection policy.

If patch j^* is finally selected, the household agent will take the next step, otherwise he will repeat the second sub-step (equation 3.62) to find another nearest patch, but the previous patch j^* is now excluded from the patch set DF.

Determine amount of timber for logging

Household agent h determines the amount of logged timber for patch j^* as follows:

$$G_{logged}\big|_{h,\,j^*} = n \times g_{logged} \tag{3.64}$$

where n is the number of trees in patch j^* logged by household agent h, and g_{logged} is the average basal area at the breast height of logged trees. Noting that the trees selected for logging normally have a relatively large size, thus the interger n can be one among a few small number with the patch size of 30 m × 30 m. g_{logged} can be easily estimated by interviewing local loggers. For example, assuming that g_{logged} is about 0.28 – 0.64 m^2 (i.e., a diameter at breast height (*dbh*) of about 60 – 90 cm), then it is reasonable to set n=1 for a patch of 900 m^2 covered by a dense natural forest with a stand basal area of about 30 - 35 m^2 ha^{-1}. The logging intensity for this case is about 3 – 7 m^2 ha^{-1}, which is a common range. If n is equal or more than 2 trees of such a size, the logging intensity now is too high (i.e., 6 – 14 m^2 ha^{-1}) and thus unrealistic for selective logging.

Generate income, modify vegetation status, and account labor budget

After logging tree(s), the household agent generates income from logged timber by multifplying the amount of the logged timber with the local price of timber. At the same time, patch j^* is asked for modifying its stand basal area (P_G) using the sub-model *ForestYieldDynamics* (see Section 3.2.2). The household agent also subtracts the labor amount spent for the logging activity from the labor budget L_{timber}. If the balance of the labor budget is zero or becoming negative, he will stop logging and quit the *ForestChoice* loop to go to the next time step, otherwise he will repeat the cycle to find another forested patch to log (see Figure 3.9).

3.2.4 The GLOBAL-POLICY module

Module overview

The module GLOBAL-POLICY represents policy and other global level parameters in the form of *tunable parameters* that the model users set according to scenarios they want to explore. The module also stores some parameters of household or landscape agents so that users can input values. Parameters stored in this module are accessible to or affect all household and land-patch agents, thus are also called global parameters.

Policy factors specified in this model version represent some of the most important governmental regulations directed to agricultural development and forest management in protected watersheds in the Vietnam upland. The policy factors include:

i) Spatial zoning regulations for watershed forest protection,
ii) Agricultural extension program(s) and a agrochemical subsidizing policy that encourage ethnic minority farmers to intensify agricultural production, with the hope to increase agriculture-based income and thereby reduce the human pressure on the protected watershed areas.

The detailed descriptions of these policy issues (with regard to land use) and the needs of policy decision supports will be addressed in Chapter 6. In this section, we only parameterize these policy factors at an approximate level for modeling.

Watershed protection zoning regulation(s)

Watershed zoning regulation is here formulised as a *set of rules for defining the spatial extent* of the protection zone. As the purpose of watershed protection is to prevent accellerating irreversible surface movement, such as soil erosion/sedimentation, most of the rules for zoning the protection area consider the factors of landform (e.g., slope geometrical indices) and surface stability (e.g., soil texture). As zoners often face a number of criteria, zoning rules usually deal with a function to calculate a single combinational *zoning index*. Given a zoning index, a threshold for the rule is required. For example, the simplex protection zoning rule may be: "*a patch will be protected* *from any non-forestry activity IF its slope > 25°*", or formally expressed as: IF P_{slope}>25° THEN $P_{protect}$ = 1 ELSE $P_{protect}$ = 0. In this case, the zoning index is simply the slope angle, and the threshold for protection is the slope angle 25°.

Which zoning index to use is more or less a technical issue and nation/locality-specific, as each nation or locality normally has its own zoning code, often defined by law. To avoid such a local specification, this module of VN-LUDAS does not deal with any procedure for computing any particular zoning index. The grid of the zoning index should be calculated outside the model, where users can apply any index they wish, then the grid of this index will be input into the model and stored in the PATCH-LANDSCAPE module as a normal variable of landscape agents P_{zoning} (see Equation 3.29). For example, in the case of the Hong Ha watershed (Vietnam), the used zoning index is calculated based on a lookup table including four factors (i.e., rainfall, slope, relative elevation, and soil physical condition) proposed by the Forestry Inventory and Planning Institute of Vietnam (FIPI).

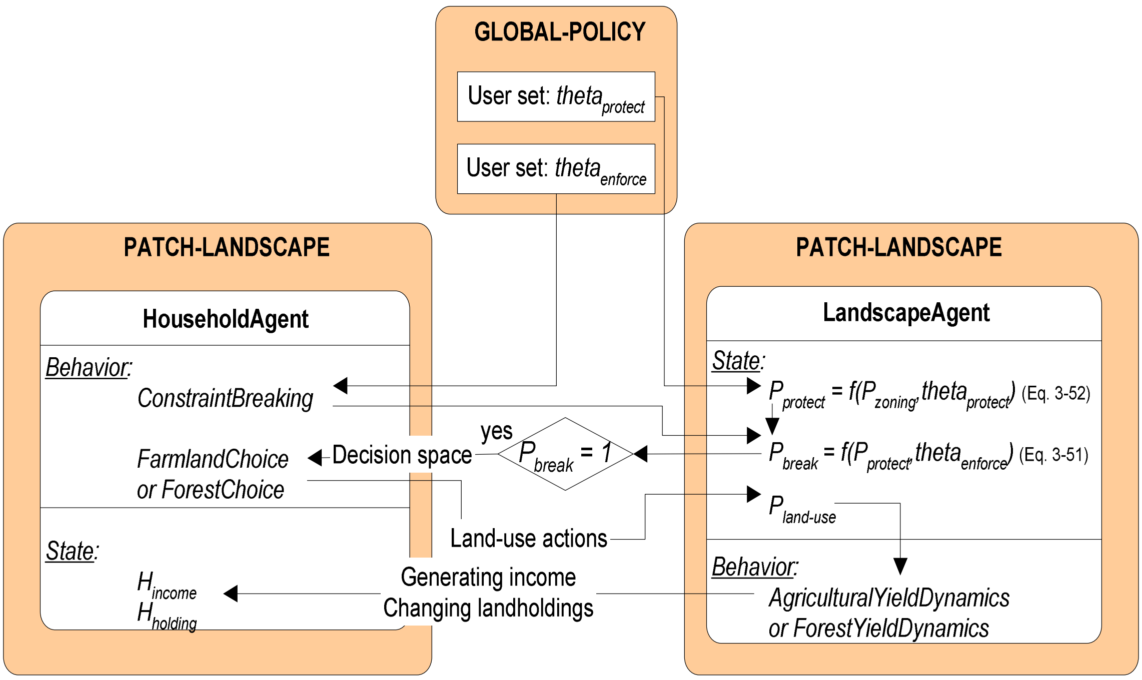


Figure 3.10 Graph shows interactive routes of zoning policy (represented by $\theta_{protect}$ and $\theta_{enforce}$) in the system

Given a surface (grid) of a zoning index [P_{zoning}], defining the threshold value used in protection rules is definitely a debatetable issue in physical planning policy.

This can be seen at the two following points. First, by tuning the threshold, the planners control the extent of the protection zone. Second, the shifting of these thresholds at only few units will create massive consequences in the living space of local people, as spatial distribution of institutional accessibility to land resources will change. Govermental organizations who are responsible for watershed protection normally tend to reduce the slope threshold to include a larger area for protection, whereas local communities want to increase the threshold to have more space for agricultural production and harvesting timber. This is a conflict point in watershed planning. Assuming that the watershed planning process is facilitated following a participatory approach, then, during negotiation processes, stakeholders may wish to see how the changes in protection zoning rules affect the dynamics of the environment and local livelihoods. These explorative trajectories will provide a basis to assist participatory processes in obtaining better consensuses.

The protection zoning policy is therefore characterized by two parameters (which are input by users):

i) The threshold of zoning index P_{zoning} for defining protection area ($\theta_{protect}$), and

ii) The enforcement degree of the protection regulation ($\theta_{enforce} \in [0,1]$).

The interactive routes of changing the threshold of protection zoning on the household agent behavior in the system are formally expressed as in Figure 3.10. Related equations are in Section 3.2.2.

Agricultural extension policy

Access to agricultural extension services directly affects land-use decisions of upland farmers (see chapter 4). Farmers with access to extension schemes may change their incentives in adopting agricultural land uses or may have better opportunities to intensify land use, thus their land-use decision space may be changed accordingly.

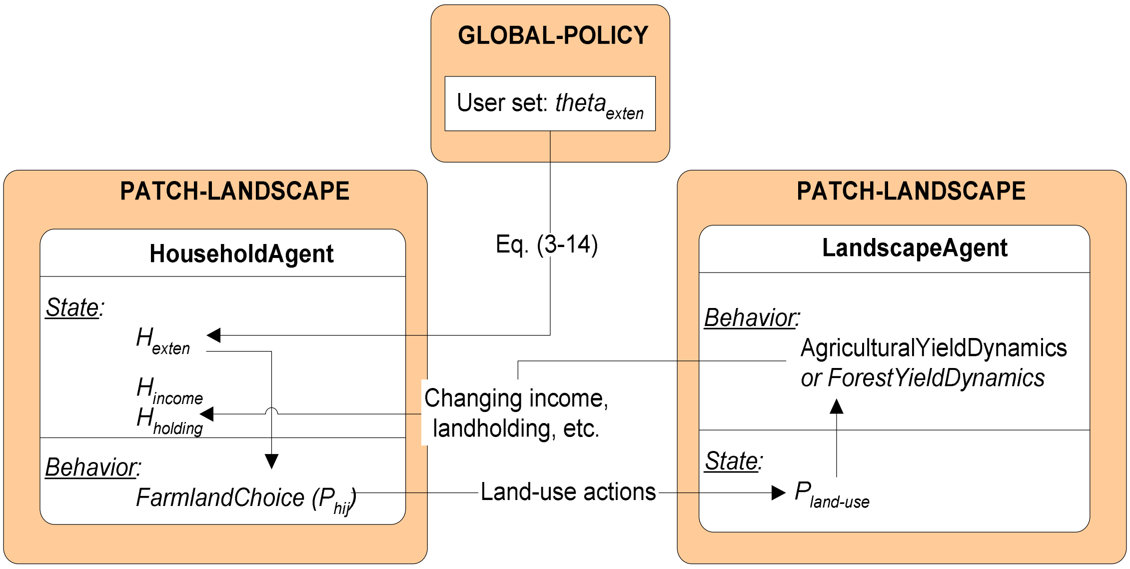


Figure 3.11 Graph shows interactive routes of extension policy (represented by θ_{exten}) in the system

In VN-LUDAS, the extension policy factor is approximately represented by the percentage of farming households who have access to extension services (θ_{exten}), as expected or planned by agricultural extension schemes. This parameter is defined by users as they wish to explore the effects of these policy factors on the system performance. The interactive routes of the exension policy factor on the household agent bebaviour are shown in Figure 3.11. The related equations are in Section 3.2.1 (equation 3.14) and section 3.2.3 (equations 5.54 and 5.56).

Agrochemical subsidizing policy

In association with extension programs, the government may subsidize industrial fertilizers and pesticides to encourage poor farmers to intensify their agricultural production, with the hope of stabilizing upland livelihoods and reducing pressure on forest resources. Access of households to agrochemical subsidy programs is a variable strongly affecting land-use choices of poor farmers (see chapter 4). Thus, it needs to be explored how changes in the subsidy factor affects the land-use decision of upland farmers, and thereby the landscape dynamics.

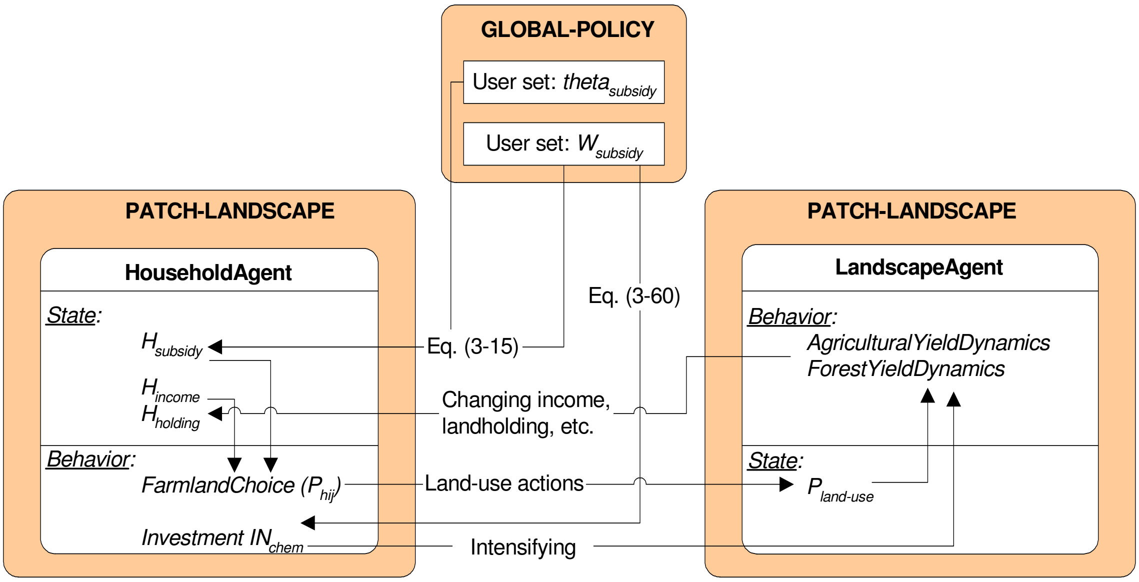


Figure 3.12 Interactive routes of the agrochemical subsidy policy (represented by $\theta_{subsidy}$ and $W_{subsidy}$) in the system

The agrochemical subsidy factor is represented by two parameters:

i) The percentage of farming households who have received subsidies ($\theta_{subsidy}$) (i.e., subsidy access), and

ii) The subsidy amount ($W_{subsidy}$) as planned by the governmental subsidy program.

The subsidy factor plays along the two main routes (Figure 3.12). First, a change in the subsidy variable of household agents ($H_{subsidy}$) will effect the computed land-use choice probability P_{hij}. Second, those household agents receiving agrochemical subsidy will increase agrochemical input to crop production, thus increasing crop yield. The increased crop yield will add income (H_{income}) to the household, thus will affect his land-use choice P_{hij} (Figure 3.12). The related equations are in Section 3.2.2 (equations 3.15, 3.35, and 3.60).

3.3 Main steps of the simulation process: The simulation protocol of VN-LUDAS

At the implementation level, the simulation program consists of ten main steps (Figure 3.13). The main time-loop of the simulation program, called *annual production cycle*, includes sequential steps, which are agent-based and integrated with patch-based processes. In most cases, all household agents and landscape agents (patches) are called and perform tasks in parallel (i.e., synchronising actions).

Brief descriptions of the main simulation steps are as follows:

i) *Set up the initial state of the system:*

- Set up initial states of household and landscape agents. This task is performed by the *Initialization* routine, including the following steps:
    - o Import the sampled household data, which was gathered through a intensive household survey (see Section 3.2.1 of this chapter and Section 4.3.3 of chapter 4).
    - o Regenerate the remaining fraction of the total population by propagating the sampled household dataset up to the exact size of the total population.
    - o Import the spatial dataset of the study landscape (see Section 3.2.2 of this chapter, and chapter 5), and establish links between the sampled household agents and the spatial dataset. These links were established and calibrated using data from a plot-based land-use survey for all sampled households (see Section 4.3.3 of chapter 4).
    - o Generate land parcels hold by the new regenated households using spatial random rules. According to these rules, given a number of land parcels (with explicit land-use types) of a new generated household, the corresponding locations of such parcels are randomly selected among the patch-set bounded by the polygons of the household's village territory and the relevant land-use type.
- Set up policy and other global parameters, according to the scenario defined.
- Set up: initial time step = 0, annual labor budget = 0, annual income = 0

Start the main time-loop (i.e., Annual Cycle), the sequential steps are:

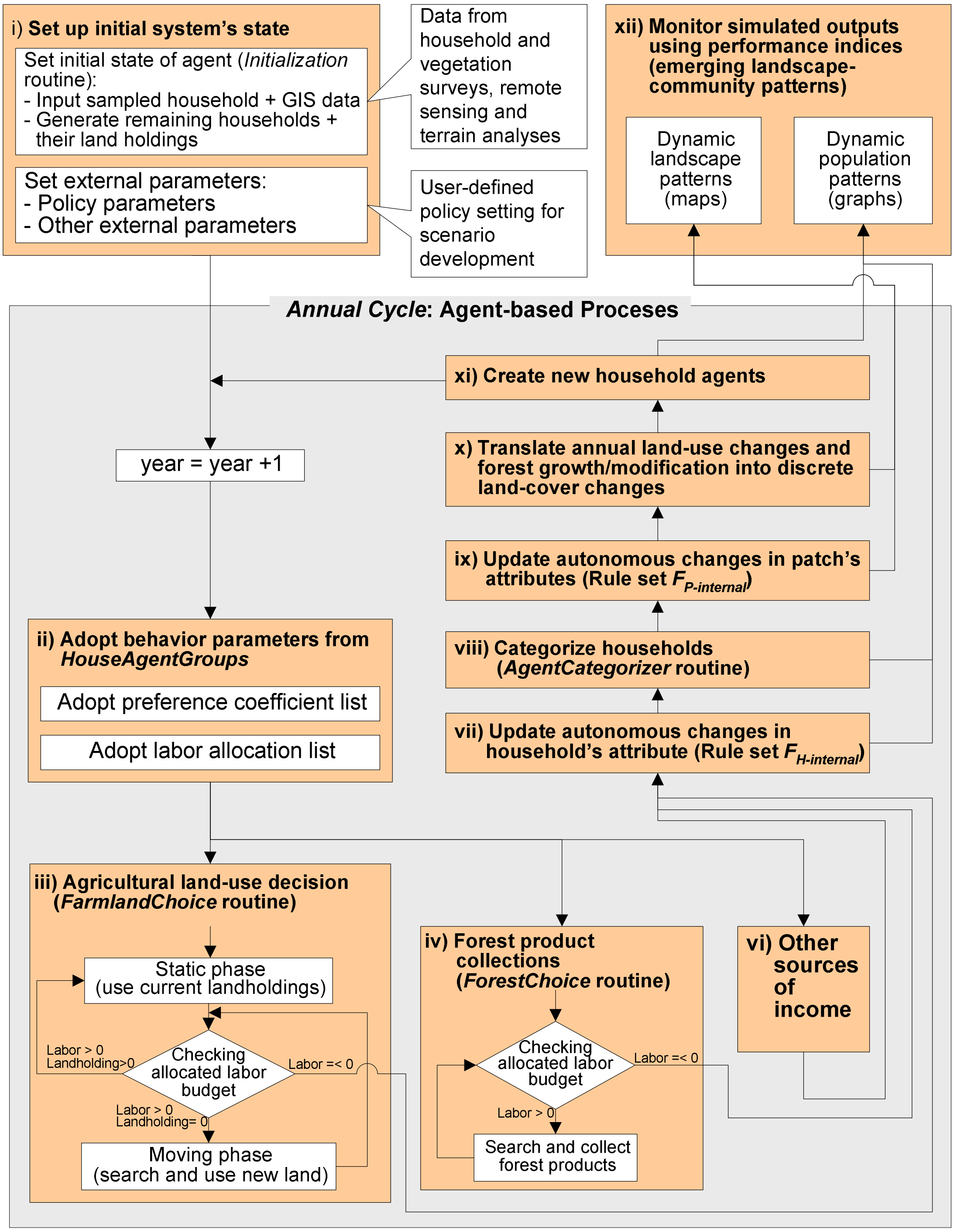


Figure 3.13 Flow chart showing main steps of multi-agent-based simulation process

ii) Adopt behavior parameters from household agent groups:

- Adopt the vector of preference coefficients of land-use choice functions: the household agent parameters vary bounded-randomly around the deterministic parameters of household group ($\beta_x \pm \sigma_x$) (see equation 3.17, Section 3.2.1) The group's preference coefficients β_x and their standart errors σ_x are taken from emprical land-use choice analyses (Section 4.4.2 of chapter 4).

- Adopt the vector of labor allocation percentage of production lines: The vector of labor allocation percentages of production lines is taken from the labor allocation vector of household group (see equation 3.18, Section 3.2.1). This group vector were derived from descriptive statistics of the sampled household groups, which were pre-calculated outside the simulation. During the simulation run, if the household changes his group membership, his labor allocation vector will change accordingly.

Parallel decision-making and action processes (agent procedures and patch's procedures within context of household agent procedure):

iii) Decision about use of landholdings for agriculture (FarmlandChoice procedure): This is a goal-driven decision-making process, directed to the utility-maximizing goal and bounded by allocated labor budget. The process includes two sequential phases:

- <u>*Static phase*</u>: Agents prioritise using their current landholdings as much as possible, due to no cost of moving/opening new lands and to utilise their past optimal site selection.
  - o Decide fallowing holding patches (if any): agents consider which patches should be fallowed, based on fallowing rules, and will optimise land-use choice only on the active own-patches.
  - o Calculate land-use decision vector (*DecisionVectorCalculate* procedure) for each active own-patch: agents calculate land-use choice probabilities distributed over *M* land-use alternatives for each active own-patch.
  - o Select best land use *i** for each active holding patch: agents select most utilisable land use, which is with highest choice probability within the decision vector.
  - o Decide input amount: agents decide amount of inputs (i.e., labor and agro-chemicals) used, linked to subsidising level given by the intensification policy.

- Generate income and convert land use: Given input by the household and natural conditions of the site, yield of the selected agricultural type is calculated using *AgriculturalYieldDynamics* sub-model. Subsequently, household agents accumulate annual income based on the production yield calculated, and convert land cover of their patches.
- Account labor budget.

If all agents' active own-patches have been used and the balance of labor budget is still positive, the agents go to the next step to search for new patches.

- *Mobile phase*: Household agents loop over their personal LandscapeVision to find a pair comprising location and land-use type that returns him the maximal utility, following the six main steps:
  - Breaking spatial constraints (*ConstraintBreaking* procedure): household agents check physical and institutional constraints over the *LandscapeVision*, and at every patch, random-proportionally decide whether to break or accept these spatial constraints. Only constraint-broken patches are included in land-use evaluation in the next step.
  - Calculate land-use decision matrix (*DecisionMatrixCalculate* procedure): household agents calculate spatial land-use choice probabilities distributed over two-dimensional space: M land-use alternatives $\times$ N accessible patches within *LandscapeVision*, i.e., generating household agent's personal land-use decision matrix.
  - Select the best pair of patch + land-use: Based on their land-use decision matrix, household agents select the patch and coupled land-use type giving highest choice probability.
  - Decide input amount: similar to static phase.
  - Generate income and convert land-use: similar to those in the static phase.
  - Account labor budget

The process is repeated until the labor budget is finished.

iv) Decision about collection of forest products (*ForestChoice* procedure): This is a heuristic rule-based decision-making process consisting of three steps:

- *Search patches* that satisfy particular conditions to cut timber.
- *Determine the amount* of logged timber.
- *Generate income, modify vegetation status, and accounting labor budget*.

The process is repeated until the labor budget is finished.

v) Generate income from other activities: as we do not model non-land-based production (e.g., livestock production, off-farm activities, NTFPs, etc.), household agents simply generate income from production based on the amount of labor allocated.

vi) Update basically autonomous changes of household agent attributes: Household attributes (e.g., age, labor availability, accessibility to projects/programs, etc.) are updated by the functioning of the $F_{H\text{-}internal}$ rule set (see Section 3.2.1 of this chapter).

vii) Categorize agents: Categorize agents into *groups of livelihood typology* for changing agent behavior structure:

- Execute *AgentCategorizer* to categorize agents into their closest groups.
- If any agent is assigned into another group, he will update his behavior structure from that group, otherwise he maintains the past behavior structure. In both cases, the household agent parameters vary bounded-randomly around the deterministic parameters of household group ($\beta_x \pm \sigma_x$).

viii) Update basically autonomous changes in patch attributes: Gradual changes in forest yield on forested patches (either intervened or not intervened by households) are calculated using the *ForestYieldDynamics* sub-model.

ix) Translate land-use change and annual forest growth to discrete land cover change: As land cover changes are direct results of land-use changes or gradual vegetation growth, this step employs the *NaturalTransition* sub-model to translate either land-use changes or annual forest growth to land cover changes, looping over all patches of the whole landscape.

x) *Create new household agents*: The number of new agents born every year can be calculated based on an empirical equation. It is assumed there is no migration[7] of households, thus the number of annual new agents can be approximated by the annual increment of total agents empirically observed.

New agents generate their attributes *randomly-bounded* around the average values of a "young" household agent class. New household agents will find their patches in the next time step following all sequential steps above, except that they will ignore the static phase in step (v) as they have no landholding yet.

xi) Draw monitoring graphs of performance indices: Annual changes in the performance indices of human system (e.g., total population, average household income, income structure, income distribution via Lorenz curve and Gini index) and landscape environment system (areas of land-use/cover types, productivities of agricultural types and forest stand) are plotted. Spatio-temporally explicit dynamics of land use/cover are shown on a movie window.

Repeat in the next time step.

3.4 Conclusion

In this chapter, a multi-agent system for simulating land-use and land-cover changes, named VN-LUDAS (Vietnam Land Use Dynamics Simulator), in a mountainous rural

[4] In some cases such as in our study area, individual persons in fact sometimes leave their households to migrate to other communities/cities for working, studies, etc. However, it is very rare that a whole original household migrates. Thus, the assumption of no household agent's migration is quite reasonable.

landscape is constructed and specified. The goal of developing this model is to explore alternative scenarios to improve mountainous livelihoods and mitigate deforestation rates, thereby supporting the negotiation process of multi-stakeholders in land-use planning and management in the upland in the Central Coast of Vietnam and other similar areas.

We constructed an interactive human-landscape system where farming households (i.e., household agents) and land patches (i.e., the landscape agents) play as constituent units of the system. Basic objects and their linkages, especially the linkages perception-knowledge presentation – action with agents, were mapped out. The framework provides a platform where many techniques already developed in spatial modeling can be integrated. For instance, in the first development of the model, we nested the bounded-rational decision mechanism (e.g., the maximisation of parameterised utility functions) with the reflex mechanism (set of relfex rules) to represent more relevantly the decision-making mechanisms of farming households about land use. Although many features of the complex processes of human decision-making are not included in the model yet, the agent-based system has flexibility for adapting to upgrading and modifications.

The model specification, module-by-module and object-by-object, clearly shows an explicit and fully parameterized architecture which accounts for the evolution of the coupled human-environment systems. The proposed agent-based architecture allows integrating diverse personal, environmental and policy-related factors into upland farmers' decision-making about land use and the subsequent accumulated outcomes in terms of spatially explicit patterns of the natural landscape and population. As the model architectures are illustrated using graphic languages and the parameterization is in algebraic languages prior to any empirical estimation, the model has better applicability to different contexts. For instance, with particular additional site-specific data, the model is potentially applicable for communes in the uplands of the Central Coast of Vietnam.

4 LAND-USE DECISIONS BY HETEROGENEOUS HOUSEHOLD AGENTS: THE CASE OF HONG HA COMMUNITY, THE CENTRAL COAST OF VIETNAM

4.1 Introduction

Land-use dynamics involve considerations of behaviors of human agents – e.g., farming households at forest margins - that take specific actions governed by their own decision rules. These human agents are engaged in a highly complex game in which they evaluate land-use alternatives. For any major geographical region, there are different typologies of land users with possibly important differences in land-use strategies (Lambin *et al*., 1999). The importance of diversity in the agent's decisions on land use suggests that it is worth making an effort to characterize the heterogeneity of human populations (Fernandez *et al*., 2003). Characterizing these human agents - in their rich diversity - for better understanding of their decision-making on land use has been identified as a priority task in the Implementation Strategy of the Land-Use and Land-Cover Change (LUCC) project, developed under the auspices of IGBP and IHDP (Lambin *et al*., 1999).

Recent developments in the multi-agent system for modeling land-use and land-cover changes (MAS-LUCC) have created new requirements to calibrate and validate the diversity of agent decision-making upon empirical data and to make these models more comprehensive, realistic and rigorous (Parker *et al*., 2002). Although the validity of MAS-LUCC models is recognised to depend on the strength of human decision-making and interactions (Verburg *et al*., 2002), most current MAS-LUCC models are still quite simplified, dealing with hypothetical landscapes and agents (Kanaroglou and Scott, 2002). Thus, efforts are currently underway to build operational multi-agent models for realistic land-use change simulations (Parker *et al*., 2003; Verburg *et al*., 2002; Polhill *et al.,* 2001; Rouchier *et al.* 2001; Barreteau *et al*., 2001). While observed LUCC outcomes may be not sufficient to calibrate such models (Verburg *et al*., 2002), the tendency in this area is to develop well-parameterised and validated models of human decision-making based on sufficient data/information obtained at household and farm levels.

Although requirements for validating MAS-LUCC models remain to be completely determined, the use of decision parameters derived from empirical household datasets for generating agent behavior is a prospective approach to model's calibration (Fernandez *et al.*, 2003; Verburg *et al.*, 2002). Some recent studies have shown that statistically causal analyses of observed data can be used to derive classes of different agent typologies, as well as specific behavior with respect to land-use decisions for each human agent group (e.g., Fernandez *et al.*, 2003). Although the capture of agent diversity using statistical analyses is often limited at the level of agent group only, it still gives potentially less bias (more objective) calibrations of human agent's behavior, and more statistical power with information about the distribution of decision variables and parameters. Moreover, statistical analyses may help defining important criteria and indicators that guide the design of further analyses at smaller-scales to capture individual-specific behavior.

This research assumes: if causal relationships exist between the biophysical environments, the socio-economic characteristics of farmers and the land-use actions they take, farmers with different livelihood typologies living in different environmental and policy conditions will have different behavioral patterns about land-use choices. Based on this hypothesis, the chapter has the two interrelated objectives:

i) to identify livelihood typologies of households and endogenous factors that differentiate such household typologies; and
ii) to assess the combinational effects of socio-economic characteristics of farming households (including particular policy-related variables) and environmental attributes of land for land-use decision-making of each typical household group.

Additionally, the analysis results should identify indicators for how specific groups of farmers respond to governmental policies and environmental attributes in terms of land-use decisions.

From a methodological point of view, the study is expected to illustrate the application of relevant quantitative methods for calibrating household decision-making for use within MAS-LUCC models.

4.2 Socio-economic setting of the study site

4.2.1 Geographic location and boundary of the study area

Given the research objectives, the study was conducted in the Hong Ha commune in the A Luoi district, Thua Thien - Hue province. The study area is the watershed of the Rao Nho river within the Hong Ha commune on the Central Coast of Vietnam. The watershed lies about 70 kilometers west of Hue City, at 16°15'04'' – 16°20'17'' N latitude, 107°15'01'' – 107°23'06'' E longitude, and covers an area of about 90 km^2 (Figure 4.1). The Rao Nho river flows from the east side of the summit area of A Xom Mountain (1242 m a.s.l), through the Hong Ha valley and empties into the Bo River at c. 30m a.s.l.

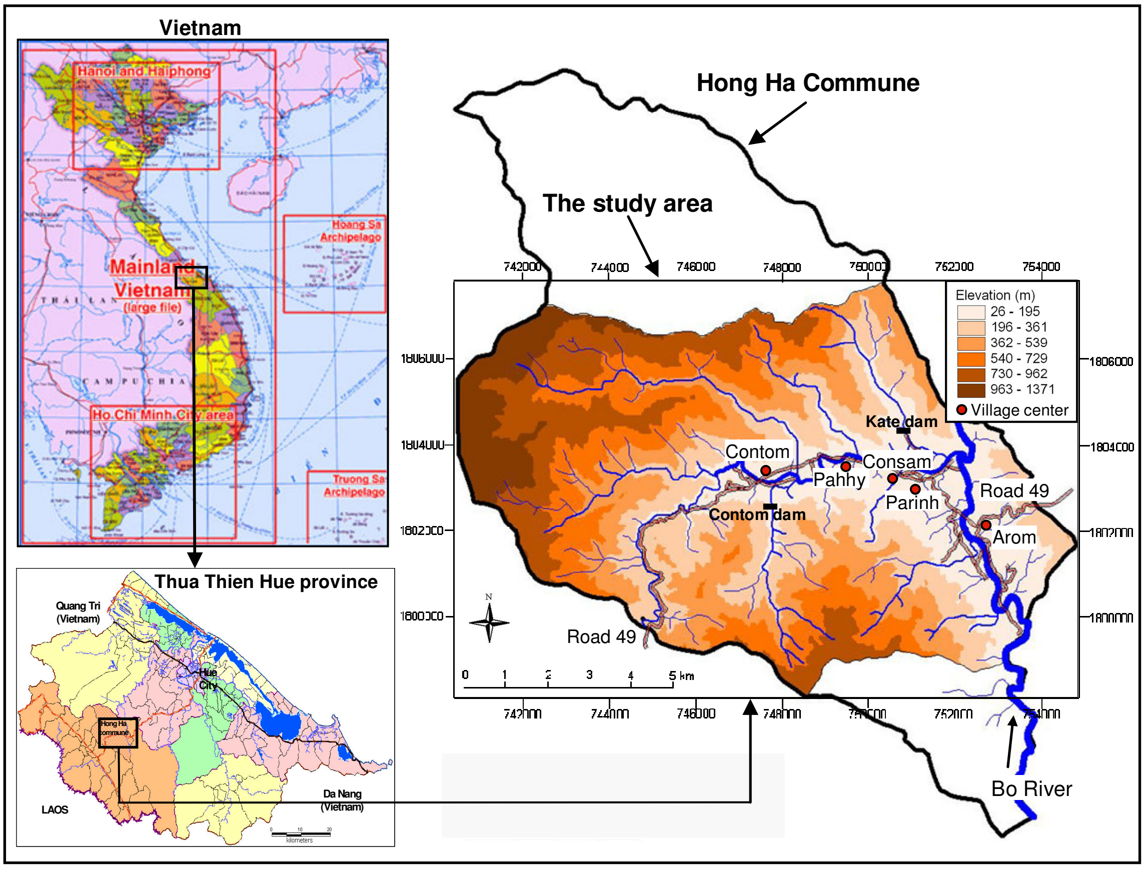


Figure 4.1 Location and boundary of the study area. Map coordinate system: UTM projection, Zone 48 North, Datum WGS84

The commune was chosen because it is highly representative for the uplands in the Central Coast of Vietnam, in terms of land-use system and socio-economic features. Firstly, the commune is the home of ethnic minority groups, fairly representing the population in the region. Secondly, land-use systems found in the commune are typical for the uplands in this area. Thirdly, as in many parts in central Vietnam, Hong Ha commune has been the focus of many pilot interventions by the Vietnam government for promoting agriculture production and forest protection. Therefore, research findings based on the site can be potentially generalized to most areas of the uplands on the Central Coast of Vietnam (see Le Van An *et al*., 1999).

We defined the extent of the study area using both natural boundaries (i.e., the ridges of high mountains) and administrative boundaries to ensure that Hong Ha's villagers do not, or very rarely, use land outside the study area.

4.2.2 Population

Hong Ha commune was resettled in 1974 at the end of the Vietnam-US war, with an initial population of 450 in 74 households. By 2003, the commune composed of 5 villages, namely Contom, Pahhy, Consam, Parinh and Arom, with a total population of about 1200 inhabitants in 240 households. The changes in total population and households over the past 30 years show the same pattern, as depicted in Figure 4.2. The avarage population growth is relatively rapid, about 4.4 ± 1.7 %. There are three ethnic groups in the commune (K'tu, Ta-Oi (including Pa-Ko and Pa-Hy), and Kinh) where K'Tu and Ta-Oi are mountainous ethnic minority groups (Nguyen Xuan Hong, 2002; EMWG, 2004). The K'Tu group is a majority, contributing about 69 % of total population. Agricultural production and collection of forest products (e.g., firewood, timber, rattan, trapped wild animals) are the main livelihoods of most villagers.

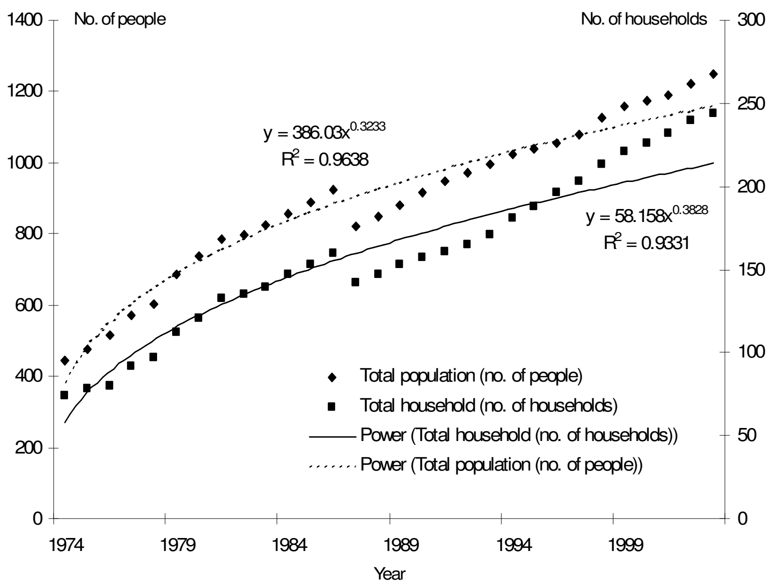


Figure 4.2 Changes in total population and total households in Hong Ha commune over the past 30 years (1974 - 2003). Data source: People Committee of Hong Ha commune (2003)

4.2.3 Main land-use types

As in many areas in the uplands along the Central Coast of Vietnam, four types of land use are most commonly found in the study area are upland crops, paddy rice, fruit-based agroforestry and forest plantation. The upland crop system, practiced traditionally by the ethnic minorities, is a type of shifting cultivation. The system on hillsides takes a crop-fallow cycle of about eight years. Upland rice (*Oryza sativa* L., local variety) is planted in the first one or two years after slash and burn, while the topsoil is still fertile. As the topsoil becomes eroded and degraded, cassava (*Manihot spp.*) is often a successor crop in the follow-up two or three years. After these seasons of cassava, the soil becomes exhausted, and the field is fallowed for a period of three or four years. The crop system on flat land mostly has a longer cropping period, about 7-8 years. The cultivation of upland crop is rain-fed and with almost no or very low input of chemical fertilizers.

Paddy rice (*Oryza sativa* L.) was introduced to the villages at the early stage of the resettlement by the mid 1970s, and is widely practiced by most households. Today, most of the paddy rice fields have two crops a year, mainly using the TH-30 and Khang Dan varieties introduced by the province Department of Agriculture and Rural Development (DARD). About half of the paddy area is irrigated with water from the two small dams (Figure 4.1), and the remaining paddy area is irrigated directly from streams using indigenous irrigation techniques (e.g., bamboo pipe-line). Chemical fertilizers (mostly NPK) and pesticides have been increasingly used in the paddy rice system during the past five years, along with agricultural extension programs. These extension programs, conducted by a variety of governmental agencies, have provided villagers technological guides and subsidized agrochemicals (fertilizers and pesticides) to encourage production intensification, including paddy rice.

Fruit-based agroforestry has been widely practiced in Hong Ha since the mid 1980s. The main fruit products include bananas (*Musa spp.*), pineapples (*Ananas comosus* (L.) Merr.) and jackfruits (*Artocarpus heterophylla* Lam.). In the last 5 years, some households have adopted lemon (*Citrus limon* (L.) Burm.f.), longan trees (*Nephelium longana* Cambess.) and black pepper (*Piper nigrum* L.), which are usually planted in association. A small proportion of land within these agroforestry farms is often used to grow annual crops, such as sweet potato (*Ipomoea batatas* (L.) Poir.) and vegetables. In addition, a few natural forest trees or jackfruit trees are planted along the edge of the plots. Weeding and mulching are done regularly. NPK fertilizer is sometimes applied when fruit crops are first planted. Fruit harvest often begins with bananas and pineapples after two or three years.

Private forest plantations were first introduced to the community in the early 1990s with the support of the United Nations' World Food Program (PAM) and the governmental reforestation program 327 (Program 327). Based on the support through these programs, villagers claimed open areas (shrubs or grass land) for reforesting, or converted their fallowed land to forest plantation. However, the private forest plantation area is a very small fraction compared to the large areas of state-owned forest plantations in the commune. *Acacia auriculiformis* Benth. and *Acacia mangium* Willd. (both Leguminosae) are the only two tree species used for forest plantation. The density of the plantation is ranging from 1600 to 2000 trees ha^{-1}. Management includes pruning main trees, slashing shrubs and tall grass in the first three years, when the planted trees are still small. The harvest of *Acacia* trees can be principally done when the tree is at least 7 years-old. However, the actual production cycle of these forest plantations in the area is not quite clear, as most of them are still at an early stage and have not yet been harvested.

4.3 Methodology

4.3.1 Methods for categorizing household agents

Potential criteria for group classification

We use the livelihood framework concept as a basis for selecting criteria that represent the livelihood typology of households. The livelihood framework includes five core asset categories: *human*, *social*, *financial*, *natural* and *physical* capital (see Ashley and Carney 1999; Bebbington 1999; Campbell *et al.* 2001). This spectrum of livelihood assets is the basis of people's capacity to generate new activities in response to needs and opportunities (Farrington *et al.*, 1999). The concept, which has been vigorously debated in the literature, forms a theoretical basis for deriving indicators/criteria for assessing the performance of natural resource management, and helps to avoid bias selection of indicators from one particular discipline (Campbell *et al.*, 2001).

Within the livelihood framework and based on reviewing previous related studies in the uplands of Vietnam and Indochina, we selected the following variables to represent overall livelihood typology of a farming household (see also Table 4.1):

- Four variables of human resources for representing human capital, i.e., size, educational status, labor availability, dependency ratio (see Castella *et al.*, 2002b; Fatoux *et al.*, 2002; Bui Dung The, 2003; Tran Duc Vien *et al.*, 2001; Gomiero and Giapietro, 2001);
- One variable representing social capital, i.e., leadership (see Tan Nguyen Quang, 2003).
- Three variables of land resources for presenting natural capital, i.e., total managed land, holding area, holding area per capita (see Castella and Erout, 2002; Fatoux *et al.*, 2002; Bui Dung The, 2003; Tran Duc Vien *et al.*, 2001).
- Two variables representing financial capital, i.e., annual gross income and gross income per capita (see Bui Dung The, 2003; Alther *et al.*, 2002; Tran Duc Vien *et al.*, 2001).

Household livelihood strategy is also taken into account in the classification of household types. Livelihood strategies are defined as those activities undertaken by smallholder households to provide a means of living (Koczberski *et al.*, 2001). In the context of the Vietnam uplands and the livelihood strategies of farming households, emphasis is placed on the range of income sources pursued by households (Alther *et al.*, 2002; Castella and Erout, 2002; Gomiero and Giampietro, 2001), or access to means of agricultural production, e.g., cultivated lands, cattle or irrigation system (see Sadoutlet, *et al*. 2002; Beckmann *et al*., 2002). Hence, it is reasonable that we approximated the livelihood strategy of farming households in the study area by the income composition of the household, i.e., percentage incomes generated from different production lines.

Table 4.1 Potential variables representing livelihood structure and strategy of households in Hong Ha commune.

Variable	*Definition*
H_{edu}[a]	Education status of the household head: 1 if more than primary school, 0 if otherwise
H_{size}	Size of a household (number of household members)
H_{labor}	Availability of household labor (number of workers)
H_{depend}	Dependency ratio (number of dependants/H_{labor})
H_{leader}[a]	Leadership of the household: 1 if a household member holds a leadership position, 0 if otherwise
H_{land}	Total land that household manages, including land holdings and contracted forest plantation[b] (m^2)
$H_{holding}$	Total land holdings of a household (m^2)
$H_{holding/pers}$	Landholding per capita (m^2 $person^{-1}$)
H_{income}	Annual gross income of a household (1000 VND $household^{-1}$ $year^{-1}$)
$H_{income/pers}$	Gross income per capita of a household (1000 VND $person^{-1}$ $year^{-1}$)
$H_{\%in\text{-}paddy}$	Percentage income from paddy rice (%)
$H_{\%in\text{-}upcrop}$	Percentage income from annual upland food crop (%)
$H_{\%in\text{-}af}$	Percentage income from fruit-based agroforestry (%)
$H_{\%in\text{-}livestock}$	Percentage income from livestock production (%)
$H_{\%in\text{-}forest}$	Percentage income from natural forests (%)
$H_{\%in\text{-}others}$	Percentage income from other activities (%)

[a] Variables are not included in principal component analysis because their values are in dummy scale.
[b] Contracted forest plantation land is owned by state agency, but contracted to local farmers for taking care.

Statistical analyses for discovering grouping criteria and agent groups

Principle Component Analysis (PCA) for deriving independent composite criteria

It would be complicated to use all 16 variables in Table 4.1 to detect the agent groups. Moreover, our exploratory correlation analyses showed there is a high multi-collinearity between these potential criteria. Thus, an integrative summary of the overall set of potential criteria is neccessary.

We reduced the dimensionality of 14 potential criteria[8] in Table 4.1 using Principle Component Analysis (PCA). PCA is a multivariate statistic method that is often used to condense information in a large number of original variables into a smaller set of new composite dimensions, with a minimal loss of information (McGarigal *et al.*, 2000; Campbell *et al.*, 2001). The basic assumption of the technique is that it is possible to explain the correlation pattern between two or more variables in terms of a few underlying factors, called principle components. The principle component is a linear combination of the original variables that accounts for the maximum possible information in the original set of variables, i.e.,the first principle component $PC1 = b_1X_1 + b_2X_2 + \ldots + b_nX_n$, where X_1, X_2, …, X_n are the standardized original variables, and b_1, b_2, …, b_n are weight parameters (i.e., loadings). The meaning of each principle component is interpreted in terms of those original variables with higher weights/loadings, i.e.,the most important variables. The first principal component (PC1) directs along the greatest variation, then the second principal component (PC2) has the direction with maximum variation in the remaining data, which is orthogonal to the PC1, and so forth. Because these extracted principle components are independent from each other, the use of component scores for subsequent analysis will help to avoid the multi-collinearity problem.

We ran PCA with Varimax rotation and Kaiser normalization. Only components with Eigenvalues over 1.0 were interpreted and used for subsequent analyses. The component scores were saved and standardized. Based on the weight parameters in the rotated component matrix, we selected the most meaningfully original variables corresponding to each principle component for use as key categorizing variables of the *AgentCategorizer* routine in VN-LUDAS. Standardized component scores were used for subsequent cluster analysis to derive household groups.

K-Mean Cluster Analysis using Principle Component scores to derive groups of agents

To derive typical agent groups shaped by livelihood criteria, we used the standardized scores of principle components to run K-Mean Cluster Analyses (K-CAs). K-CA is a

[8] Educational status (H_{edu}) and leadership (H_{leader}) were not included in the PCA because they are in dummy scale. However, these two variables were included in the models of land-use choices later.

non-hierarchical and divisive clustering method that attempts to minimize the intra-cluster variances while maximizing the inter-cluster distances (Kintigh and Ammerman, 1982). The K-CA works by searching for cluster formations that minimize the global Sum of the Squared Error (SSE), where SSE is defined as the total of the squared distances between the cluster's center and each of its members (measured in Euclidean distance) (Kintigh and Ammerman, 1982). Unlike hierarchical methods, K-CA methods avoid problems of chaining and artificial boundaries and work on the original input data rather than on a similarity matrix (Kintigh and Ammerman, 1982). Moreover, in our case K-CA was chosen because we have quite a number of cases (n = 69), thus it would be difficult to interpret grouping results using hierarchical cluster analysis.

4.3.2 Methods for estimating spatial behavior of categorized households to land-use choices

Multi-nominal logistic regression model for land-use choices

After agent groups were derived, we employed multi-nominal logistic regression (called M-logit regression) to identify determinants of land-use choices by each agent group, based on plot-based dataset. The M-logit model is a type of a general model of utility maximization (Green, 1995; Wu, 1998; Kitamura *et al.*, 1997; Bui Dung The, 2003). When the stochastic component of the utility function adopts the Gumbel distribution (or log Weibull), and there is no correlation across land-use types (McFadden, 1973, cf. Nelson *et al.*, 2004); the probability (P_i) that an individual chooses land use i is expressed by the multi-nominal logistic (M-logit) function of the form:

$$P_i = \frac{\exp(V_i)}{\sum_j \exp(V_j)} \tag{4.1}$$

Assuming that the deterministic term (V_i) of the utility function (U_i) is a linear function, then utility V_i for alternative i can be written as:

$$V_i = \beta_{i1} X_{i1} + \beta_{i2} X_{i2} + \ldots + \beta_{iK} X_{iK} + \beta_{i0} = \sum_k \beta_{ik} X_{ik} + \beta_{i0} \tag{4.2}$$

where X_{i1}, X_{i2}, ..., X_{iK} are the attributes (dependent variables) of alternative i, and β_{i1}, β_{i2}, ..., β_{iK} are their parameters, respectively.

From equations (4.1) and (4.2), the form of the M-logit model is expressed as:

$$P_i = \frac{\exp(\sum_k \beta_{ik} X_{ik} + \beta_{i0})}{\sum_j \exp(\sum_k \beta_{jk} X_{jk} + \beta_{j0})} \tag{4.3}$$

Parameters β_{ik}, also called *preference coefficients*, were estimated by the maximum likelihood method based on the plot-based dataset of each household agent group, using SPSS package version 11.0.

Specification of variables of M-logit models

Dependent variable

The dependent variable of the M-logit models is land-use choice (P_{use}) by farming households. The choice is one among four land-use alternatives: annual upland crops, paddy rice, fruit-based agroforestry and (private) forest plantation. The descriptions of these land-use types are given in Section 4.3.2. The spatial distribution of these land-use types are shown in Section 5.4.1.

Explanatory variables

Explanatory variables, which are hypothesised to be important in the choice of land use, include three groups (see Table 4.2). The first group of independent variables is environmental attributes of land plots (spatial variables). The second is socio-economic characteristics of farming households. The third includes household's access to important policies related to land use. The justification of the roles of independent variables in land-use choice is summarised as follows.

Variables of natural environment

Crop production, especially fruit-based agroforestry, requires a transportation infrastructure for accessing the land plots and for exporting products from the plots. Thus, it is important to include the distance from each plot to nearest road ($P_{d\text{-}road}$) and to farmer's house ($P_{d\text{-}house}$) in the models of land-use choice (e.g., see Fox *et al*., 1994;

Briassoulis, 2000; Wu and Yeh, 1997; Wu, 2000; Kitamura *et al*., 1997). However, the selected upland crop fields sometimes may not necessarily be close to a road or the house either because the exposure of the swidden plots to the government agencies is to be avoided, or the areas along roads are occupied by other land use. Thus, the directions of these effects may be ambiguous, rather depending on the specific nature of each land-use type (e.g., see Fox *et al*., 1994; Bui Dung The, 2003).

Water availability critically affects agricultural production. The wetness index (P_{wet}) is a terrain variable indicating approximate spatial pattern of soil moisture content, which is important in agricultural production (De Roo, 1998; Wilson and Gallant, 2000). A land plot with a higher index value means the land is more likely to store water. The distance from the land plot to rivers/streams can influence the choice of land-use (see Fox *et al*., 1994) in some different ways. Paddy fields are normally located near rivers/streams, as the paddy rice needs to be irrigated. Upland fields are more likely to be chosen near streams/rivers, and are mostly associated with other activities on rivers/streams (e.g., fishing) or water for domestic uses, not necessarily for irrigating crops. Thus, the direction of effect of distance to river may be not consistent among different land-use choices.

Slope gradient (P_{slope}) determines overland and subsurface flow velocity and runoff rate, thus indicating soil erosion potential (Wilson and Gallant, 2000; Pallaris, 2000). Land plots with steeper slopes have a higher risk of soil erosion, and subsequently return lower crop yield (see Section 5.4.2). Therefore, we hypothesised that slope gradient significantly and negatively influences the land-use choice.

Variables of household characteristics

The age of the household head (H_{age}) influences his attitude towards risk and uncertainty in land-use choice (Bui Dung The, 2003). Older farmers likely become more "*conservative*" and less susceptible to new land-use technologies, such as fruit-based agroforestry (Bui Dung The, 2003). However, this issue seems to be ambiguous in a general context.

Educated farmers may take into account environmental risk (e.g., soil erosion), and have better understanding of the trade-off between current and future benefits/costs in land-use planning (Bui Dung The, 2001). Common sense says that, on the one hand, farmers with better access to education tend to use land more intensively and adopt new technologies earlier, thus saving land to achieve a given income (Müller, 2003). On the other hand, they have better understanding of the trade-offs between current and future benefits/costs in land-use planning, and are thus more likely to select fruit-based agroforestry (Bui Dung The, 2001). However, educated farmers also may clear more forest land as shown by some empirical evidence (Pichon, 1997). In the context of central Vietnam uplands, better access to education is likely to be associated with less swidden practice and more permanent land use (Müller, 2003). Therefore, fruit-based agroforestry is anticipated to be more likely with educated farmers.

Local leadership (H_{leader}), indicating the political power of a household, was anticipated to be important in land-use choice. Political power of farmers relates to their access to government programs and outside assistance, which could not be measured directly (Müller, 2003). A household with leadership may have more possibilities to intensify their farming, thus reducing the extension of upland crop fields and pressure on forests. Moreover, in rural Vietnam, local leaders often receive a government salary and are thus less dependent on agriculture than others. Therefore they may also be more likely to select land-use types that return benefits in the long term and are less demanding with respect to labor, such as fruit-based agroforestry or forest plantation, rather than paddy rice and upland crops.

Labor availability, i.e., number of workers of the household (H_{labor}), is an important input for agricultural production, thus is hypothesised to significantly influence land-use choice (e.g., see Fox *et al*., 1994; Gomiero and Giapietro, 2001). The dependency ratio (H_{depend}), i.e., ratio of dependants to workers, reflects the number of mouths each worker feeds, thus relating to the urgency in food demands of the household (see Fatoux *et al*., 2002; Tan Nguyen Quang, 2003). Households with a high dependency ratio normally have a more urgent need of food, thus they may choose land use that potentially meets their food demands in the short term. A high dependency ratio should be a reason to favor upland crops or paddy rice rather than fruit-based agroforestry.

Table 4.2 List of explanatory variables used in multi-nominal logistic (M-logit) regression models for land-use choices

Variable	*Definition*	*Data source*	*Direct linked module*
Dependent variable: Land-use choice by households			
P_{use}	1 if annual upland crop is chosen, 2 if paddy rice is chosen, 3 if fruit-based agroforestry chosen, 4 others (i.e., forest plantation)	Interview + field observation	PATCH-LANDSCAPE
Environmental attributes of land plots:			
P_{road}	Distance from the plot to roads (m^2)	GIS-based calculation	PATCH-LANDSCAPE
P_{house}	Distance from the plot to owner's house (m^2)	Map-based calculation	PATCH-LANDSCAPE
P_{river}	Distance from the plot to rivers/streams (m^2)	GIS-based calculation	PATCH-LANDSCAPE
P_{slope}	Slope angle (degree) of the plot	Field measurement	PATCH-LANDSCAPE
P_{wet}	Wetness index of the plot	GIS-based (DEM-driven) calculation	PATCH-LANDSCAPE
Characteristics of plot's owner (i.e.,household):			
H_{age}	Age of the household head (year)	Interview	HOUSEHOLD-POPULATION
H_{leader}	Leadership of the household: 1 if a household member holds a leadership position, 0 if otherwise	Interview	HOUSEHOLD-POPULATION
H_{edu}	Education status of the household head: 1 if more than primary school, 0 if otherwise	Interview	HOUSEHOLD-POPULATION, GLOBAL-POLICY
H_{labor}	Labor availability of the household, i.e.,number of workers of the household. An extra worker is accounted as 0.5.	Interview	HOUSEHOLD-POPULATION
H_{depend}	Dependency ratio of the household = number of dependants/total household members	Interview	HOUSEHOLD-POPULATION
$H_{holding/pers}$	Land holding/person (m^2 $person^{-1}$)	Calculation based on interview data	HOUSEHOLD-POPULATION
$H_{income/pers}$	Annual gross income/person (1000VND $person^{-1}$ $year^{-1}$)	Calculation based on interview data	HOUSEHOLD-POPULATION
Policy-related variables:			
H_{extens}	Accessibility of the household to agricultural extension services in the past 3 years: 1 if the household got extension services, 0 if otherwise	Interview	HOUSEHOLD-POPULATION, GLOBAL-POLICY
$H_{subsidy}$	Accessibility of the household to fertilizer subsidy = amount of subsidized fertilizers the household received in the cropping year 2002/2003 (1000 VND $year^{-1}$)	Interview	HOUSEHOLD-POPULATION, GLOBAL-POLICY

Current availability of land holding is often a driving force for land-use choice (Fox *et al*., 1994; Bui Dung The, 2003; Tran Duc Vien *et al*., 2001). Land holding per capita ($P_{holding/pers}$) is often an indicator for measuring availability of land holding. However, the affecting direction of $H_{holding/pers}$ is not consistent and depends on the concrete land-use types.

Wealth status, represented by annual gross income per capita ($H_{income/pers}$) is an important factor affecting land-use choices of upland farmers (Fox *et al*., 1994; Bui Dung The, 2003), or the extent of land-use types (Tan Nguyen Quang, 2003).

Policy variables

Accessibility of the household to agricultural extension services (H_{extens}) was hypothesised to be an important factor influencing land-use choice. It is widely recognized that farmers with better access to these services will likely adopt the land use promoted by the extension program.

In the Vietnam uplands, changes in agricultural technology are linked with the introduction of industrial fertilizer (mainly NPK) and pesticide for increasing crop yields. Along with many government programs aiming to improve the livelihoods of poor upland communities, ethnic minorities often receive subsidy in the form of fertilizers. As the application of agrochemicals increases the productivity of land, the provision of agrochemicals allows farmers to reduce costs for farming input, and thus the agricultural profit of subsidized farmers is increased. Given labor constraint and the intensification goal directed by the government (i.e., the main subsidizer), increased yields and more input on existing plots of encouraged land-use should be more likely. However, on the other hand, farmers may distribute the agrochemicals on other plots to increase the yields in other land use.

4.3.3 Data sources

The socio-economic data used in this study were derived from an intensive household survey conducted in the Hong Ha commune during the summer 2003. The topics addressed in the survey are household livelihood characteristics, information about holding plots, and accessibility of households to agricultural policies/programs. The survey was carried out through face-to-face interviews using a structured questionnaire.

This questionnaire, especially the answer choices and the related coding system, was developed and localised through topical Participatory Rural Appraisals (PRAs) in the five villages of the commune in December 2002, and then pre-tested with 5 households (one in each village). Then, the fully coded questionnaire was developed (Quang Bao Le, 2003).

Random sampling stratified by the 5 villages of the commune selected 69 households - accounting for about 30 % of the population of the communes – for an intensive interview campaign. As a result, a data panel of 69 households, containing coded information about their demographic and livelihood structures, was developed. The household database was used for statistically deriving household groups of typical livelihood structure in the study population, as well as for characterising these groups.

Plot-explicit data were gathered using *enhanced participatory mapping*, which is a participatory mapping process supported with UTM topographic map, Geographic Information System (GIS) and Global Positioning System (GPS) technologies to enhance the spatial accuracy of the mapping. The procedure is briefly described in the following. First, interviewees, village heads and commune's cadastral officers were asked and facilitated to map all holding plots of the interviewed household on a localized topographic map scale 1:10,000 (enlarged from the orgininal scale 1:50,000) with all map figures marked with local names. This mapping process was iterative, and both printed maps and GIS interface of *aerial photograph-based maps* on a laptop were used as mapping interfaces. An aerial photograph-based map is a geo-referenced aerial photograph with map grids, contour lines, roads, rivers/streams network, bridges, and other recognizable features with local names. The aerial photographs were taken in June 1999, covering the whole study area. Educated farmers and local cadastral officers played a key role in communicating with the interviewees. Second, GPS checking was done in field visits to validate uncertain points. All relevant landholding plots were coded and co-registered into the GIS database. Locations of the surveyed plots are shown in Figure 4.3a.

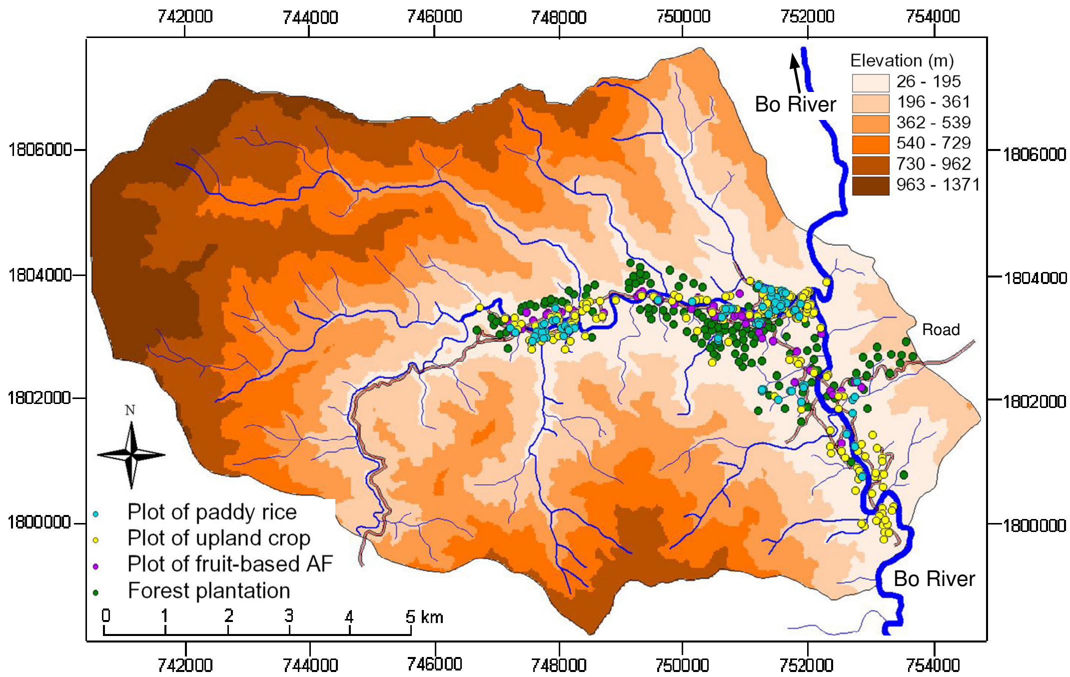


Figure 4.3a Locations of surveyed holding plots in the study area. Map coordinate system: UTM, Zone 48 North, Datum WGS84

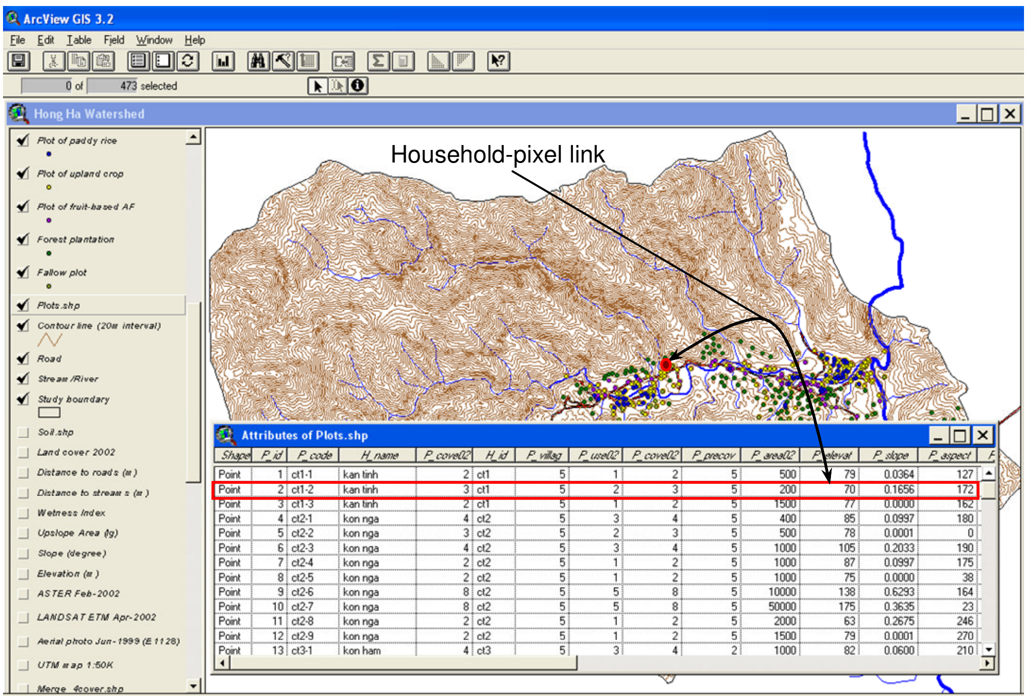


Figure 4.3b GIS interface and the underlying nested household-plot data panel showing the explicit linkages between households and land pixels

As all plots were geo-referenced, the plot variables of terrain indices, distances to roads, rivers/streams and farmer houses were extracted from the thematic GIS raster grids. The raster grids of terrain indices were calculated from a Digital Elevation Model (DEM) of 30m×30m cell-size, which was interpolated from a digitised UTM topographic map scale of 1:50,000. Networks of rivers/streams and roads were also digitized from the topographic map and geo-referenced aerial photographs. The development and products of these GIS grids are shown in Sections 5.3.1 and 5.4.1.

By nesting the extracted physical data with the socio-economic data from the interviews for every survey plot, a spatially explicit database of 367 plots, i.e., the land holdings of the 67 interviewed households, was developed. In the database, plot-specific attributes are nested with the characteristics of plot owners (including policy-related variables) (see Figure 4.3b), thus allowing to conduct spatially causal analyses to identify and estimate determinants of land-use choices for each typical group of households.

4.4 Results and discussions

4.4.1 Identification of typological agent groups

Underlying factors explaining the differences in household typology

The PCA run extracted six (6) principle components with total Eigenvalues greater than 1.0, explaining 82 % of the total variance of original independent variables (Table 4.3). The rotated component matrix then helps to determine what the components represent (Table 4.4).

In Table 4.4, the principle component 1 (PC1) is strongly related to *land* variables, i.e., total landholding $H_{holding}$ (loading b = 0.942), landholding per capita $H_{holding/pers}$ (b = 0.920), and total managed land H_{land} (b = 0.825). Thus, we named this component "*land factor*". The land factor accounts for for 23.3 % of the total variance of the original dataset. Pair correlations among these three variables show they are highly multi-collinear. Because *landholding per capita* ($H_{holding/pers}$) has a very high loading and a more economic meaning than the two others, this variable is the best representative for the land factor.

Table 4.3 Total variance explained by extracted components, using Principal Component Analysis (PCA) as the extraction method

Component	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	3.267	23.337	23.337	3.267	23.337	23.337	2.766	19.757	19.757
2	2.730	19.500	42.837	2.730	19.500	42.837	2.759	19.705	39.462
3	1.895	13.535	56.372	1.895	13.535	56.372	2.250	16.070	55.532
4	1.373	9.808	66.180	1.373	9.808	66.180	1.274	9.103	64.635
5	1.100	7.858	74.038	1.100	7.858	74.038	1.242	8.872	73.507
6	1.097	7.834	81.871	1.097	7.834	81.871	1.171	8.365	81.871
7	0.805	5.750	87.621						
8	0.761	5.432	93.054						
9	0.507	3.622	96.676						
10	0.213	1.518	98.194						
11	0.122	0.869	99.063						
12	0.058	0.412	99.475						
13	0.040	0.285	99.759						
14	0.034	0.241	100.000						

The principle component 2 (PC2) is most weighted by *income* variables, i.e., annual gross income per capita $H_{income/pers}$ (loading b = 0.896), annual gross income H_{income} (b = 0.862), and percentage income from other off-farm activities $H_{\%in\text{-}other}$ (b = 0.798). Thus, we labelled the component "*income factor*". The income factor accounts for 19.5 % of total variance of the original dataset. Annual gross income per capita ($H_{income/pers}$) is the best representative of the income factor, because the variable has the highest weight and is most economically meaningful among the other variables of income.

The component 3 (PC3) is most highly correlated with labor availability P_{labor} (b = 0.830) and household size P_{size} (b = 0.848), thus called "*labor factor*". Pair correlation of the two variables showed they are strongly correlated (r = 0.696). Because of the high loading value and being an important input for production, labor availability (P_{labor}) was selected to represent the labor factor.

Table 4.4 Rotated Component Matrix (i.e., loadings) using Varimax rotation method and Kaiser normalization of first six principle components

Variable	*Principle Component*					
	1- Land factor (23.3 %)	*2- Income factor (19.5 %)*	*3- Labor factor (13.5 %)*	*4- Depend- -ancy factor (9.8 %)*	*5- Livestock factor (7.9 %)*	*6- Paddy rice factor (7.8 %)*
Household size (H_{size})	0.173	-0.064	**0.848**	0.391	0.006	-0.029
Labor availability (H_{labor})	0.256	0.003	**<u>0.830</u>**	-0.322	0.047	-0.051
Dependency ratio (H_{depend})	-0.171	-0.085	-0.061	**<u>0.935</u>**	-0.126	0.002
Total managed land (H_{land})	**0.825**	0.099	0.365	0.071	0.109	-0.042
Total landholdings ($H_{holding}$)	**0.942**	0.016	0.112	-0.070	-0.080	-0.082
Total landholding/person ($H_{holding/pers}$)	**<u>0.920</u>**	0.024	-0.125	-0.193	-0.095	0.019
Annual gross income (H_{income})	0.272	**0.862**	0.143	0.089	0.145	0.070
Annual gross income/person ($H_{income/pers}$)	0.162	**<u>0.896</u>**	-0.215	-0.072	0.069	0.079
Percentage income from upland crops ($H_{\%in\text{-}upcrop}$)	0.187	-0.525	0.323	-0.056	-0.483	0.367
Percentage income from paddy rice ($H_{\%in\text{-}paddy}$)	0.056	-0.127	0.063	-0.033	-0.048	**-<u>0.921</u>**
Percentage income from agroforestry ($H_{\%in\text{-}af}$)	-0.139	-0.092	0.472	-0.136	0.264	0.383
Percentage income from livestock ($H_{\%in\text{-}livestock}$)	-0.042	-0.042	0.089	-0.128	**<u>0.818</u>**	0.104
Percentage income from upland crops ($H_{\%in\text{-}forest}$)	0.104	-0.501	-0.517	0.154	0.295	0.067
Percentage income from upland crops ($H_{\%in\text{-}other}$)	-0.240	**0.798**	0.065	-0.143	-0.330	0.046

Notes: - Numbers in parenthesis are percentages of total variance of original variable set explained by the principal components.
- Bold numbers are the high loadings, indicating most important original variables representing the principle components.
- Bold and underlined numbers indicate the variables selected for household categorization.

The components 4, 5, and 6 are strongly explained by the dependency ratio H_{depend} (b = 0.935), percentage income from livestock $H_{\%in\text{-}livestock}$ (b = 0.818) and percentage income from paddy rice $H_{\%paddy\text{-}rice}$ (b = -0.921), respectively. Thus, these components are named after the original variables that represent them best (see Table 4.4).

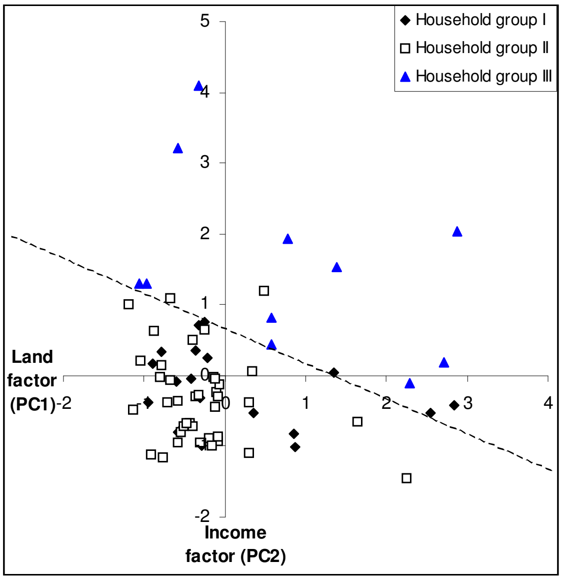
Table 4.5 Descriptive statistics for 6 key categorizing variables for each classified agent group

Categorizing variable	Agent group		$\overline{X}$	σ_x	$S.e_x$	95% confidence interval for mean		X_{min}	X_{max}
						Lower bound	Upper bound		
H_{labor}	I	19	2.7	1.2	0.3	2.1	3.2	0.5	5.5
	II	39	2.8	1.1	0.2	2.4	3.2	1.5	5.5
	III	11	2.3	1.0	0.3	1.7	3.0	0.5	4.0
H_{depend}	I	19	1.27	0.95	0.22	0.81	1.73	0.00	3.67
	II	39	1.31	0.71	0.11	1.08	1.54	0.20	3.00
	III	11	1.26	0.69	0.21	0.80	1.72	0.60	3.00
$H_{holding/pers}$	I	19	3149	3348	768	1535	4763	500	13175
	II	39	1622	2029	325	964	2280	167	11200
	III	11	5418	5000	1508	2059	8777	383	15450
$H_{income/pers}$	I	19	1358	834	191	956	1760	496	3715
	II	39	1159	627	100	955	1362	381	2858
	III	11	4000	2222	670	2507	5493	1872	9529
$H_{\%in\text{-}livestock}$	I	19	4.2	7.0	1.6	0.8	7.6	0.0	27.2
	II	39	11.6	10.4	1.7	8.2	14.9	0.0	35.9
	III	11	4.8	5.8	1.7	1.0	8.7	0.7	15.1
$H_{\%in\text{-}paddy}$	I	19	25.6	6.8	1.5	22.4	28.9	13.5	38.5
	II	39	5.5	5.7	0.9	3.6	7.3	0.0	21.0
	III	11	3.5	3.7	1.1	1.0	5.9	0.0	10.4

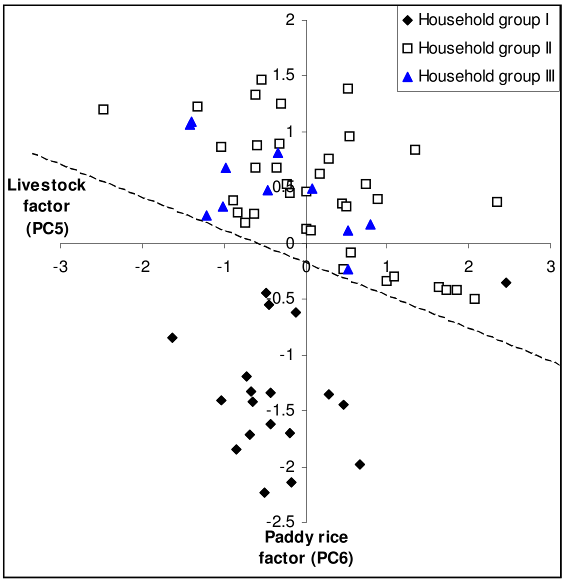
Note: *N*: group size (i.e.,number of households for each group), $\overline{X}$ *:* Mean value of the variable X, σ_x: standard deviation of the mean, $s.e_X$: standard error of the mean, X_{min}: minimal value of the variable *X*, X_{max}: maximal value of *X*

The K-CA run - using standardized scores of the six principle components - with $k = 3$ resulted in three household agent groups I, II, and III with group sizes 19, 39 and 11, respectively. The descriptive statistics of the key variables for each agent group are given in Table 4.5.

The two scatter diagrams in Figure 4.4 illustrate that the four factors *land* (PC1), *income* (PC2), *livestock* (PC5), and *paddy rice* (PC6) are enough to separate the three agent groups. Plotting samples along the *land factor* (PC1) and the *income factor* (PC2) distinguished agent group III from the two other groups (Figure 4.4a). Moreover, ordination of sampled households against the *livestock factor* (PC5) and the *paddy rice factor* (PC6) discriminates household group I from the two others (see Figure 4.4b).



(a)



(b)

Figure 4.4 (a) The land factor (PC1) and income factor (PC2) discriminate household group III from the two other groups
(b) The livestock factor (PC5) and paddy rice factor (PC6) distinguish household group I from the two other groups

Livelihood typologies of household agents

Based on classification results and descriptive statistics for the three identified household groups we found three household typologies of different livelihood structure and production strategies (see Figure 4.5). T-statistic tests show that there is no significant difference in labor availability and dependency ratio among the three household groups.

Household type I: The "paddy rice-based farmers"

The radar diagrams of standardised scores of basic livelihood indicators show that this group consists of households with limited land holdings and low income (Figure 4.5a1). Each household of this type has from 1,535 to 4,763 m^2 $person^{-1}$, and an annual total revenue about from 0.96 to 1.76 million VND $person^{-1}$ (see Table 4.5). The group constituted about 28 % of the population.

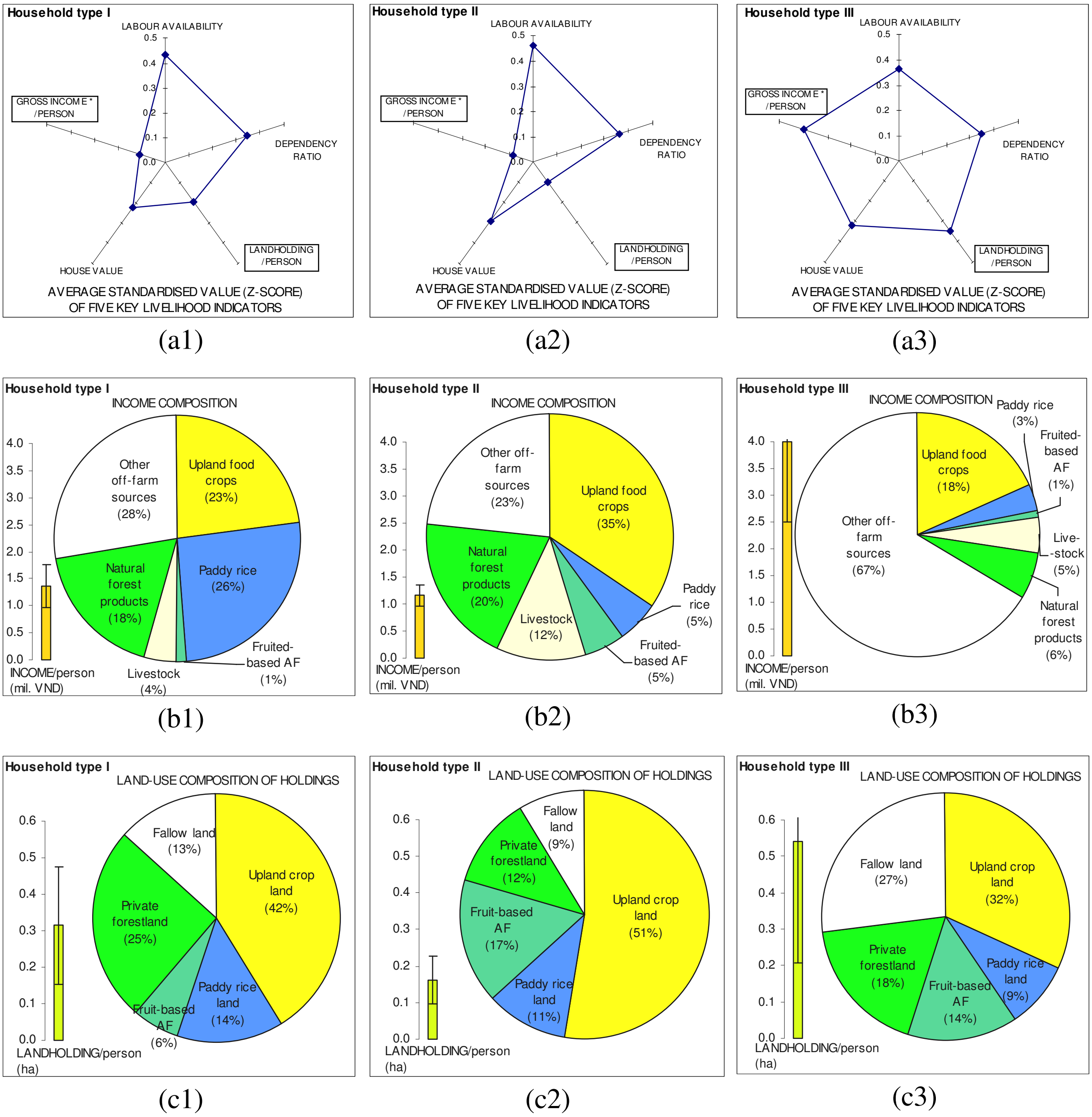


Figure 4.5 (a1), (a2), (a3): Key indicators of livelihood dimensions of three household types
(b1), (b2), (b3): Income composition, reflecting different livelihood strategies of three household types
(c1), (c2), (c3): Land-use composition of three household types

The main factor that differentiates this household group from the two others is paddy rice production. Income from paddy rice production provides about 26 % of their annual income, which significantly higher than in the other groups (see Figure 4.5-b1). Upland crop production and collection of forest products are still important components in livelihood structures, comprising about 23 % and 18 % of the annual income, respectively. A further analysis shows that the main source of off-farm income of this group is the monthly social subsidy from the government, providing about 25 % of the annual income. Both fruit-based agroforestry and livestock production are only a small proportion of the total income.

Household type II: The "upland crop and livestock farmer"

This group comprises the poorest households having the lowest amount of land and income (see Figure 4.5-b2). Each household holds about 1000 – 2,800 m^2 $person^{-1}$, and has an annual income of about 0.96 – 1.36 million VND $person^{-1}$. The group constitutes about 57 % of the total population.

Livestock production is the main factor that differentiates this group from group I (Figure 4.5b2). Percentage income from livestock, although still smaller than income from crops, is significantly higher than in the two other groups ($p < 0.05$). Income proportion from upland crops is dominant, comprising about 35 % of the annual income. Collection of forest products and social subsidies are about 20 % and 14 % of the total income, respectively; thus they still play a quite important role in the household livelihood.

Household type III: The "off-farm and better-off farmers"

Households of this group are richer than others in terms of both land resources and income. Each household of this type holds from 2,000 to 9,000 m^2 $person^{-1}$, and an annual income from 2.5 to 5.5 million VND $person^{-1}$. Other livelihood indicators of this group are not significantly different from those of the two other groups. Households of this type comprise about 16 % of the total population.

Income composition indicates that the livelihood of this household type is much less dependent on land and forest resources than the others, as their non-agriculture/forest proportion of income is dominant, ranging from 42 % to 81 %. Further analysis of the off-farm income fraction showed most of the off-farm income is from salaries for public services or social welfare. The household members are local administrators, village teachers, or agricultural extensionists who receive governmental salaries, or families receiving a monthly allowance from the government for their contribution in the war for national reunification.

Having a high income and being rich in land, households of this type are not under pressure to shorten fallow periods to meet their daily food demand, which explains why fallowed land occupies such a high proportion in their land-use composition (see Figure 4.5c1).

4.4.2 Modeling land-use choices for each typological household agent group

Factor affecting land-use choices of "*paddy rice-based farmers*" (household type I)

The results of the M-logit analysis of land-use choice for household type I are summarized in Table 4.6. The effect coefficients were estimated with respect to the private forest plantation, i.e., the base case. Thus, the inference from the estimated coefficients for each choice category is also made with reference to the base case.

The chi-square test shows that the empirical M-logit model is highly significant (p <0.01) in explaining land-use choice by farmers of the group. The Nagelkerke's pseudo-R^2 (also sometimes called livelihood ratio index) of 0.759 means that 75.9 % of the total variation in the probability of land-use choice is explained by the selected explanatory variables. The model also has good predictive power, where the choice of upland crop, paddy rice, fruit-based agroforestry and private forest plantation is correctly predicted for 82 %, 63 %, 85 %, and 55 % of the sample, respectively.

Table 4.6 M-logit estimations of land-use choices by paddy rice-based households (household type I) (n = 100 plots), using private forest plantation as a base case

Variable	*Upland crop*	*Paddy rice*	*Fruit-based agroforestry*
(constant)	-1.135	2.221	-25.887*
	(8.066)	(8.480)	(14.463)
<u>*Environmental attributes of land plots*</u>:			
Distance to roads (P_{road})	-0.004	0.002	-0.018**

	(0.004)	(0.005)	(0.008)
Distance to house (P_{house})	0.003**	0.003*	-0.008**
	(0.001)	(0.001)	(0.004)
Distance to rivers/streams (P_{river})	0.003	-0.019*	-0.012
	(0.008)	(0.011)	(0.014)
Slope angle (P_{slope})	-0.229**	-0.488***	-0.120
	(0.097)	(0.170)	(0.164)
Wetness index (P_{wet})	-0.197	-0.118	-0.276 (1)
	(0.157)	(0.163)	(0.206)
<u>*Household characteristics*</u>:			
Age of the household head (H_{age})	-0.032	0.028	0.061
	(0.064)	(0.069)	(0.114)
Leadership (0/1) (H_{leader})	-4.240*	-1.727	-5.716*
	(2.876)	(3.148)	(3.640)
Education status (0/1) (H_{edu})	0.399	-0.105	13.081**
	(3.165)	(3.428)	(6.653)
Labor availability (H_{labor})	1.734	1.154	4.671**
	(1.529)	(1.601)	(2.009)
Dependency ratio (H_{depend})	2.233	1.681	4.222*
	(1.960)	(1.995)	(2.503)
Landholding/person ($H_{holding/pers}$)	0.000	0.000	0.001*
	(0.000)	(0.000)	(0.001)
Gross income/person ($H_{income/pers}$)	0.001	0.001	-0.004*
	(0.002)	(0.002)	(0.002)
<u>*Policy variables*</u>:			
Accessibility to extension services	-1.401	-0.195	-1.027
(H_{extens})	(2.727)	(2.874)	(3.623)
Fertilizer subsidy ($H_{subsidy}$)	0.173***	0.172***	0.173***
	(0.008)	(0.008)	(0.000)

<u>*Fitness and accuracy assessment of the model*</u>:

Likelihood ratio test (chi-square statistics): 124.532*** $df = 42$ $p = 0.000$

Pseudo $R^2 = 0.759$ (Nagelkerke); 0.692 (Cox and Snell); 0.486 (McFadden)

Percentage correct predictions:		
	Upland crops:	81.6 %
	Paddy rice:	63.0 %
	Fruit-based agroforestry:	84.6 %
	Private forest plantation:	54.5 %
	Overall percentage:	74.0 %

Notes: Numbers in parenthesis are standard errors of the estimated preference parameters.
***, **, and * indicate statistical significance at the 0.01, 0.05, and 0.1 level, respectively.
(1) p-value of the estimation is barely higher than the cut-off 0.1 level (0.1-0.2)

Variables significantly influencing the decision to grow upland crops include P_{house}(+), P_{slope}(-), H_{leader}(-), and $H_{subsidy}$(+). Fields of upland crops are located farther from the house of the plot owner than plots of other land-use. This reflects the fact that most farmer houses are settled in the valley bottom and next to the main road, as a result of the Resettlement Program during the 1970s, while areas favourable for swidden farming are near forest margins. Upland crops are more likely selected for plots with less steep slopes, which coincides with the hypothesized effect. The probability that upland crops are chosen decreased if the household had a leader. The reason may be that the households are quite sensitive to the governmental policies that aim to reduce swidden agriculture to protect the forests and sedentarize farming activities. Households receiving more agrochemical subsidies more likely select upland crops. This implies that subsidizing fertilizer and pesticides encourages not only sedentary farming types as intended by the subsidizers, but also the swidden practice concurrently.

Variables significantly affecting the choice of paddy rice of this household group are P_{house}(+), P_{river}(-), P_{slope}(-), and $H_{subsidy}$(+). The paddy rice is more likely chosen for plots farther from the plot owner's house, closer to rivers/streams, and with less steep slopes. As irrigation water is vital for paddy rice production, paddy rice fields need to be located near to water sources rather than to the farmer house. Notice that the choice of paddy rice by rice-based farmers is mainly shaped by the spatial pattern of river/stream net and the terrain, and not tied to any household characteristic. This supports the fact that paddy rice is widely practiced and quite an important component in the land-use and income structures of this household type. Agrochemical subsidy also creates more incentives for households of this group to grow paddy rice, as expected by subsidizers.

Variables significantly influencing the selection of fruit-based agroforestry are P_{road}(-), P_{house}(-), H_{leader}(-), H_{edu}(+), H_{labor}(+), H_{depend}(+), $H_{holding/pers}$(+), $H_{income/pers}$(-), and $H_{subsidy}$(+). Fruit-based agroforestry is more likely selected for plots closer to roads and the farm house. Fruit-based farms located near to roads give the farmers better access to the markets to sell their fruit products (mainly bananas and pineapples). Moreover, if the fruit-based farms are close to the house, households can utilize daily some of the free time of their workers, and even children and elders, for taking care of and tending their fruit gardens.

In contrast to the anticipated effects, households with a leadership position likely leave fruit-based agroforestry. This "*surprising*" effect is explained when looking more closely at differences in non-cropping production between farmers with and without leadership. *T*-statistic tests[9] showed households with leadership have a significantly higher income fraction from fishponds ($p < 0.05$) than households without

[9] These (extra) statistical test results are available from the author.

leadership, while there is no significant difference in income percentage on agroforestry farms between households with and without leadership. This suggests that farmers with leadership have better access to fishponds, a choice not included in the choice model, than to fruit-based agroforestry.

Educated farmers seem to adopt fruit-based agro-forestry more easily, probably because of their better understanding of the potential benefits of this land use. Households with more labor and land holdings are more likely to select fruit-based agroforestry.

The positive effect of H_{depend} on the choice of fruit-based agroforesty seems to be a surprising result. However, this effect relates to the fact that the household's dependants (i.e., children and old members) are often mobilised to serve as extra workers on fruit-based plots located near their house.

Farmers who hold more land are more likely to select fruit-based agroforestry. This is understandable as the larger holding area normally gives the farmers some chance to practice farming systems returning benefits in the longer term, while annual croplands are still maintained at a reasonable level to satisfy in part their present food demands.

The negative effect of $H_{income/per}$ on the choice of fruit-based agroforestry is contrary to the hypothesised effect. The remarkable correlation[10] between $H_{income/per}$ and income percentage from off-farm activities ($r = 0.446$, $p < 0.05$) probably explains the negative effect of $H_{income/pers}$ on agroforestry choice. Households with more income are more likely to shift to off-farm activities, as off-farm jobs often return higher labor efficiency, or comparable to that obtained with farming activities. Better access to fertilizer subsidy also increases significantly the incentive of farmers to choose fruit-based agro-forestry.

Factors affecting land-use choices of *"upland crop and livestock farmers"* (household type II)

A similar M-logit regression was done for the group of household type II and the results are summarized in the Table 4.7. The likelihood ratio test showed the empirical choice

[10] Pair correlation analysis between $H_{income/per}$ and every income components showed the Pearson coefficient of $H_{income/pers}$ and income percentage from off-farm activities is positively highest. This (extra) statistic analysis result is available from the author.

model is highly significant ($p < 0.01$). The test for goodness-of-fit showed the model has a good fit with respect to the empirical dataset, with Nagelkerke's pseudo-R^2 of 0.676. The model also has good power of prediction, as the choice of upland crop, paddy rice, fruit-based agroforestry and private forest plantation is correctly predicted for 84 %, 41 %, 83 % and 73 %, respectively.

The set of variables significantly affecting the land-use choice of the farmers of this type is different from those of farmers of type I. With regard to the choice of upland crop, significantly affecting factors are P_{house}(+), P_{river}(-), H_{age}(-), $H_{holding/pers}$(+) and $H_{subsidy}$(+). Differing from the household type I, upland crop is more likely chosen for plots closer to streams/rivers. This relationship relates to the fact that swidden cultivators usually associate their swidden farming with fishing activity, and also need water for drinking and cooking as they often stay some days on the fields[11]. The probability that upland crop is chosen decreases if the farmer is older. Significantly positive coefficient of $H_{subsidy}$ indicates that the subsidy of agrochemicals increases the likelihood for selecting upland crops, including swidden fields.

Variables that significantly influence the decision to select paddy rice are P_{river}(-), P_{slope}(+), H_{age}(-), $H_{holding/pers}$(+), and $H_{subsidy}$(+). Obviously, paddy rice is more likely selected for plots closer to rivers/stream and less steep slopes. Preference to be near to rivers/streams is explained by the need of irrigation for the paddy fields. Directions of the effect of other variables are the same as in the case of the upland crop choice. The significance ($p < 0.1$) of the intercept and the low percentage of correct prediction for paddy rice choice (41 % only) mean that the selected independent variables may not be enough to explain correctly the choices. However, the limitation of our sample size does not permit including more explanatory variables in the model.

Table 4.7 M-logit estimation of land-use choices by household type II (upland crop and livestock farmers) (n = 165 plots), using private forest plantation 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)
<u>*Environmental attributes of land plots*</u>:			
Distance to roads (P_{road})	0.000	0.002	-0.006
	(0.004)	(0.004)	(0.006)

[11] Information acquired through field visit and interviews.

Distance to house (P_{house})	0.001*	0.001	-0.004***
	(0.001)	(0.001)	(0.001)
Distance to rivers/streams (P_{river})	-0.023***	-0.028***	-0.004
	(0.009)	(0.010)	(0.011)
Slope angle (P_{slope})	0.039	-0.233*	0.050
	(0.113)	(0.142)	(0.135)
Wetness index (P_{wet})	0.180	0.084	0.241 (1)
	(0.157)	(0.160)	(0.171)
Household characteristics:			
Age of the household head (H_{age})	-0.079*	-0.098**	-0.041
	(0.044)	(0.045)	(0.049)
Leadership (0/1) (H_{leader})	-2.385	-1.313	-1.855
	(1.963)	(2.033)	(2.072)
Education status (0/1) (H_{edu})	-2.012 (1)	-1.872	-2.925*
	(1.540)	(1.588)	(1.757)
Labor availability (H_{labor})	0.692	0.639	-0.671
	(0.635)	(0.655)	(0.757)
Dependency ratio (H_{depend})	1.994	1.674	3.323*
	(1.700)	(1.726)	(1.783)
Landholding/person ($H_{holding/pers}$)	0.000**	-0.001**	0.000
	(0.000)	(0.000)	(0.000)
Gross income/person ($H_{income/pers}$)	-0.000	-0.000	-0.002*
	(0.001)	(0.001)	(0.001)
Policy variables:			
Accessibility to extension services (H_{extens})	0.166	-0.437	0.645
	(1.540)	(1.596)	(1.679)
Fertilizer subsidy ($H_{subsidy}$)	0.171***	0.171***	0.171***
	(0.002)	(0.003)	(0.000)

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 predictions:		
	Upland crops:	84.1 %
	Paddy rice:	40.5 %
	Fruit-based agroforestry:	82.8 %
	Private forest plantation:	72.7 %
	Overall percentage:	73.3 %

Notes: Numbers in parenthesis are standard errors of the estimated preference parameters.
***,**, and * indicate statistical significance at the 0.01, 0.05, and 0.1 level, respectively.
(1) The p-value of estimation of education variable is barely higher than the cut-off 0.1 level (0.1 - 0.2)

Variables that significantly affect the decision to select fruit-based agroforestry are P_{house}(-), H_{edu}(-), H_{depend}(+), $H_{income/per}$(+), and $H_{subsidy}$(+). Coinciding with the anticipated direction, distance from the farmer house has a significant and negative effect on the choice of fruit-based agroforestry. The more education, the more likely the households leave fruit-based agroforestry. This effect is explained when looking more closely at differences in non-cropping production between educated and non-educated farmers. T-tests showed educated farmers have a significantly higher income fraction from cattle ($p < 0.1$) and from fish ponds ($p < 0.05$) than non-educated farmers. This means the more they are educated, the more farmers this group concentrate on livestock components, not so on crop production.

The effect of $H_{income/per}$ on fruit-based agroforestry choice also seems not to fit common sense. However, the remarkable[12] correlation between $H_{income/pers}$ and income percentage from cattle ($r = 0.374$, $p < 0.05$) explains partly that households with more income are more likely to concentrate on cattle raising.

Factors affecting land-use choices of *"off-farm and better-off farmers"* (household type III)

Because the sample size of this household group is rather small (i.e., 61 plots nested with 11 households), the regression could not be run with a full set of explanatory variables as in the regression analysis for the agent groups I and II. Therefore, the three dummy independent variables H_{edu}, H_{leader}, and H_{extens} were not included in the regression, as these variables do not highly vary among households of this group (Table 4.8).

A similar M-logit regression was done as for the household type II (i.e., the poor and swidden cultivator) and the results are summarized in Table 4.8. Likelihood ratio test indicated that the empirical choice model is highly significant (p <0.01). Test for goodness-of-fit showed the model has a good fit to the empirical dataset, with Nagelkerke's pseudo-R^2 of 0.684. The model also has fairly good prediction power, as the choice of upland crop, paddy rice, fruit-based agroforestry and private forest plantation is correctly predicted for 89 %, 43 %, 50 % and 57 %, respectively.

Table 4.8 M-logit estimates of land-use choices by household type III (off-farm and better-off farmers) (n = 61 plots), using private forest plantation as a base case

Variable	Upland crop	Paddy rice	Fruit-based agroforestry
(constant)	-15.138 (39.276)	-39.587 (40.213)	-38.240 (39.967)
Environmental attributes of land plots:			
Distance to roads (P_{road})	-0.010 (1)	-0.004	-0.046***

[12] Pair correlation analysis between $H_{income/per}$ and every income component showed the Pearson coefficient of $H_{income/pers}$ and income percentage from cattle is positively highest.

	(0.008)	(0.009)	(0.017)
Distance to house (P_{house})	0.001	-0.001	0.001
	(0.001)	(0.002)	(0.001)
Distance to rivers/streams (P_{river})	-0.001	0.006	0.019
	(0.013)	(0.014)	(0.017)
Slope angle (P_{slope})	-0.319*	-0.422**	-0.038*
	(0.170)	(0.182)	(0.216)
Wetness index (P_{wet})	0.275	0.269	0.374
	(0.382)	(0.388)	(0.394)
<u>*Household characteristics*</u>:			
Age of the household head (H_{age})	-0.038	0.067	0.110
	(0.210)	(0.215)	(0.213)
Labor availability (H_{labor})	5.749	9.119	7.456
	(7.272)	(7.377)	(7.395)
Dependency ratio (H_{depend})	7.864	12.811*	8.847
	(8.651)	(8.804)	(8.845)
Landholding/person ($H_{holding/pers}$)	0.000	0.001**	0.001**
	(0.000)	(0.000)	(0.001)
Gross income/person ($H_{income/pers}$)	-0.002	-0.001	-0.001
	(0.002)	(0.002)	(0.002)
<u>*Policy variables*</u>:			
Fertilizer subsidy ($H_{subsidy}$)	-0.009	-0.016 [(1)]	0.000
	(0.012)	(0.013)	(0.014)

<u>*Fitness and accuracy assessment of the model*</u>:

Likelihood ratio test (chi-square statistics): 93.625*** $df = 33$ $p = 0.002$

Pseudo R^2 = 0.684 (Nagelkerke); 0.629 (Cox and Snell); 0.393 (McFadden)

Percentage correct predictions:		
	Upland crops:	89.3 %
	Paddy rice:	42.9 %
	Fruit-based agroforestry:	50.0 %
	Private forest plantation:	57.1 %
	Overall percentage:	67.3 %

Notes: Numbers in parenthesis are standard errors of the estimated preference parameters.
***,**, and * indicate statistical significance at the 1%, 5 %, and 10 % level, respectively.
[(1)] p-value of the estimation is barely higher than the cut-off 0.1 level (0.1-0.2)

The choice of upland crop by this household type is significantly influenced by slope only (p <0.1). This means the choice of upland crop in this case is less “*selective*”, or more “easy” than the choices of the other household groups. These choice characteristics may explain partly why households of this type have large areas of upland crop fields.

Variables significantly affecting the choice of paddy field are P_{slope}(-), H_{depend}(+), and $H_{holding/pers}$(+). Paddy farming is more likely chosen for plots with less steep slopes. Households with a high dependency ratio seem to concentrate on paddy rice, which has two crops a year, to meet their monthly food demand. Also, the likelihood for selecting paddy rice increases significantly if the farmer holds more land. This also implies that newly claimed plots are more likely to be used for paddy rice.

Variables affecting the choice of fruit-based agroforestry are P_{road}(-), P_{slope}(-), $H_{holding/pers}$(+). The proximity to roads and land slope have significant and negative effects on the choice of fruit-based agroforestry. Farmers who hold more land are more likely to select fruit-based agroforestry. Through holding a large area of land, farmers of this type normally have a greater chance to practice farming systems returning benefits in the longer term, while annual croplands are still maintained at a moderate level to satisfy their actual food demands.

4.5 Conclusions

Key variables differentiating household livelihood categories in Hong Ha communities have been objectively identified. Statistic inferences, i.e., PCA and k-CA, of household data revealed six endogenous factors that differentiate livelihood typologies of farming households in Hong Ha, namely: land holding per capita, annual income per capita, labor availability, labor dependency, livestock and paddy rice productions. These findings also agree with particular previous studies on processes of household differentiation in the Vietnam uplands (Castella *et al*., 2002; Fatoux *et al*., 2002; Tran Duc Vien *et al*., 2001; Gomiero and Giapietro, 2001; Castella and Erout, 2002; Alther *et al*., 2002). As these key variables explain most variations of household livelihoods, they should be used as criteria for regularly categorizing households in the VN-LUDAS (i.e., used for the *AgentCategorizer* routine in the HOUSEHOLD-POPULATION module). Moreover, since these key variables are highly independent of each other, using them in further computations, such as in randomised generations of new household agents, will help avoid co-variation problems. With respect to the development work in the study area, these variables are also helpful in designing community surveys and identifying target groups of households.

Classification using extracted principle components resulted in the three livelihood typologies of households in the study area, namely: the "*paddy rice-based farmers*" (household type I), the "*upland crop and livestock farmers*" (household type II), and the "*off-farm and better off farmers*" (household type III). Further land-use adoption analyses for each household type show different household types have different patterns of land-use choices/adoptions. The differences in land-use choices among the three household groups are observed in terms of three aspects: the one affecting the direction of land-use choice determinants, the one affecting the magnitude of these choice determinants, and most importantly, the set of choice determinants. These findings clearly show considerable heterogeneities in land-use choice behavior in the studied community, and rigogously parameterized these heterogeneities. In general, households of all groups choose land use based on the mutual considerations of a range of personal characteristics, natural conditions of the environment, and particular policy factors. Therefore, the developed model of land-use choice is one of the bases for coupling the human-environment systems under particular policy circumstances when simulating land-use changes.

When applying these land-use choice analysis results to VN-LUDAS in the study area, both the estimated preference coefficients (β) and the standard errors of the estimated (σ_{β}) groups are used for the computation of land-use choice probabilities (P_{hij}) in the *FarmlandChoice* procedure of the DECISION program. Each household agent will adopt random values of preference coefficients around the group's coefficients, bounded by the standard errors. Therefore, the land-use choice behaviors of households fluctuate within a behavior template if they are in the same group, but are structurally different if they are in different livelihood groups.

With respect to the development work in the study area, the findings regarding the determinants of land-use choices provide a better understanding of the reality of land-use adoption processes, which is valuable for many development activities in the uplands. The results suggest that determinants of land-use choice not only are site conditions, household's tangible resources and land-use related policy factors, but also social factors such as education status and political powers. The direction and likelihood of the effects of these factors can be different among specific groups and may not always agree with the conventional beliefs of outsiders.

5 ECOLOGICAL DYNAMICS OF HETEROGENEOUS LANDSCAPE AGENTS: THE CASE OF HONG HA WATERSHED, CENTRAL VIETNAM UPLANDS

5.1 Introduction

Complex processes of land-use and land-cover change (LUCC) arise from not only the diversity of human decision-making, but also from the heterogeneous dynamics of the landscape environment (Parker *et al.*, 2003). Different locations of the landscape normally have different conditions of climate, soil, topography, and subsequently different land-use capability. This spatial heterogeneity often changes over time due to either the impacts of past land-use activities, or natural processes that are in ways beyond human control. The combination of such spatio-temporal dynamics and the diversity of human behavior (see Chapter 4) often drive the complex and non-linear dynamics of LUCC process (Verburg *et al.*, 2002; Parker *et al.*, 2003). With the increasing awareness that bio-complexities can be represented by MAS-LUCC models, modeling of spatio-temporal dynamics has gained importance for better understanding on how human-environment system evolves.

Recent developments in MAS-LUCC modeling have also created new requirements and challenges to represent the realism of the dynamic environment. According to agent-based design, a natural landscape is represented in the form of a grid of congruent land patches that are autonomous landscape agents. Landscape agents are attributed with internal state variables storing heterogeneous spatial data, and equipped with ecological sub-models, which employ the emerging internal state, possibly the neighborhood state, and flexible inputs/intervention from human agents (Box, 2002; Chapter 2 and 3 of this thesis). This agent-based presentation of the landscape thus treats landscape dynamics as a self-organized phenomenon, which evolves from micro-autonomous processes. However, putting this agent-based design into simulations with reflections of landscape realism has seen little progress. Although the biophysical environment is highly dynamic due to its nature (Wooldridge, 2002), many recent MAS-LUCC models are still based on the static environment that is assumed to remain unchanged except when affected by the agent's actions (e.g., Westervelt and Hopkins, 1999; Lintenberg *et al.*, 2001; Loibl and Toetzer, 2003). Moreover, many current MAS-

LUCC models are still only able to simulate very simplified and hypothetical landscapes (Kanaroglou and Scott, 2002; Parker *et al*., 2003). Thus, empirical parameterisation of landscape agents for representing more realistically the landscape dynamics is gaining importance in MAS-LUCC research.

To do so, the first task is the collection of real biophysical data for the initialization of the state of landscape agents. This means that the landscape environment has to be characterised in a spatial explicit way, e.g., in forms of GIS raster layers, using real data. In order to obtain a relevant understanding, the landscape characterisation should focus on essential characteristics of relevant ecological processes, e.g., those that indicate common environmental concerns and/or are main drivers of human decision-making regarding resource use.

The second task, which is the most challenging, is the development and calibration of *ecological sub-models* for each landscape agent typology, for instant productivity functions of patches with a particular land-cover type. Because ecological dynamics of the patch are the combined results of both heterogeneous natural constraints/potentials (e.g., erosion risk and natural vegetation growth) and interventions of human agents (e.g., crop management practices and logging activities), the ecological sub-models should include both the natural and human drivers, i.e., they are rather bio-economic than purely bio-physical models. In the context of human-ecological system modeling, the priority should be given to formulate and approve ecological processes that play important roles in building human-environment relationships.

In forest margins, the dynamics of agricultural and forest yields are widely recognised as the most immediate ecological responses that constitute the human-environment interaction loop (Vanclay, 1997; Haggith *et al*., 2003). Through generating benefits for rural inhabitants, agricultural and forest yields are important drivers of land-use decision-making (Haggith *et al*., 2003; Park and Vlek, 2003; Bousquet *et al*., 2000; Vanclay, 1994). Forest yield also is a parameter indicating directly the modification of the vegetative covers, and thus the health of the terrestrial environment. Moreover, because it is the combined outcome of bio-physical conditions and actor-specific interventions, yield function is one of the ways of coupling the landscape dynamics with human dynamics. Therefore, modeling the dynamics of agricultural and forest yields becomes relevant to the representation of the complexity of the LUCC process.

This study aims at specifying more details about the landscape component, i.e., the PATCH-LANDSCAPE module, its framework developed within the VN-LUDAS model (see Chapter 3), and calibrating it using the empirical spatial data gathered from our study site. The major assumption of this study is that different landscape patches have different potential productivities (i.e., agricultural or forest yields) in response to natural conditions and human interventions. Based on this assumption, the chapter has four interrelated specific objectives:

i) to characterize variables that have ecological and economic relevance to site productivity and land-use decision-making processes, including the categorization of landscape agents (i.e., patches) into functional types in term of land covers,
ii) to formulate and calibrate agricultural yield sub-models for agricultural landscape agents (i.e., cultivated patches),
iii) to formulate and calibrate forest yield sub-models for forested landscape agents (i.e., forested patches), and
iv) based on the achievements of the three objectives above, to formulate and calibrate rules of natural transitions among land-cover types that in part generate the spatio-temporal pattern of LUCC.

Moreover, this case study is expected to illustrate in particular relevant quantitative methods to bring the realism of environmental dynamics (in coupling with the human system) into the MAS.

5.2 Bio-physical setting of the study area

5.2.1 Climate

The study area is within the zone of the tropical monsoon climate with the two distinct seasons: the hot-dry season from April to August, and the cool-rainy season from September to March (see Figure 5.1). Average annual temperature is nearly 25^0C. The hottest months are June, July, and August with monthly mean temperatures of nearly 29^0C. The coolest period is from December to February with average temperature of about 19 – 20^0C. Annual rainfall ranges from 2700 to 5600 mm, with an average about

3,000 – 3,300 mm, but distributed unevenly over the months. About 80 % of the total rainfall falls during the first four months of the rainy season (September – December). The wettest months are October and November with a monthly rainfall of about 750 – 1400 mm, often associated with tropical storms and subsequent flash floods (Figure 5.1).

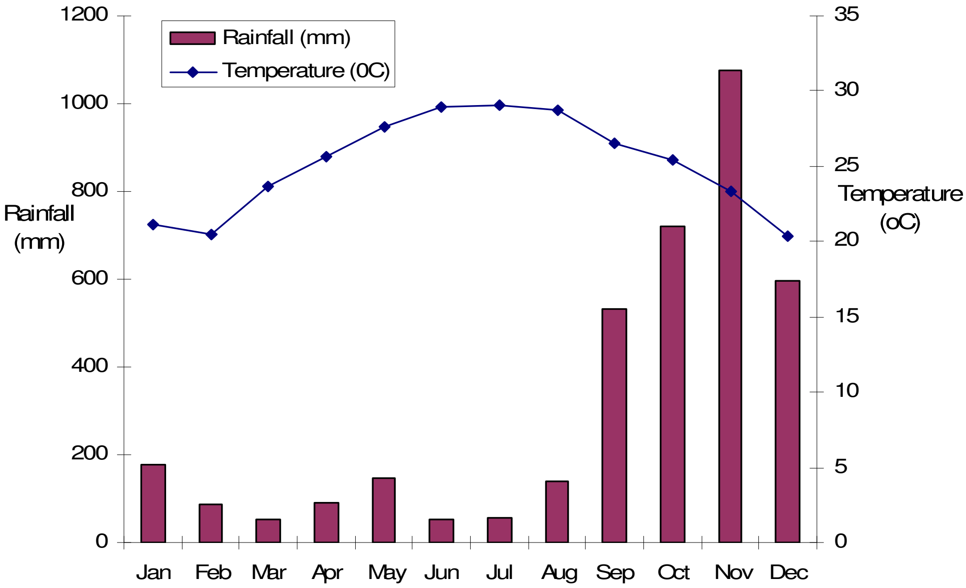


Figure 5.1 Monthly distribution of rainfall and temperature. Average data of five years (1996 - 2000) recorded at Hue Station. Data source: Thua Thien – Hue Statistical Office (2001)

5.2.2 Soils

There are four soil types, according to Vietnam's Soil Classification System[13] (see Nguyen Ngoc Binh, 1996), found in the study area, namely: yellowish-red ferralitie soil developed on argillaceous or metamorphic rocks (F_s), yellowish-red ferralitie soil developed on acid magma (F_a), reddish-yellow humus ferralitie soil on acid magma (H_a), and alluvial soil along streams/rivers (P_b) (Figure 5.2). The F_s soil covers most of the study area (67 % total area), occurring in both northern and southern sides of Rao Nho Stream. The F_a soil occupies relatively large zones on the north side of the Rao Nho Stream and the east side of the Bo River, covering about 24 % of the total area.

[13] The soil classification system of Vietnam, which were formulated based on the soil genesis processes prevailing throughout the country, has been applied widely since 1978. Although the Vietnamese Soil Scientists Association have recently shifted that system to the soil classification and mapping system of FAO-UNESCO, the former system has still used commonly for studies of soils developed on mountainous and hillside landscapes (Do Dinh Sam and Nguyen Ngoc Binh, 2000).

The humus ferralitie soil (H_a) is found only in the summit area of Dong A Xom (c. 900-1200 m). Alluvial soils (P_b) are apparent in tiny trips mainly along the Rao Nho River.

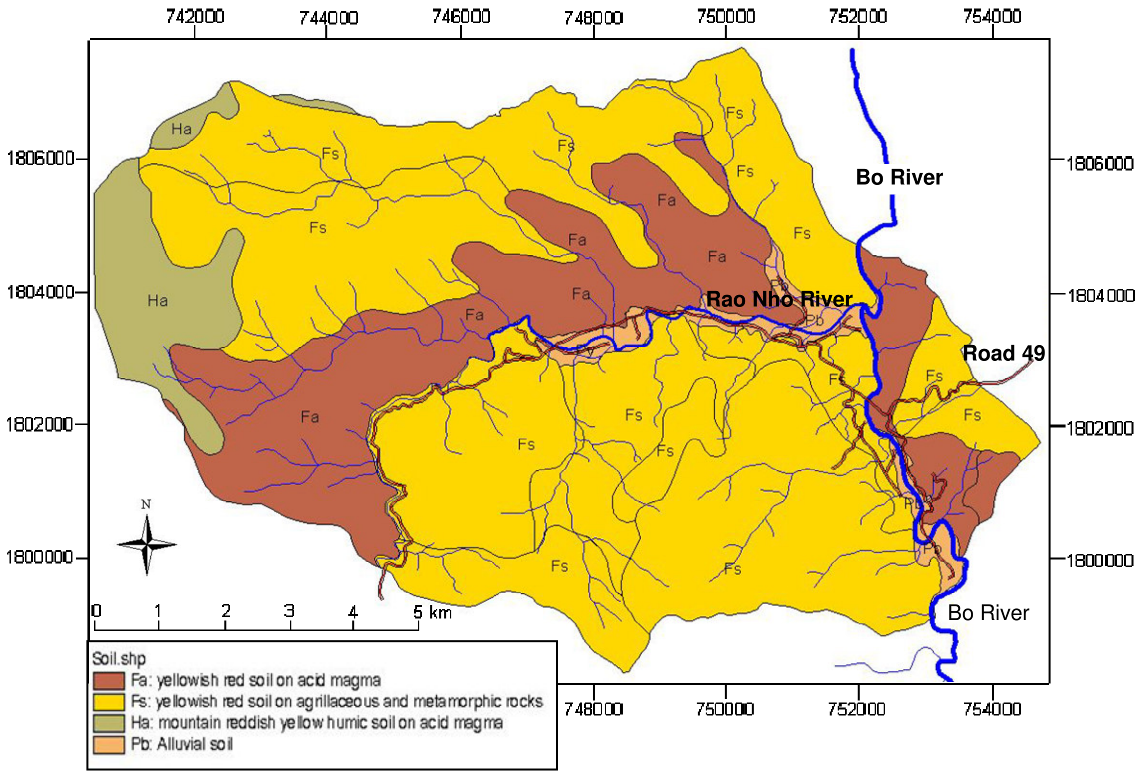


Figure 5.2 Map showing major soil types in the study area. Note: Soil types are according to Vietnam Soil Classification System. Map coordinate system: UTM, Zone 48 North, Datum WGS84. Data source: Soil map scale 1:100,000 of the Department of Agriculture and Rural Development, Thua Thien – Hue province

Characteristics of these soil types on Vietnam's mountains and hillsides are described in detail in Nguyen Ngoc Binh (1996), and Do Dinh Sam and Nguyen Ngoc Binh (2000). Based on these studies, the most basic features of these soils are as follows. F_s soils have a texture varying from clay to loam, and normally show a medium productivity. F_a soils are sandy loam, with low contents of humus (about 2 – 4 %) and nitrogen (about 0.05 – 0.10 %), thus less fertile. However, the productivity of both F_s and F_a soils are further reduced remarkably if the forest cover is converted to create non-forested land (e.g., shrubs, tall grasses or swidden fields). The soil type H_a has a high humus content (i.e., 8-10 % or above). Changes in the productivity of the H_a soil between forested and non-forested sites are not so apparent as in the case of F_s and F_a

soils, because of high humidity and low evaporation in the high mountain zone (above 900 m a.s.l).

5.2.3 Vegetation

As it is situated within the Northern Vietnam Coastal Moist Forests and the Annamite Range (i.e., Truong Son) Moist Forests Ecoregions (Wikramanayake *et al.* 1997), the vegetation formation in the study area is tropical moist forest. Under complex interventions of human beings - such as wartime spraying with defoliants, logging and the expansion of swidden agriculture in the post-war periods - the primary tropical moist forest, which formerly covered most of the area, has been modified or converted to other secondary vegetation types, typically: mature secondary forest, immature secondary forest, regenerating forest, shrubs and grasses (Le Trong Trai *et al*., 2001; Vu Van Dzung, 2002). Consequently, the dense forest cover has been reduced, degraded and fragmented. However, the remaining fragments of disturbed primary forest and mature secondary forest in the area are still relatively large (about 50 – 60 % of the total area). Species compositions are still highly similar between forests of different succession stages (Le Trong Trai *et al*., 2001).

The human-introduced vegetation cover in the area, i.e., *Acacia* forest plantations and agricultural types, was briefly described in Chapter 4. In general, these man-made cover types are distributed along the Rao Nho and Bo rivers.

5.3 Methodology

5.3.1 Methods of landscape characterization

Terrain analyses for characterizing soil-water conditions

Topography as a proxy predictor for soil-water landscape variations

Being one of the major determinants of the ecosystem's primary productivity, the inclusion of spatial variation of soil/water status is essential for ecological models on the landscape scale (Park and Vlek, 2002). However, considering the fact that soil/water occurrence is highly variable over space and time (Park *et al*., 2001; Park and Van de Giesen, 2004), it is greatly challenging to capture efficiently both spatial and temporal heterogeneities of such complex phenomena. The more accurate and detailed the soil map is to be determined, the less precise the temporal dynamics of such spatial details is to be predicted. Therefore, according to the agent-based modeling approach, if a phenomenon intended to be modeled is complex, it still needs to be modeled with respect to its constituent drivers (see Bonabeau, 2002). This implies that the soil/water phenomena should be represented in terms of their primary drivers, i.e., environmental variables that play decisive roles in soil-forming processes. The driving variables must have a strong, inherent and proven role in regulating soil/water-landscape patterns, and be more easily determined.

It is clearly recognized that the topography inherently regulates water flows and redistributes soil materials in both the short and long term, thereby strongly determining landscape patterns of soil/water conditions (Gessler *et al*., 2000; Wilson and Gallant, 2000). The theoretical basis for understanding the decisive role of landform in soil development on hillslopes lies at the heart of the catena concept, which was initiated by Milne (1935), cf. Park *et al.* (2001), and conceptually reformulated by Conacher and Dalrymple (1977). The catena concept has been frequently used as the qualitative framework to guide modern soil surveys or landscape process monitoring and interpretations (Conacher and Dalrymple, 1977; Schoeneberger, 1998). Mathematical reinterpretation of the qualitative model of catenary soil development has been well justified by previous studies (e.g., Carlson and Kirkby 1972, Moore *et al*., 1993; Park *et al*., 2001).

Recently, empirical realities of strongly causal relationships between topography and landscape patterns of soil/water have been proven through rigorous spatial correlations of field-measured soil/water properties against terrain parameters. Soil properties that proved to respond strongly to terrain factors are soil depth (e.g., Gessler *et al*., 2000; Park *et al*., 2001), soil moisture content (Westtern *et al*., 1999; Park and Van de Giesen, 2004), soil carbon content (Gessler *et al*., 2000; Creed *et al*., 2002), soil total and potentially mineralizable nitrogen contents (Creed *et al*., 2002), soil pH and total exchangeable bases (Park and Vlek, 2002). These soil attributes are obviously important indicators of soil productivity for crop production or natural vegetation growth. In general, terrain factors are most useful for the prediction of soil attributes whose spatial distributions are strongly influenced by lateral hydrological and slope processes with relatively simple depth functions (Park and Vlek, 2002).

Selection of primary terrain indices: upslope contributing area and slope gradient

Key primary terrain indices for soil/water-landscape characterization were selected by reinterpreting the well-known continuity equation of Carson and Kirkby (1972). The following argument is after Park *et al*. (2001). In a transport-limited slope, the mean rate of soil material transport (*S*) by diffusive slope process (i.e., slope wash, soil creep, and subsurface leaching of weathering products) can be substituted with the transportation capacity (T_h) of soil at a given location (Carson and Kirkby, 1972):

$$S = T_h = kf(x)^m g(x)^l \tag{5.1}$$

where $f(x)$ is the distance of position x from the hilltop, $g(x)$ is the change of elevation with the change of position, and k, m, and l are constants. The function $f(x)$ refers to water flow factor, which is possibly replaced by the upslope contributing area per unit of contour length (P_{As}) (Carlson and Kirkby, 1972; Moore and Wilson, 1992). The function $g(x)$ refers to the slope shape factor influencing transportation processes, which is usually approximated by slope gradient (P_{slope}) (Park *et al*., 2001; Park and van de Giesen, 2004).

Upslope contributing area (P_{As}) is defined as the total catchment area above a point on the landscape. For a grid cell i of a DEM, P_{As} is computed from the grid cells from which the water flows into the cell i:

$$P_{As} = (1/b)\sum_{i=1}^{n} \rho_i . A_i$$

where A_i is the area of the grid cell i, n is the number of cells draining into the cell i, ρ_i is the weight depending on the runoff generation mechanism, and b is the contour width approximated by the cell size (Park *et al*., 2001). Therefore, P_{As} theoretically determines runoff volume, steady-state runoff rate, and water flow accumulation at a landscape position (Willson and Gallant, 2000). Recent empirical studies proved a strongly positive relationship between P_{As} and soil depth – both total and *A* horizon thicknesses - (Gessler *at al*., 2000; Park *et al*., 2001), soil moisture content (Park and Van de Giesen, 2004), and soil organic content (Gessler *et al*., 2000). Therefore, it is supported that P_{As}

indicates the accumulation potential of soil and water, thereby positively affecting soil productivity at a site.

The slope gradient (P_{slope}) determines the kinetic energy of the water flow, i.e., the velocity of overland flow and subsurface flow, and runoff rate (Wilson and Gallant, 2000; Pallaris, 2000), thus creating an overall physical force of soil erosion. The role of slope in limiting the overall productivity of a site is impressive, as this parameter has been traditionally used for zoning the landscape capabilities of land uses. Thus, we use P_{slope} for indicating soil degradation potential of a site.

Therefore, the coupling of P_{As} with P_{slope} in particular ways can approximate soil/water landscape variability in modeling ecological dynamics of landscape agents, such as the dynamics of crop yields. From an equilibrium viewpoint, the actual productivity of a site is more or less a balance between the accumulation potential, which is represented by P_{As}, and the degradation risk, which is approximated by P_{slope}.

Topographical wetness index

We realize that there are two common methods for coupling P_{slope} and P_{As} in modeling ecological responses to soil-water conditions. The first coupling method is the use of compound terrain indices, such as sets of wetness or power stream indices (Moore *et al.*, 1993; Wilson and Gallant, 2000). These compound indices are fixed or consistent in terms of their equation form and coefficients for every application. The second coupling method comprises empirical functions, which are estimated based on local-specific dataset. In contrast to the compound index method, this empirical coupling is local specific in terms of the function forms and coefficients. Both coupling methods are used in VN-LUDAS for different purposes. Wetness index is used as a variable of land-use choice models, to reduce the number of explanatory variables for improving the robustness of the choice model. The empirical coupling method is used for modeling crop yield responses, as described in a later section (see Section 5.3.2).

Wetness index (P_{wet}) is a compound terrain index that has been used extensively to approximately delineate the spatial pattern of soil moisture content that is important in agricultural production (De Roo, 1998; Wilson and Gallant, 2000). The index is calculated based on upslope contributing area and slope gradient as follows:

$$P_{wet} = \ln(P_{As} / \tan P_{slope})$$

Computational methods

We computed the grids [P_{slope}], [P_{As}], and [P_{wet}] from a Digital Elevation Model (DEM), using the grid-based algorithm developed by Zevenbergen and Thorne (1987). The DEM has a spatial resolution of 30 m × 30 m, which was interpolated from a digitized UTM map scale 1:50,000, using the TOPOGRID routine in ARC/INFO 8.0. The selected spatial resolution is relatively fit to the scale of the original maps, the average size of landholding parcels (see Chapter 4), and the resolution of land cover data extracted from Landsat ETM images.

Spatial Accessibility Analyses

Spatial accessibility can be defined as the ease with which a target location may be reached from another location (Goodall, 1987). Because reaching a target location is a precondition for the satisfaction of almost any need there, accessibility is often a key variable determining land-use choices (Angelsen and Kaimowitz, 1999; Nelson, 2000). More than just roads, spatial accessibility has social, economic and environmental dimensions, which all can be seen to be important in the development processes (Nelson, 2000; Burrow and Nelson, 2001), including land use and management. In this chapter, we only focus on characterizing the economic and environmental aspects of spatial accessibility. The social aspects of spatial accessibility will be later characterized in Chapter 6, as linked to institutional and policy issues (i.e., customary boundaries and zoning policies).

From an economic viewpoint, access to transportation is a critical function for an economy, as it affects the movement of goods, and people approaching markets, schools and information needed (see Burrow and Nelson, 2001). Better access to roads will reduce the transaction costs in agriculture and forestry, e.g., the costs of moving products from used lands to markets/home and vice versa for input materials. This often supports the choice for fruits or cash crops (e.g., see Chapter 4 and Fox *et al*., 1994), or facilitating forest exploitation (Liu *et al*., 1993; Komitz and Gray, 1996; Cropper *et al*., 1997; Angelsen and Kaimowitz, 1999). To measure accessibility to roads, we used approximate distance from land patches to roads (P_{road}).

Access to water bodies, represented by the approximate distance from the land parcel to rivers/streams (P_{river}), can influence the choice of land use (see Fox *et al.*, 1994) in different ways. Paddy fields are normally located near rivers/streams as the paddy rice needs to be irrigated. Upland fields may be more likely chosen in plots near stream/rivers, but possibly associated with other livelihoods on rivers/streams (e.g., fishing) or domestic uses (e.g., drinking, cooking), rather than for irrigating crops (see Chapter 4).

We calculated the grids [P_{road}] and [P_{river}] using the *Find Distance* routine of the *Spatial Analysis* module in ARCVIEW GIS 3.2 package. The road network built before 1999 was digitized directly from geo-referenced aerial photographs at the original scale 1:33,350, taken on June 17, 1999. Roads built after 1999 were digitized from a false composite of a geo-referenced ASTER image (15 m × 15 m resolution), taken in February 2002, and additionally mapped using data points tracked by a handheld Global Positioning System (GPS) unit. The river/stream network was digitized from the UTM topographic map (scale 1: 50,000). Small streams/creeks not shown on this map were mapped by local experts, i.e., forestry engineers of Hong Ha Station for Forest Protection and key informants from the commune, with the support of the aerial photographs and the flow accumulation map derived from the DEM.

Land cover classification

Because land cover is clearly an important variable of MAS-LUCC models, an accurate mapping of this variable is critically important for the calibration of system initialization. More important, in the context of VN-LUDAS, classification of current land cover plays a role as initial categorization of landscape agents (i.e., land patches) into ecologically functional types in terms of land-covers. This classification will create a basis for further development of yield response functions for each cover type, as well as spatial extrapolation of variables measured on limited sampling units.

Land cover can be reliably derived from data obtained through remote sensing. Automatic classification methods, which are mostly based on spectral information, are often used to extract main land cover types at regional scale in a fast and objective way. However, because some land cover classes may exhibit similar spectral properties, it is difficult to differentiate such cover classes using automatic classification algorithms alone. Therefore, automatic classifications are usually used in association with an interpretation procedure that utilizes other supporting information - e.g., ground truth data, finer-solution spatial data, thematic maps, and etc. - to develop the land-cover/use database as needed (e.g., Hafeez, 2003).

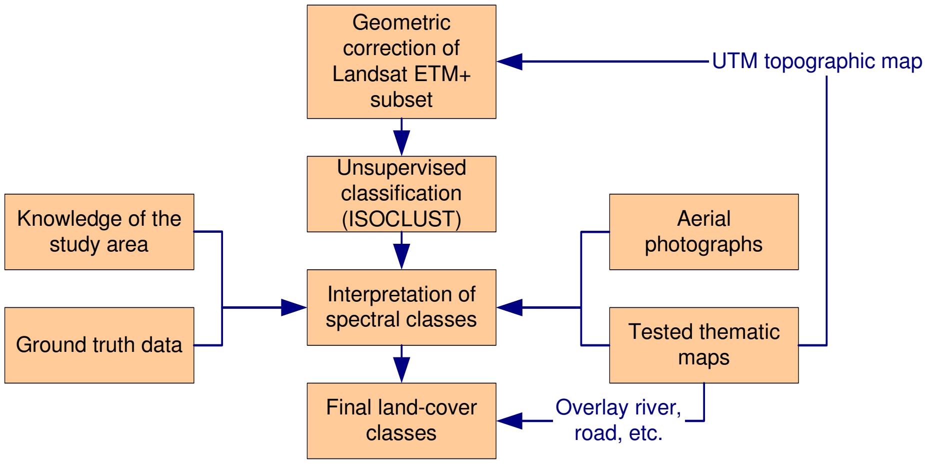


Figure 5.3 Procedure of land-use classification

The procedure of land-cover classification in this study is shown in Figure 5.3. First, we created a subset of Landsat ETM+, taken on April 2, 2002, covering the Hong Ha study area and geo-referenced the subset to the UTM topographic map, with a root mean square error (RMSE) of less than 1 pixel size. Second, we conducted unsupervised classification, using the ISOCLUST routine of IDRISI 32 (Eastmann, 2001), to extract objectively 14 spectral classes. Third, these 14 spectral classes were interpreted using ground truth data, and evidence from the aerial photographs taken in June 1999. The color map of these unknown spectral classes, overlaid with contour lines, stream/river and road networks, was used for a ground truth survey and discussions with local experts to interpret these unknown classes. The interpretation of 14 spectral classes distinguished 5 major land cover types, namely: i) dense natural forest, ii) open natural forest, iii) shrubland, iv) grassland, and v) bare land. However, the human-introduced cover classes were still mixed in with these 5 major classes.

In order to distinguish human-introduced cover classes from the 5 major classes above, we used other data sources than Landsat ETM$^+$ data. Areas of forest plantations, which are often mixed with the open natural forests, or even with shrublands, were digitized from plantation maps of the Bo River Forest Management

Division (BRFMD), i.e., the governmental organization that manages forests in the study area. Parcels of agricultural cover types, which are often mixed with the grasslands and shrublands, were drawn based on GPS point data of ground truth surveys and visualized mapping in the fields. Road areas, which are mixed with the bare land category, were extracted using the road map. River/stream areas were created based on the drainage map. When converting from shapes (lines) to grids, the main rivers take a width of 2 pixel sizes (i.e., 60 m width), while roads and streams to be a width of 1 pixel size (i.e., 30 m width).

After the land-cover types were mapped, productivity/yield functions for each cover type were developed and calibrated to represent in part the ecological dynamics of landscape agents.

5.3.2 Method for modeling agricultural yield response

Modeling approach

To model crop yield responses to given environmental conditions and management options, there are two different approaches, namely: empirical and process-based models, and each approach has its own merits and limitations (Park and Vlek, 2003). The empirical approach attempts to derive the patterns of crop yield responses from empirical datasets of driver variables using statistical analyses (i.e., regression or correlation analyses), without any postulating of biological/ecological processes underlying crop growth and development. The empirical models are relatively simple to build and develop, and often have a good predictive power within ranges of empirical data. However, purely empirical functions often do little to further understanding of the ecological processes underlying yield dynamics, and provide less capabilities for extrapolating yield beyond the data ranges (i.e.,spatial and temporal ranges) on which the model is based (Vanclay, 1994; Park and Vlek, 2003).

The process-based modeling approach attempts to model behavior of crop yield through using mathematic equations built on agro-climatic, physiological biochemical theories and to quantitatively model plant-soil-atmospheric interactions (Godwin and Vlek, 1985, cf. Park and Vlek, 2003; Mathew, 2002). Theoretically, this approach has a relatively high power of extrapolation over space and time scales (Jame and Cutforth, 1996), and gives a better explaining power by taking into account the biological/ecological mechanisms of plant growth and development (Vanclay, 1994). However, in practice the requirements of extremely intensive calibration-verification procedures at a detailed resolution limit a wider application of this model type (Sinclair and Seligman, 1996; Stephens and Middleton, 2002). Alternatively, model parameters may be set by experts or adopted from previous researches in different environments, but uncertainties will then greatly limit the predictive power and reliability of model application (Penning de Vries *et al.*, 1989; Stephens and Middleton, 2002).

In this study, we selected the empirical approach to model agricultural yield for the following reasons. First, as our modeling scale is agricultural types as a whole rather than detailed crop species/varieties, it would have been unnecessarily complicated if the process-based approach is applied. Second, because the main modeling objective is rather to anticipate agricultural yield response than to understand underlying processes of crop growth, empirical models are normally more robust (Vanclay, 1994). Third, if necessary data are available, such as the plot-specific data panel in our case, empirical models even offer a more reliable yield response than poorly calibrated process-based growth models (Park and Vlek, 2003). Fourth, there are ways to overcome the limitations of the empirical approach. The careful choice of explanatory variables across a wide range of potential yield drivers can help formulating yield functions behaving in an ecologically realistic way (Vanclay, 1994). Spatial and temporal dynamics of agricultural yield can be represented by including spatial and time variables into yield functions.

An empirical model for predicting yields of agricultural types (the *AgriculturalYieldDynamics* sub-model)

Defining response variable: yield of agricultural land-use type

The response variable is the yield of an agricultural land-use type, which is one among the three identified agricultural land-use types (see Table 5.1), not explicit for each crop. Since each agricultural type can include more than one crop, harvested crop products are converted to an equivalent amount of rice, then the yield unit is kg of rice per hectare per annum (i.e., kg rice ha^{-1} $year^{-1}$). Because the crop products can include vegetables and/or pepper that not suitable for the caloric converting method, we used monetery converting method with the local price in the summer 2002 as the base price.

Selection of driver variables

Regardless of genetic factors, the agricultural yield of a plot ($P_{a\text{-}yield}$) can conceptually be a function of climate conditions (*CL*), soil/water conditions – or site productivity - (*SW*), land management practices (*M*) and time (*t*):

$$P_{a\text{-}yield} = f(CL, SW, M, t) \tag{5.2}$$

Because of the relatively small size of the study area (about 90 km^2) and the narrow elevation range of the potential cultivation area (from 50 to 300 m a.s.l), it is reasonable to assume that the climate factor *CL* is uniform over the study area. Because time-series data on agricultural yield were lacking, the climate factor was also assumed to be stable during the simulation period.

The soil-water factor *SW* of the patches can be approximated by slope gradient (P_{slope}) and upslope contributing area (P_{As}), as justified above. Thus, site productivity is approximated in the model following a geocentric view (Leary, 1985). In modeling agricultural yield responses, we used these two primary indices through empirical coupling rather than through a single compound terrain index. The use of a single compound index would not explain the effect of component primary variables. Moreover, previous studies have shown that a compound terrain index alone, such as the wetness index, does not always give a good representation of soil-water patterns (e.g., Western *et al*., 1999; Gessler *et al*., 2000; Park *et al*., 2001; Park and Van de Giesen, 2004). In contrast, the empirical coupling method is more flexible in terms of functional forms and the coefficient of each primary index, and thus should have a better ability to explain the crop response. Also, the development of the model on a data-fitting basis may help improve the prediction accuracy. Because of the correlation with soil erosion risks, the slope gradient (P_{slope}) is expected to significantly inhibit agricultural yield. Because it reflects soil/water accumulation potentials, a greater upslope contributing area (P_{As}) is anticipated to support a higher agricultural yield.

Among land management factors (M), labor (I_{labor}) and agrochemicals (i.e., NPK fertilizer and pesticides) (I_{chem}) inputs should be the prior variables for consideration because these resources are used directly for cultivation. It is widely recognized that crop yield increases if farmers apply more fertilizers, pesticides, and spend more time managing their croplands. However, the sensitivity of crop yield to increments of agrochemical and labor inputs (i.e., agrochemical and labor efficiency) may be different between agricultural types, depending on the nature of each land-use type and actual natural conditions. The instant values of I_{labor} and I_{chem} are determined by household agents, whose behaviors are governed by the DECISION module, and affected by the policy factor (fertilizer subsidy).

The time factor t is represented by current cropping time of the plots (P_t) (usually $P_t \leq$ cropping period). The fluctuation of crop yield during the cropping period can occur in different directions, depending on the nature of the cropping systems. Yield of the annual upland crop fields, which are a type of swidden cultivation, is anticipated to decrease with the *decline of soil fertility* over cropping years. In contrast, the yield of fruit-based agroforestry farms principally increases according to the *growth of the fruit tree/crop components*, and thus is expected to correlate positively with P_t. Although P_t is a plot variable, its values are determined by human agents.

Brief definitions of the variables selected for agriculture yield modeling are listed in Table 5.1. By replacing these specified variables, equation 5.2 becomes:

$$P_{a\text{-}yield} = f\,(P_{slope}, P_{As}, I_{labor}, I_{chem.}, P_t) \tag{5.3}$$

Selection of function form: power function

The power function is the most frequently used production function in empirical work (Wallenius, 2004). The function makes it easy to express different non-linear patterns and to quantify the elasticity of yield response. Let us consider the simplest form of power function: $Y = aX^{\beta}$ (a>0), where Y is the response yield, X is an explanatory variable, a and β are coefficients. The power coefficient β is very meaningful for interpretation of the behavior of the yield response Y to the explanatory variable X in a qualitative and quantitative manner. Qualitatively, simple mathematics show that the response pattern of yield (Y) is flexible, depending on β (Figure 5.4). The coefficient β shows not only the direction of the yield change, but also the acceleration behavior of the yield increment. A limitation of modeling yield using the power function is that the function becomes undefined when β is not an integer and the explanatory variable X receives negative values (Sit and Poulin-Costello, 1994). Thus, the power function may not be suitable for including explanatory variables that can be negative, such as surface curvature indices.

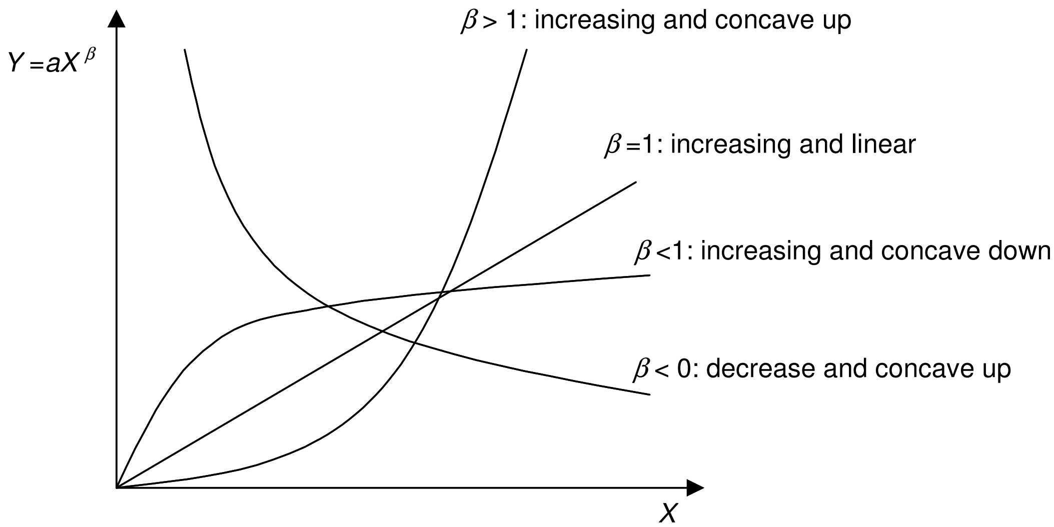


Figure 5.4 Power function showing different behavior of response variable *Y* according to β. Source: Sit and Poulin-Costello (1994)

Quantitatively, the power β measures directly the elasticity of agricultural yield *Y* to the change in its predictor *X*. By definition, elasticity is the percent change in the dependent variable as a result of a 1 % change in the explanatory variable (Franklin, 1999; Wallinus, 2004). Based on that definition, the simple mathematic justification for yield elasticity measurement in the power function is as follows:

$$Yield\ elasticity = \frac{\%\Delta Y}{\%\Delta X} = \frac{\Delta Y / Y}{\Delta X / X} = (\frac{dY}{dX}) \cdot \frac{X}{Y} = (\beta a X^{\beta - 1}) \cdot \frac{X}{Y} = (\frac{\beta Y}{X}) \cdot \frac{X}{Y} = \beta \qquad (5.4)$$

We used the yield elasticity coefficient β for comparing the magnitude of the effects of different explanatory variables, which have different units of measurement.

The simple power function can be readily extended to a multivariable power function (Wallenius, 2004). Given the five selected explanatory variables above, agricultural yield in the conceptual equation 5.3 can be expressed in the power function of the form:

$$P_{a-yield} = a.I_{chem}^{\beta_1}.I_{labor}^{\beta_2}.P_{slope}^{\beta_3}.P_{As}^{\beta_4}.P_{t}^{\beta_5} \qquad (5.5)$$

where a is a constant; β_1, β_2, β_3, β_4, and β_5 are yield elasticities to agrochemical input (I_{chem}), labor input (I_{labor}), patch slope (P_{slope}), patch's upslope contributing area (P_{As}) and cropping time (P_t). This extended power function still allows investigating the effect of any explanatory variable while holding all other variables constant.

Function 5.5 is a relevant representation of spatial bio-complexity, since it is a non-linear combination among variables of the natural landscape and human interventions. If the landscape and time variables are assumed to be constant, equation 5.5 becomes the well-known Cobb-Douglas production function (see Tochombe, 2002).

Multiple log-linear regression analysis

One of advantages of the power function is that the non-linear relationship can be easily transformed into a log-linear form for implementing simple statistical estimation (Wallenius, 2004). By taking logarithms for both sides of equation 5.5, we have a log-linear function that can be easily estimated using multiple linear regressions (equation 5.6).

$$lnP_{a\text{-}yield} = lna + \beta_1 lnP_{slope} + \beta_2 lnP_{As} + \beta_3 lnI_{labor} + \beta_2 lnI_{chem} + \beta_5 lnP_t \qquad (5.6)$$

Care must be taken when performing any computation that uses equations 5.5 and 5.6. The power function will be zero, or the log-linear function will be undefined, if one of the explanatory variables is zero. This is really a problem because the fact that P_{slope} = 0 (i.e., absolutely flat land) or I_{chem} = 0 (i.e., farmers do not apply any agrochemicals) does not necessarily mean the crop yield has to be the zero. To fix this problem, we change zero value of explanatory variables to 1 in the computation processes. This change does not affect the computed results.

The explanatory capabilities of the models were validated by the F-statistic test for the model as a whole, and by T-statistic tests for each explanatory variable. The goodness-of-fit of the regression models was measured by the standard error of the estimate(s) and the coefficient of the determination (R^2) (Retherford and Choe, 1993; Maddala, 1992). To achieve an acceptable capability for prediction with a multiple linear regression model, the standard error of the estimate should be less than 10% of the mean value of prediction (Gupta, 1999). The higher the value of R^2, the better the fit of data.

Table 5.1. Variables used for the *AgriculturalYieldDynamics* sub-model

Variable	*Definition*	*Data sources*	*Direct linked module*
Yield responses (dependent variables)			
$P_{y\text{-}paddy}$	Yield of paddy rice plot (*kg rice ha^{-1} year^{-1}*)	Interviewing plot-owners (n= 73 plots)	Patch-landscape
$P_{y\text{-}upcrop}$	Yield of annual upland crop plot (*kg rice ha^{-1} year^{-1}*)	Interviewing plot-owners (n= 134 plots)	Patch-landscape
$P_{y\text{-}af}$	Yield of fruit-based agroforestry plot (*kg rice ha^{-1} year^{-1}*)	Interviewing plot-owners (n= 47 plots)	Patch-landscape
Natural predictors			
P_{slope}	Slope angle of the plots (*degree*)	Field measurement	Patch-landscape
P_{as}	Unit upslope contributing area of the plots (*m^2/m*).	Dem-driven	Patch-landscape
Management predictors			
I_{chem}	Monetary value of agrochemical input (mainly NPK fertilizers + pesticide) on the plot (*1000VND ha^{-1} year^{-1}*)	Interviewing plot-owners	Decision and global-policy
I_{labor}	Man days used for production activities on the plots (*day ha^{-1} year^{-1}*)	Interviewing plot-owners	Decision and global-policy
Temporal factor			
P_t	Continuously cultivating time-length of the plot (*year*)	Interviewing plot-owners	Patch-landscape and decision

Representing uncertainty of yield prediction: random-bounded yield functions

When the calibrated agricultural yield models are applied in the VN-LUDAS model, uncertainties of the predictions are also accounted for. Uncertainty of yield prediction can be subjective, and due to either the limitation of the conceptual model, or the limited size of dataset, or errors in data collection/conversion. Crop yield uncertainty can also be objective, because many regulating factors may occur stochastically by nature, e.g., droughts, incidence of plant diseases, etc. Therefore, uncertainty cannot be avoided in the predictions of crop yields, and is inherent. If we expect to reflect realism in the model, uncertainty should be represented in association with the deterministic element in the model.

To do so in the VN-LUDAS model, crop yield functions are expressed in the form of random-bounded functions. Given an agricultural land-use decision for a particular patch, the prediction of the *instant* agricultural yield is computed as follows:

$$predicted\ lnP_{a\text{-}yield} \in [lnP_{a\text{-}yield} - CI_{0.05},\ lnP_{a\text{-}yield} + CI_{0.05})]$$

or

$$predicted\ lnP_{a\text{-}yield} = lnP_{a\text{-}yield} - CI_{0.05} + random(2CI_{0.05}) \qquad (5.7)$$

where $lnP_{a\text{-}yield}$ is the deterministic log-yield estimated by equation 5.6, $CI_{0.05}$ is the confidence interval at 95% of the estimated log-yield, and $random(2CI_{0.05})$ generates a random number within the bounds $[0,\ 2CI_{0.05}]$ following a uniform distribution. The $CI_{0.05}$ is calculated as: $CI_{0.05} = t_{0.25} \times s = 1.96 \times s$, where s is standard error of the estimate.

In case of good estimations of the yield, e.g., high R^2 and low s, the uncertainty range becomes rather narrow and the predicted yield is more deterministic. Otherwise, we have rather highly stochastic predictions of crop yield.

Data sources for parameter estimation

The parameters in equation 5.6 were estimated using a plot-based data panel acquired through an intensive household survey in summer 2002 (see Chapter 4). Although the survey investigated a total of 367 plots (belonging to 69 households selected randomly) for different study purposes, full data records that permit yield modeling are available for 254 plots only. For every plot, we asked plot owners about the amounts of products harvested, agrochemicals (i.e., NPK fertilizers and pesticides) and labor used during the agricultural year 2002/03, cropping time and other related information. Yield and agrochemical data were, respectively, converted to rice and currency, using local price units in summer 2003. Slope angles of the plots (P_{slope}) were measured directly in the field. Since all plots were geo-referenced, upslope contributing area of the plots (P_{As}) was extracted from the P_{As} grid, which was calculated from the DEM (see Section 5.3.1).

The fully parameterized yield model of the three agricultural land-use types in Hong Ha was named *AgriculturalYieldDynamics* routine, which is used as a sub-model of the PATCH-LANDSCAPE component of the VN-LUDAS model. The summary of variables of the *AgriculturalYieldDynamics* routine is shown in Table 5.1.

5.3.3 Method to specify forest yield functions

Selection of forest growth modeling approach

Similar to crop yield modeling, empirical and process-based approaches are mutually exclusive in forest growth modeling (Mäkelä *et al*., 2000). Empirical models have a long tradition in forest growth prediction, but they require time-series data obtained from permanent sample plots (Vanclay, 1994). Process-based models derive the growth of a forest stand based on the underlying physiological processes (e.g., photosynthesis and respiration) (Johnsen *et al*., 2001; Mäkelä *et al*., 2000; Bartelink, 2000), and components of stand dynamics (i.e., size increment, mortality and recruitment), thus taking single trees or parts of trees as basic modeling units. Here, the lack of field-measured data on forest growth does not permit developing a purely empirical forest growth model. Moreover, since our modeling unit for forest growth is at stand level as a whole, it does not make good sense to apply sophisticated process-based models.

Alternatively, we specified the forest yield function at stand level using a theoretical approach. This selection was based on our conviction that a "*theoretical guess*" remains a good choice when data are lacking. Moreover, theory-based equations may be more reliable for predictions that involve extrapolations beyond the range of empirical data (Vanclay, 1994). We formulated the function of forest yield response based on basic concepts of forest growth and succession, principles of biological system theory (Von Bertalanffy, 1942, 1968), and other reasoned assumptions. The equation parameters should be characterized using common values revealed in forest sciences, rather than empirical growth data, which were not available.

A theoretical model of forest yield dynamics (the *ForestYieldDynamics* sub-model)

Selection of forest yield variable: Basal area of forest stand

We selected stand basal area (P_G) to quantify the yield/stock/growth of a forest stand, because of the following advantages. Stand basal area (P_G) is the sum of the individual tree basal area (G_i), which is the basic parameter for calculating tree volume, biomass and crown. Because G_i is calculated directly from tree diameter d_i (or girth) using a straightforward geometric formula ($G_i = \pi d_i^2/4$ [cm^2]), basal area may be less ambiguous than volume (Alder, 2000). From an ecological viewpoint, P_G indicates not only the forest yield, but also the density of a forest stand, which is strongly correlated with the competition status that is important for the growth of forest trees (Brack, 2004). In forestry practice, the amount of timber logged is often expressed in terms of basal area.

It is important to note the relationships between the concepts of instant forest growth, yield (cumulative growth), and residue stock. *Instant growth* of a forest stand (Z_G) refers to the natural net increment of the forest stand size per time unit, at a particular time t (i.e., $Z_G = dP_G/dt$). *Natural yield* of a forest stand (${}^tP_{Gn}$) is the total cumulative size of the stand at the time t, without any removals, also called *natural cumulative growth*. *Residual forest stock* (${}^tP_{Gr}$) is the natural forest yield from which are subtracted the removals ($G_{removals}$), if any (i.e., ${}^tP_{Gr} = {}^tP_{Gn} - G_{removals}$). The yield function can be expressed either by the integration of the growth function along elapsed time (i.e., ${}^tP_{Gr} = \int {}^tZ_G dt$), or by the previous residual stock (${}^{t-1}P_{Gr}$) plus the instant growth rate (i.e., ${}^tP_{Gr} = {}^{t-1}P_{Gr} + {}^{t-1}Z_G$). The latter expression of yield is more convenient for the computing algorithm as well as for incorporating impacts of human activities (Vanclay, 1994). Thus the relationship of these concepts can be numerically expressed as follows:

$$ {}^tP_{Gr} = ({}^{t-1}P_{Gr} + {}^{t-1}Z_G) - G_{removals} \tag{5.8} $$

We used residual basal area ${}^tP_{Gr}$ as the response variable to represent forest dynamics, since the variable couples the natural growth dynamics (viz. ${}^tZ_G$) with the impacts of human interventions (viz. $G_{removals}$), thus allowing linking forest dynamics to human behavior. Moreover, as the current forest stock depends on the previous state, the dynamics of residue forest stock are accumulative and path-dependent, and depend very much on the initial state of the forest stand. Detailed developments of these three components (${}^tZ_G$, $G_{removals}$, and initial stock ${}^{2002}P_{Gr}$) are as below.

Theoretical function of stand basal area increment (Z_G)

This section begins with the well-known growth equation of an organism by Ludwig von Bertalanffy (1957), which was included in the biological system theory invented by the same scholar (von Bertalanffy, 1942, 1968). Von Bertalanffy theorized that the growth of an organism could be represented as the difference between the synthesis (anabolism) and degradation (catabolism) of its building materials. Following Pütter (1920, cf. Vanclay, 1994), von Bertelanffy assumed that the processes of anabolism and catabolism could be expressed as allometric functions of body weight (W), and thus the growth of an organism (*dW/dt*) can approximate:

$$dW/dt = \eta W^m - \kappa W^n$$

where m and n are the constants of anabolism and catabolism, respectively; η and κ are allometric constants. Von Bertalanffy's equation was latter reinterpreted and reformulated for modeling the growth of different organism types, such as plants (Richards, 1959) and fishes (Chapman, 1961).

Based on von Bertalanffy's equation, Vanclay (1994) developed a theoretical equation expressing the basal area growth of a forest stand as a whole:

$$Z_G = dP_G/dt = a(P_G)^{\varepsilon} - b(P_G) \tag{5.9}$$

where P_G is stand basal area, Z_G is instant growth rate of P_G, a and b are the constants, and ε is a coefficient of very small value ($\varepsilon \rightarrow 0$). This equation has been used as a theoretical basis for different empirical analogues of growth models for uneven-age forest stands (Vanclay, 1994). However, the determination of these three parameters in a more theoretically explicit way is still a problem. Even when empirical data are available, it is still difficult to fit the equation of this non-linear form with the data (Vanclay, 1994 and Ratkowsky, 1990).

We determined the parameters a and b in equation 5.9 using the following theoretical development. First, it is assumed that the stand growth rate Z_G is asymptotically zero in the equilibrium state. Let ${}^{equil}P_G$ be the stand basal area at the equilibrium state of the forest stand (also called *natural basal area*), according to the assumption then we have:

$$a({}^{equil}P_G)^{\varepsilon} - b({}^{equil}P_G) = 0 \tag{5.10}$$

Second, basic mathematics state that the derivative of the growth function Z_G is zero when it reaches the maximum. Let $*P_G$ be the stand basal area when Z_G is maximal (i.e., ${}^{max}Z_G$), we have:

$$(Z_G)'\Big|_{*P_G} = \varepsilon a(*P_G)^{\varepsilon-1} - b = 0 \quad \Leftrightarrow \quad *P_G = (b/a\varepsilon)^{1/(\varepsilon-1)}$$

In equation 5.9, replacing P_G by $*P_G = (b/a\varepsilon)^{1/(\varepsilon-1)}$ and Z_G by ${}^{max}Z_G$, we have:

$$^{max}Z_G = a[(b/a\varepsilon)^{1/(\varepsilon-1)}]^{\varepsilon} - b[(b/a\varepsilon)^{1/(\varepsilon-1)}] \tag{5.11}$$

Assuming that the parameters ε, ${}^{equil}P_G$ and ${}^{max}Z_G$ are known, solving the set of two equations 5.10 and 5.11 to determine the two unknowns a and b, yields:

$$a = {}^{max}Z_G\,/\,[({}^{equil}P_G)^{\varepsilon}(\varepsilon^{\varepsilon/(1-\varepsilon)} - \varepsilon^{1/(1-\varepsilon)})] \tag{5.12}$$

$$b = {}^{max}Z_G\,/\,[{}^{equil}P_G(\varepsilon^{\varepsilon/(1-\varepsilon)} - \varepsilon^{1/(1-\varepsilon)})] \tag{5.13}$$

The two parameters ${}^{equil}P_G$ and ${}^{max}Z_G$ are settable either by forestry experts or reviewing literature on tropical forest growths. From the phytocentric view, ${}^{equil}P_G$ is conceptually considered an expression of the productive capacity of the site (Havel, 1980; Vanclay, 1994). However, empirical studies on the productivity of tropical moist forests in many places around the world show that natural basal area (${}^{equil}P_G$) is one of the "*constants of tropical forest*" (Liegh, 1999: 120-122). In this study, we assume ${}^{equil}P_G$ takes a constant over space, as there is no evidence to correlate this parameter with location variables. The concrete value of ${}^{equil}P_G$ is the upper confidence limit of the mean basal area of the surveyed dense/rich natural forest plots. The value of ${}^{max}Z_G$ can be approximated from the projected outputs of empirical growth models. The constant ε can be fixed by setting a very small value (e.g., $\varepsilon = 10^{-6}$).

Defining impacts of human activities ($G_{removals}$)

Assuming that the human impact on forest quality is mainly in terms of logging, the removed basal area ($G_{removals}$) principally includes three components: harvested amount (G_{logged}), logging damage (G_{damage}) and logging-driven mortality ($G_{mortality}$):

$$G_{removals} = G_{logged} + G_{damage} + G_{mortality}/T \quad (5.14)$$

where G_{logged} is the basal area logged by human agent(s), G_{damage} is standing basal area damaged immediately by logging operation, and $G_{mortality}$ is basal area lost as tree mortality occurring over some years (T) after the logging event (see Alder, 2000).

Notice that the appearance and the intensity of logging activities are decided by human agent(s), through the functioning of the DECISION module, and influenced by forest protection zoning policy (i.e., linked to GLOBAL-POLICY module) (see Figure 5.3).

Principally, G_{damage} and $G_{mortality}$ have positive relationships with logging intensity, tree size and land slope (Vanclay, 1994). However, as far as we are aware, there has been no data about logging damage in Vietnam that can be used to directly derive empirical functions for such relationships. Therefore, we approximated G_{damage} and $G_{mortality}$ based on an empirically logging impact model of Alder and Silva (2000), developed in the Brazilian Amazon:

$$\%G_{damage} = 0.0052N_{logged} + 0.0536 \quad (R^2 = 0.8987) \quad (5.15)$$

$$\%G_{mortality} = 0.0058N_{logged} + 0.0412 \quad (R^2 = 0.9044) \quad (5.16)$$

where $\%G_{damage}$ and $\%G_{mortality}$ are percentage of standing basal area before logging that are lost due to damage and mortality following the logging activity, respectively (i.e.,$\%G_{damage} = G_{damage}/({}^{t-1}P_{\mathrm{Gr}})$ and $\%G_{mortality} = G_{mortality}/({}^{t-1}P_{\mathrm{Gr}})$); N_{logged} is the number of logged trees.

Let g_{logged} be the mean basal area of logged trees, then N_{logged} in two equations 5.15 and 5.16 can be expressed in term of G_{logged}, i.e., $N_{logged} = G_{logged}/g_{logged}$. Accordingly, through a mathematic conversion we have:

$$G_{damage} = {}^{t-1}P_{\mathrm{Gr}}\,(0.0052\, G_{logged}/g_{logged} + 0.0536) \quad (5.17)$$

$$G_{mortality} = {}^{t-1}P_{\mathrm{Gr}}\,(0.0058\, G_{logged}/g_{logged} + 0.0412) \quad (5.18)$$

Spatial extrapolation of initial forest stock from plot-measured data

Based on plot-measured data, we extrapolated the average basal area of surveyed plots over all patches with the same cover type using the empirically random-bounded rule. According to this rule, a patch *j* of forest cover type *i* will randomly receive a basal area value $P_{Gr(i)}$ within $\overline{P_{Gr\,(i)}} \pm CI_{(i)}$, where $\overline{P_{Gr\,(i)}}$ is the mean stand basal area for the forest cover type *i*, and $CI_{(i)}$ is the confidence interval at 95 % level of the mean. The mathematical expression of the extrapolation rule is:

$$P_{Gr(i)} = \overline{P_{Gr\,(i)}} + random(2CI_{(i)}) \tag{5.19}$$

where *random* ($2CI_{(i)}$) is a random function that returns a random number between 0 and $2CI_{(i)}$, following a uniform distribution. The stand basal area mean $\overline{P_{Gr\,(i)}}$ and the confidence interval $CI_{(i)}$ were derived from plot data using descriptive statistics.

In the context of VN-LUDAS, the full spatial extrapolation of the basal area for the three forest cover types found in the Hong Ha watershed were done using the following function:

$$^{2002}P_{Gr} = \begin{cases} \overline{P}_{Gr(6)} + random(2CI_{(6)}) & if \quad P_{cover} = 6 \\ \overline{P}_{Gr(7)} + random(2CI_{(7)}) & if \quad P_{cover} = 7 \\ \overline{P}_{Gr(5)} + random(2CI_{(5)}) & if \quad P_{cover} = 5 \\ 0 & if \quad P_{cover} \notin \{5,6,7\} \end{cases} \tag{5.20}$$

where 5, 6, and 7 are the cover codes of dense/rich natural forest, open/poor natural forest, and *Acacia* forest plantation (> 4 years), respectively.

The ForestYieldDynamics algorithm

Because the residual stand basal area P_{Gr} is accumulative over time, event-driven by human agents, and determined by a set of equations, its dynamics are formalized by a numerical algorithm rather than a single yield function. The above developments were the bases of our computational routine to project the dynamics of stand basal area, named *ForestYieldDynamics*. Based on equations 5.8, 5.9, 5.12, 5.13, 5.14, 5.17, 5.18, and 5.8, we constructed the pseudo-algorithm of the *ForestYieldDynamics* sub-model as in Figure 5.5.

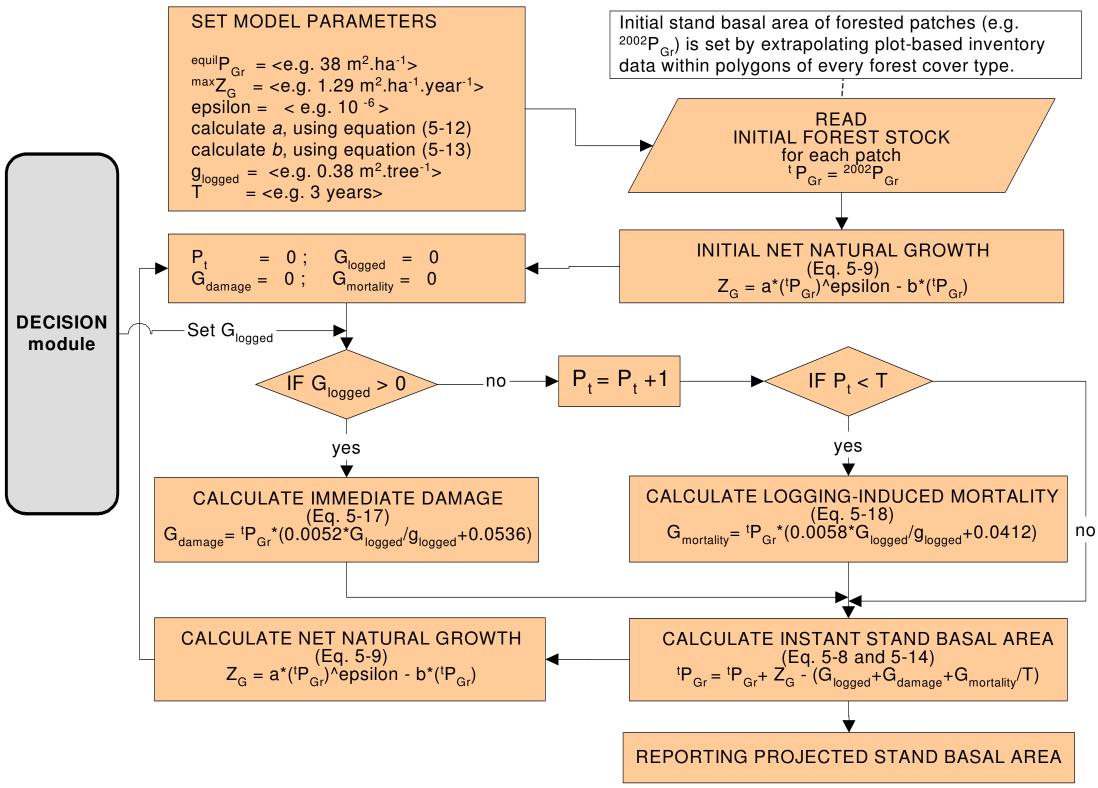


Figure 5.5 Algorithmic flowchart of *ForestYieldDynamics* sub-model. Note: P_t is years after logging, T: post-logging period with severe tree mortality due to logging impacts

Data sources and specification of parameters for *ForestYieldDynamics* routine for the study site

Extensive forest inventory

In June and July 2003, we conducted an extensive forest inventory to get data for spatial extrapolation of the initial forest basal area. After forest cover types throughout the study area were delineated, we positioned sampling plots within each forest cover type for estimating its initial residue stock. As the complex terrain conditions inhibited random sampling strategy, we only located sampling points along transects in the four representative areas of natural forests within the study watershed (see Figure 5.6). Because the forest canopy inhibited GPS signals, most sampling points were geo-referenced approximately using compass course from a few GPS-based geo-referenced points in open areas (forest gaps).

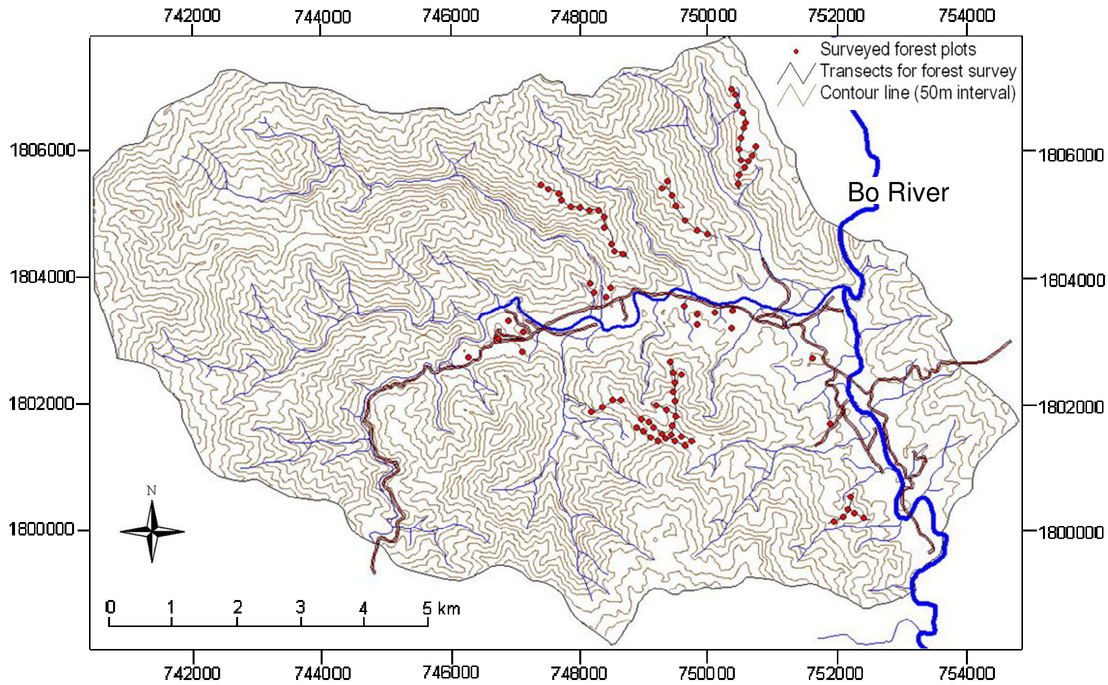


Figure 5.6 Map showing the plots for forest survey in June and July 2003. Map coordinate system: UTM, Zone 48 North, Datum WGS84

At each sampling point, one sampling unit was set for measuring current forest yield[14]. The sampling unit was a circular plot of 100 m^2 (5.64 m radius). We chose circular plots because they are easy to lay out in complex terrain conditions, they have a low edge error, and allow computing directly density/yield and rapid survey (Le Quang Bao, 1998). Within the circular plots, all trees with a girth at breast height (*gbh* [cm]) ≥ 31.4 cm (i.e., diameter breast height (*dbh*) ≥ 10cm) were measured for *gbh*, using a measuring tape. Plot basal area P_G [m^2 ha^{-1}] is the sum of individual basal areas, using the formula: $P_G = \sum(gbh^2/4\pi).10^{-2}$, where π = 3.1416. There were a total of 34 and 28 plots located within dense/rich and open/poor natural forests, respectively.

[14] At each sampling point, beside a circular plot of 100m^2 for studying tree community, we also set four quadrats of 4m^2 for studying tree seedling community. However, the floristic and tree seedling data was not used with this study.

For forest plantations, we used the dataset of the Faculty of Forestry (FOF) at Hue University of Agriculture and Forestry (2002) for estimating the current stock of such a forest cover type. The dataset was gathered from 16 rectangular plots of 1000m^2 (i.e., 20 m $\times$ 50 m), which were located in the zone of the *Acacia* plantation of the Hong Ha commune, in October 2002. The stand basal area was calculated in the same way as for the natural forests.

Specification of input parameters for the ForestYieldDynamics sub-model

A summary of the parameters of the *ForestYieldDynamics* routine is shown in Table 5.2. For natural forest, the two parameters $^{equil}P_G$ and $^{max}Z_G$ were confidentially set by reviewing literature on the growth of tropical moist forests. Dawkin (1959), cf. Leigh (1999), and Leigh (1999) reported that most of tropical forests have a stand basal area (at 1.3 m above ground of all trees $\geq$ 10 cm) around 30 m^2 ha^{-1}. However, this report does not refer to any specific sub-type of tropical forest or to whether the forest is in an equilibrium state. Previous studies on tropical moist forests in Costa Rica (Alder, 1996a), Brazil (Alder, 1996b) and Papua New Guinea (Alder, 1998) show that the stand basal area of moist tropical forests at the equilibrium state $^{equil}P_G$ is around 35 - 36 m^2 ha^{-1}. Moreover, $^{equil}P_G$ also should be adjusted by examining the statistical upper bound of the stand basal area for the dense/rich natural forests. The $^{max}Z_G$ of natural moist tropical forests can be set at 1.29 m^2 ha^{-1} year^{-1}, as derived by the empirical growth model SIRENA I in Northern Costa Rica (Alder, 1997; Alder, 1996a).

In the case of forest plantations in Hong Ha, it is quite difficult to define explicitly the parameters $^{equil}P_G$ based on literature, as most of the literature on the growth of *Acacia* forests deals with plantations younger than 8-10 years-old, i.e., before the age for harvesting (e.g., Hirasuka *et al*., 2003; Vu Dinh Huong *et al*., 2004; and Do Dinh Sam, 2001). Thus, we temporarily assumed that the $^{equil}P_{Gr}$ for *Acacia* forest plantation in the study site was about 35 m^2 ha^{-1}, more or less similar to the case of natural forest. As these two *Acacia* species are fast-growing species, $^{max}Z_G$ for the plantations should be higher than that of natural forests. We set $^{max}Z_G$ = 1.5 m^2 ha^{-1} year^{-1} for *Acacia* forests in Hong Ha through approximating values from previous studies in Vietnam (see Vu Dinh Huong *et al*., 2004; Do Dinh Sam, 2001).

Table 5.2 Summary of parameters and variables of *ForestYieldDynamics* routine when implemented within the VN-LUDAS model for the study site

Parameters and variables	*Definition*	*Sources*	*Directly linked module in vn-ludas*
Parameters			
$^{Equil}p_{gr}$	stand basal area at the equilibrium state (m^2 ha^{-1})		Global-policy (i.e., tunable and user-defined parameters)
	• Natural forest: $^{equil}p_{gr}$ =35 - 38	Alder (1998, 1996a, 1996b) + descriptive statistics of plot data	
	• Mixed *acacia* plantation: $^{equil}p_{gr}$ = 35	Assumption	
$^{Max}z_g$	Possible maximal growth rate of stand basal area (m^2 ha^{-} $year^{-1}$)		Global-policy (i.e., tunable and user-defined parameters)
	• Natural forest: $^{max}z_g$ = 1.29	Interpreted from alder (1996a, 1996b)	
	• Mixed *acacia* plantation: $^{max}z_g$ = 1.5	Approximated from Vu Dinh Huong *et al.* (2004), Do Dinh Sam (2001)	
G_{logged}	Average basal area of logged trees in natural forest (m^2. *Logged tree*$^{-1}$).	Interviews local loggers and field observations	Global-policy (i.e., tunable and user-defined parameters)
T	Post-logging period with severe mortality due to logging operations (*year*). T = 3	Alder (1996a)	Global-policy (i.e., tunable and user-defined parameters)
Variables			
$^{2002}p_g$	Forest basal area measured by 2002 as initial forest yield.	Spatially random-bounded extrapolation of plot data 2002	Patch-landscape (initial forest condition)
G_{logged}	Basal area logged each time by household (m^2)	Defined by human agents in simulation runs	Decision (decided by household agents – logging action)
P_t	Years after logged (years) (temporary variable)	Elapsed since logging event, initiated by human agent in simulation runs	Decision (decided by household agents – fallow action)

The average size of logged tree g_{logged} was set based on interviewing local loggers and field observations. Because interviews often give an approximate range of g_{logged} rather than a deterministic value, we let g_{logged} receive a random value within empirical bounds. Since Alder (1996b and 2000) reported that severe logging-induced mortality has observed for about 2 years after logging, we let T in equation 5.14 be 3 years.

Validation

Because stock increment data from permanent plots in the natural forest in Hong Ha was not available, our validation compared the stand basal area simulated by the *ForestYieldDynamics* sub-model to those projected by other empirical models, which were developed based on field-measured growth data. SIRENA-I is one of the empirical models of this type for projecting stand basal area dynamics.

The SIRENA-I model was developed on the basis of data from Northern Costa Rica (Alder, 1996a and 1996b), with a conceptual framework that is definitely different from the *ForestYieldDynamics* sub-model. The sub-model is expressed as follows (see Alder, 1996b):

$$Z_G = Z_{trees} + Z_{recruitment} - Z_{nat\_mortality}$$

where Z_G is net increment of stand basal area, Z_{trees} is sum of basal area increment of trees (with dbh > 10 cm) that survive during the specified measuring period, $Z_{recruitment}$ is basal area of new/young trees recruited each year, and $Z_{nat\_mortality}$ is basal area lost due to natural mortality. Based on fitting temporal data measured from permanent plots, the three above components were expressed as functions of stand basal area P_G (Alder, 1996b):

$$Z_{trees} = 0.0419(\mathrm{P_G})^{0.8449}$$
$$Z_{recruitment} = 0.0275\mathrm{P_G} - 0.0928$$
$$Z_{nat\_mortality} = 2.0528e^{-0.0994\mathrm{P_G}}$$

For the forest plantation in Hong Ha, because the stand age of the surveyed plots clumped at about an age of 6 years, it was not possible to validate the projection curve against time-series observed data using correlation or regression analyses. An alternative validation is to compare the average basal area of the survey plots to the basal area projected at the time point of the average age.

5.3.4 Method for modeling natural transition among land-cover types: the *NaturalTransition* sub-model

Conversion among land cover types can occur either through land-use activities or natural processes that are beyond human controls. Conversions among land covers within VN-LUDAS are illustrated in Figure 5.7. Conversions due to land-use activities (transition H_x) are the result of the *FarmLandChoice* and *ForestChoice* routines in the DECISION module, which have already been specified and calibrated in Chapter 4. Natural conversions (transition Nx) are the result of the dynamics of natural vegetation growth, performed by the *NaturalTransition* sub-model.

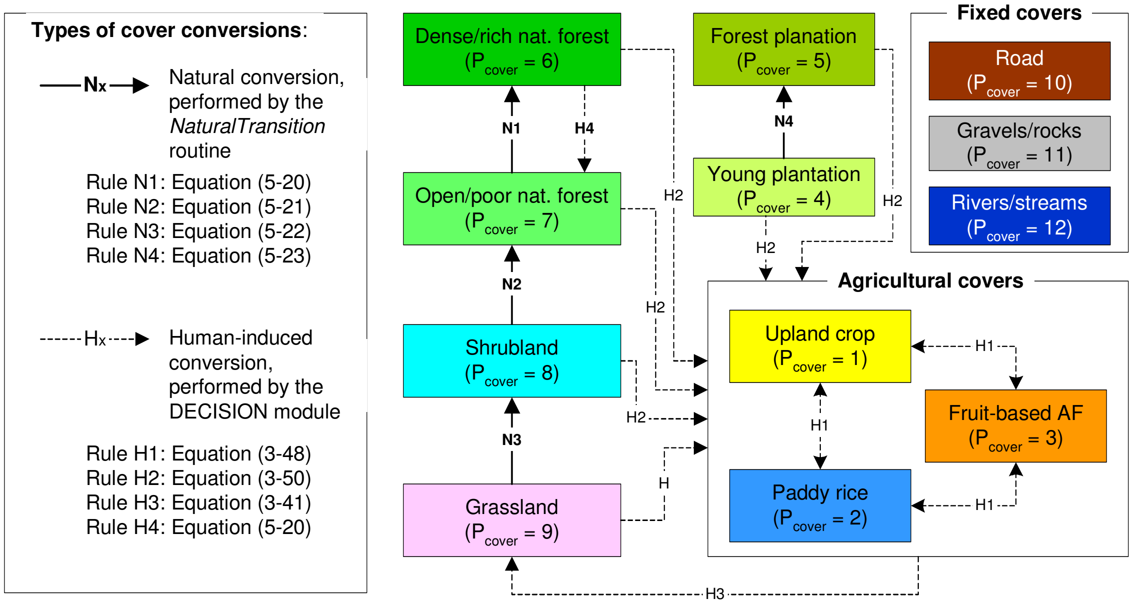


Figure 5.7 Land-cover transition in VN-LUDAS: combination of human-induced transition (influenced by DECISION module) and natural transition (viz. *NaturalTransition* sub-model)

The *NaturalTransition* sub-model is a set of transition rules that govern the natural transitions among vegetative covers. In general, the firing of these rules is based on the evaluation of the four patch variables: previous cover type (${}^{t-1}P_{cover}$), life-span of existing cover type ($P_{cover\text{-}age}$) (see Green, 1993 and Quintero *et al.*, 2004), existing stand basal area (P_{Gr}), and distance to nearest natural forest ($P_{d\text{-}forest}$). Rule evaluations for specific land-cover types are specified below.

Transition rules for natural forest cover types (transition rule *N1*)

This transition rule is defined based on: i) checking whether the previous state of the patch belonged to natural forest categories, and ii) evaluating present stand basal area P_{Gr} against the basal area thresholds for each forest-cover type. If a patch was previously natural forest and its P_{Gr} falls out of the empirical range of the current forest type, the cover state of the patch (P_{cover}) will transit to another state. The logical expression of the rule *N1* is as follows:

$$
{}^{t}P_{\mathrm{cover}} = \begin{cases} 6 & if \quad {}^{t-1}P_{\mathrm{cover}} \in \{6,7\} \;\; and \;\; {}^{t}P_{Gr} \geq \theta_{rich-forest} \\ 7 & if \quad {}^{t-1}P_{\mathrm{cover}} \in \{6,7\} \;\; and \;\; {}^{t}P_{Gr} < \theta_{rich-forest} \end{cases} \tag{5.21}
$$

where *6* and *7* are the cover codes of rich/dense and poor/open natural forests, respectively; and $\theta_{rich\text{-}forest}$ is the threshold of stand basal area for distinguishing these two natural forest types. Transitional rule (5.21) represents the cover conversion as a result of accumulative modifications. Conversion among forest cover types occurs when the magnitude of modification exceeds the threshold $\theta_{rich\text{-}forest}$. The threshold $\theta_{rich\text{-}forest}$ was defined from descriptive statistics for stand basal area in 2002.

Transitions rules for non-forest vegetative cover types (transition rules *N2*, *N3*, *N4*)

The rules of transitions among non-forest vegetative cover types or from non-forest to forest cover types are defined based on i) checking whether the patch has fallen into a non-forested vegetation category, ii) evaluating the life span of the existing cover type (${}^{t}P_{cover\text{-}age}$), and iii) the distance from the patch to the nearest natural forest area (${}^{t}P_{d\text{-}forest}$), and iv) a general understanding of vegetation succession dynamics. The first two criteria are commonly used in simple cellular automata models for vegetation succession (see Green, 1993; Quintero, 2004), based on the ecological principle that a non-forest vegetation has a capacity to recover and naturally convert back to natural forest through the positive serial succession: *grassland* → *shrubland* → *regenerating forest*, provided no human disturbance (e.g., burning) during a long enough period.

The last criterion takes into account the site-specific potential of natural forest regeneration. Patches with grass or shrub at forest edges or forest gaps have a better chance to convert back to secondary forest, since they receive seed rains from adjacent natural forests, still maintain a seed bank in the top soil, and benefit from the forest micro-environment. Patches with grass or shrub far from natural forest areas have less, or no capacity to regenerate because they lack sources of tree seeds, or are prone to soil degradation.

If a patch has previously had a shrub cover, the life-span of the existing cover is long enough (${}^{t}P_{cover\text{-}age} > \theta_{t\text{-}forest}$), and the patch is located next to a natural forest area (${}^{t}P_{d\text{-}forest} > \theta_{d\text{-}forest}$), the cover state of the patch will change to poor secondary forest, otherwise it remains shrubland. Hence, the logical expression of rule *N2* is as follows:

$$ {}^{t}P_{\text{cover}} = \begin{cases} 7 & if \quad {}^{t-1}P_{\text{cover}} = 8 \;\; and \; {}^{t}P_{\text{cover}-age} > \theta_{t-forest} \;\; and \;\; {}^{t}P_{d-forest} < \theta_{d-forest} \\ 8 & if \quad otherwise \end{cases} \tag{5.22} $$

where $\theta_{t\text{-}forest}$ is the threshold of the life-span of shrubland, which is used for deciding if the shrubland patch is converted to open/poor natural forest. $\theta_{d\text{-}forest}$ is the threshold of the distance from the patch to the nearest natural forest, which is used for determining if the shrub patch can change to open/poor natural forest.

Similarly, rule *N3* for the transition from grassland and to shrubland is expressed as follows:

$$ {}^{t}P_{\text{cover}} = \begin{cases} 8 & if \quad {}^{t-1}P_{\text{cover}} = 9 \;\; and \; {}^{t}P_{\text{cover}-age} > \theta_{t-shrub} \;\; and \;\; {}^{t}P_{d-forest} < \theta_{d-forest} \\ 9 & if \quad otherwise \end{cases} \tag{5.23} $$

where $\theta_{t\text{-}shrub}$ is the threshold of the grassland life-span used for deciding if the grassland patch is transited to shrubland.

Rule *N4* for transition from young plantations (${}^{t-1}P_{cover} = 4$) to forest plantation ($P_{cover} = 5$) is expressed as follows:

$$ {}^{t}P_{\text{cov}\,er} = \begin{cases} 5 & if \quad {}^{t-1}P_{\text{cov}\,er} = 4 \quad and \quad {}^{t}P_{\text{cov}\,er-age} > \theta_{t-plantation} \\ 4 & if \quad otherwise \end{cases} \tag{5.24} $$

where $\theta_{t\text{-}plantation}$ is the threshold of the young plantation's life-span used for deciding if the young plantation patch changes into a forest plantation.

All threshold values for rule evaluations in equations 5.22, 5.23, and 5.24 were calibrated based on field-based observations and interviews with local experts (i.e., farmers and local forestry officials).

5.4 Results and discussion

5.4.1 Landscape characterization

Results of terrain analysis

The DEM of the study area is shown in Figure 5.8a. Elevation is highly variable across space, i.e., ranging from 35 to 1370 m a.s.l within an area of 90 km^2 only. The eastern side of the area is the downstream zone where the Rao Nho and Bo rivers meet, while the western part is the high mountain area of the Annamite Range.

Figure 5.8b shows the calculated slope grid (P_{slope}). The image shows greatly complex terrains over the study area. The study area is steep, dominated by the slope classes III and IV (i.e., ranging 15 – 35°) (see Figure 5.9). Flat land or areas with gentle slope are located mainly along the downstream part of the Rao Nho and Bo rivers, which are agricultural and settlement zones of Hong Ha community.

The computed upslope contributing area (P_{As}) ($\log_{10}$ transformation) and topographic wetness index (P_{wet}) grids are shown in Figure 5.8c and 5.8d, respectively. The higher value of P_{As} indicates the higher potential of water flow accumulation. High values of P_{As} often occur in zones of foot slope or along drainage paths, while a low P_{As} is found in mountain ridges or upslope positions. The higher value of P_{wet} reflects the higher degree of water saturation of the site. In general, the spatial patterns of P_{As} and P_{wet} have a high agreement with the pattern of drainage network, indicating a dense stream network throughout the study area. Moreover, as anticipated in the methodology, the wetness index P_{wet} delineated efficiently the locations of saturated zones (i.e., the zone with dark blue colour in Figure 5.8d), which was confirmed through our field surveys.

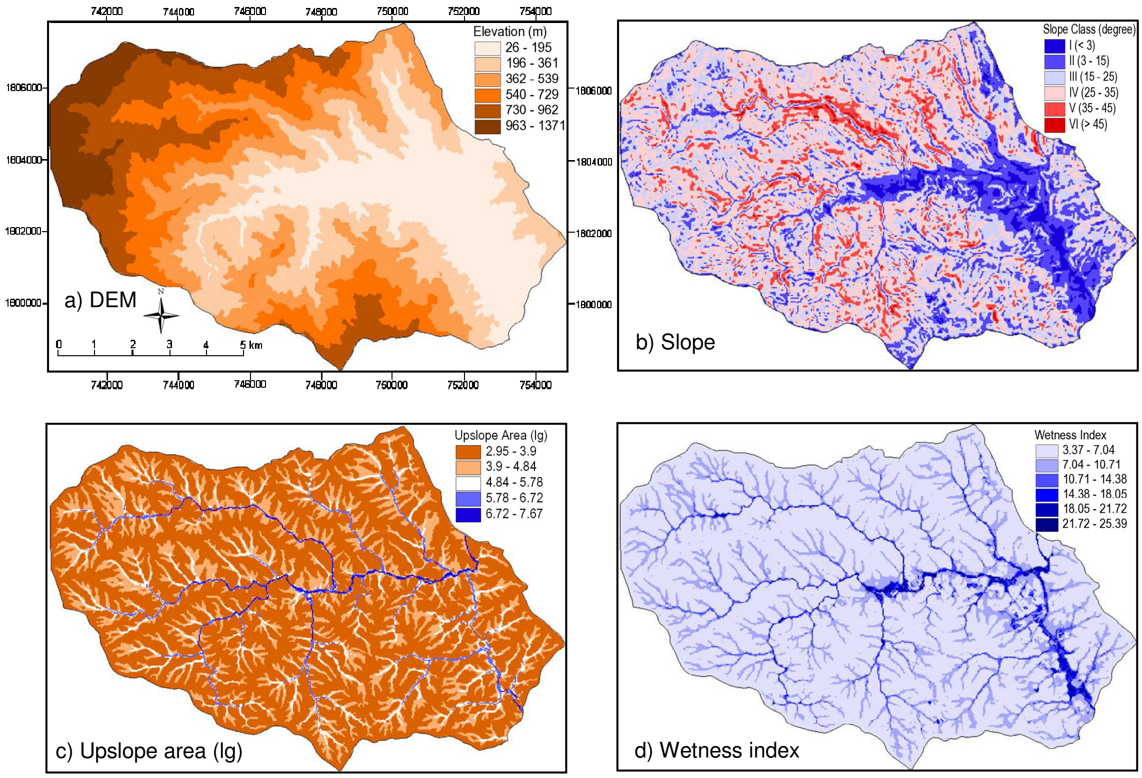


Figure 5.8 Raster images of a) elevation (m), b) slope gradient (degree), c) upslope contributing area (m^2/m) (log10 transformation), and d) wetness index in the Hong Ha watershed. Map coordinate system: UTM, Zone 48 North, Datum WGS84

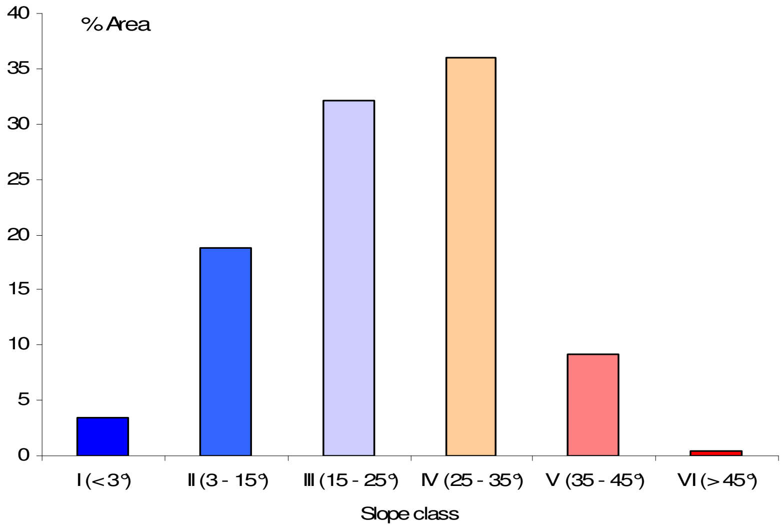


Figure 5.9 Distribution of slope classes over the study area. Data source: calculated from the slope grid

Results of accessibility analysis

The grids of proximate distances to roads (P_{road}) and stream/water (P_{river}) are shown in Figures 5.10a and 5.10b, respectively. Obviously, the difference between patterns of road and stream/river networks leads to the difference between the accessibility variables. The simple and single road network, combined with the severe terrain conditions, has resulted in a great inequality among locations with regard to access to the road network, especially in the north-south direction. In contrast, the dense and complex drainage network gives a better equality among location in access to rivers/streams.

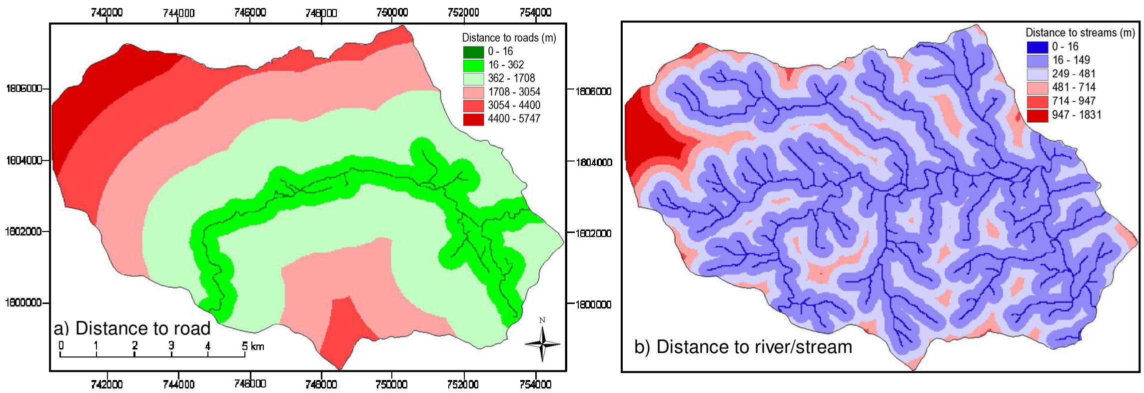


Figure 5.10 Raster images of a) proximate distance to roads (m), and b) proximate distance to rivers/streams (m) in Hong Ha watershed. Map coordinate system: UTM, Zone 48 North, datum WGS84

Result of land cover/use classification

The classification process of land cover 2002/03 delineated 12 land cover types as shown in Figure 5.11 and Table 5.3. Natural forests, including dense/rich and open/poor forests, occupy 50 % of the total area and are found in remote mountains or steep land. Shrublands cover 25 % of the total area, appearing mainly in the transitional zone between man-made cover areas and natural forests. Grassland covers about 5 % of the total area, located on the border between shrubland and forest plantations. *Acacia* plantation forests are on about 9 % of the total area, located along the main road No. 49 in forms of plantation compartments or small patches between agricultural patches. Total agricultural land takes a very small proportion, i.e., only 3 % of the total area. The remaining lands are bare land (rocky/gravel surfaces), road and water surfaces.

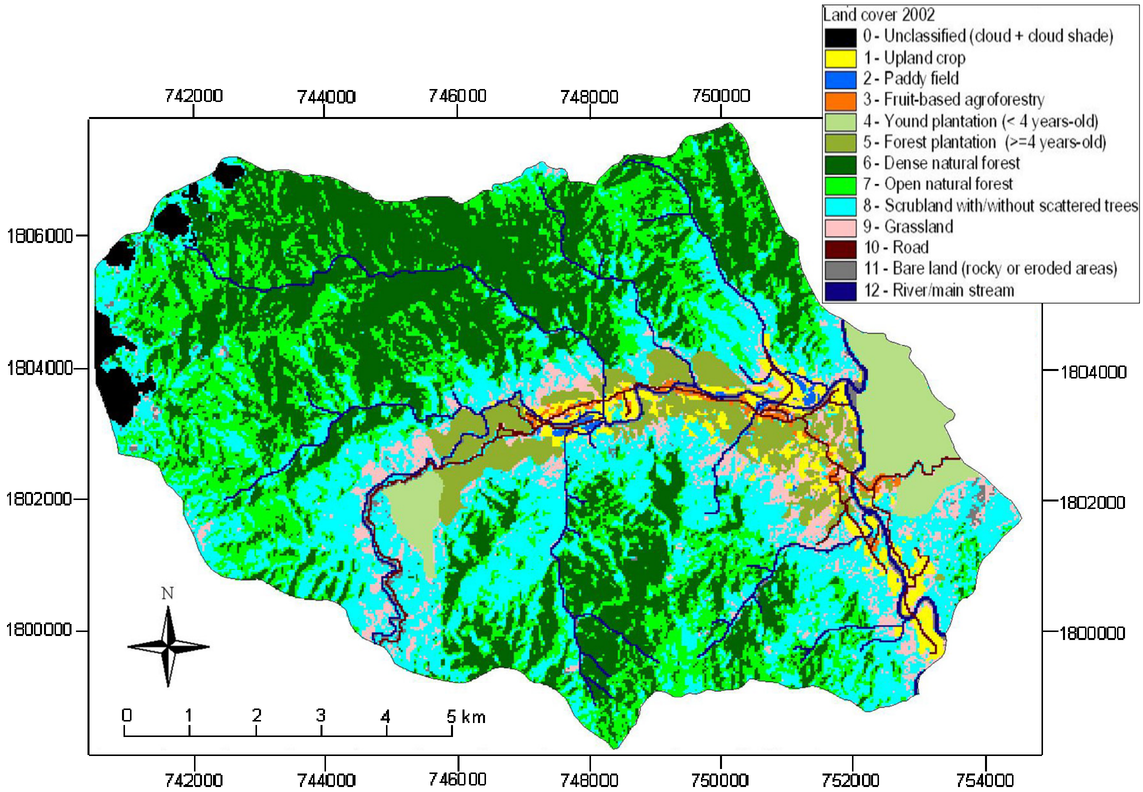


Figure 5.11 Raster image of land cover/use types in 2002 in Hong Ha watershed. Map coordinate system: UTM, Zone 48 North, Datum WGS84

Ground-truth surveys associated with interviewing local forestry experts gave brief descriptions[15] of the main cover types as follows. The dense/rich natural forests (P_{cover} = 6) include primary forests with medium disturbances or secondary forests in mature state. The forest structure is still dense, typically including five layers, with diverse floristic compositions. The canopy layer is complex and continuous, with a coverage of about 50-70 % and a height of 20-30m, formed by common tree species such as: *Quercus platycalyx* Hickel et A. Camus, *Lithocarpus ducampii* (Hickel et A.Camus) A. Camus, *Castanopsis indica* A. DC., *Quercus myrsinaefolia* Blume (Fagaceae), *Michelia balansae* Dandy, *Michelia mediocris* Dandy (Magnoliaceae), *Cinnamomum tonkinense* A. Chev. (Lauraceae), *Syzygium zeylanicum* DC. (Myrtaceae), *Polyalthia nemoralis* Aug. DC. (Annonaceae), *Scaphium macropodium* (Miq.) Beumee (Sterculiaceae).

---

[15] Floristic data crossing the study area (in forms of species-sampling unit matrices) are also collected during the ground-truth survey. Although these data are not presented here, they are available from the author.

The open/poor natural forests (P_{cover} = 7) are either immature secondary forest with remnant trees, or regenerating forests with pioneer tree communities. Immature secondary forests are severely degraded, with canopy coverage ranging from 30 % to 50%, and with no clear layer stratification. Tree species composition is quite similar to that of dense/rich forest, except that commercial tree species have become very rare due to overexploitations. Regenerating forests succeed after either fallowed or abandoned swidden fields, and is dominated by fast-growing tree species such as *Macaranga trichocarpa* Müll. Arg. and *Macaranga denticulata* Müll. Arg. (Euphorbiaceae).

Shrubland (P_{cover} = 8) is an extremely degraded formation of evergreen tropical forests. Shrublands are distributed in the shape of shrub parcels and bush sheets. Shrub parcels are located at the edge of or in the gaps within natural forests, and are dominated by shrubs, bushes and sometimes by scattered remnant trees from the previous forests. Likely, these shrub parcels are able to transform gradually to regenerating forests thanks to receiving seed rains from adjacent forests, or by maintaining an intact seed bank in the topsoil. Bush sheets are large and continuous areas of bush species, predominated by *Melastoma candidum* D. Don (Melastomataceae) of 1 - 2 m height. Because of the dense bush layer and the lack of tree seed sources, any natural forest regeneration in these bush sheets is unlikely.

Table 5.3 Areas of land cover types in Hong Ha watershed in 2002

Land cover type	*Code*	*Area (ha)*	*Percentage (%)*
No data (cloud + cloud shade)	0	131	1.4
Upland crop	1	211	2.3
Paddy field	2	18	0.2
Fruit-based agroforestry	3	40	0.4
Acacia forest plantation ≤ 4 years-old	4	379	4.2
Acacia forest plantation > 4 years-old	5	418	4.6
Dense natural forest	6	2672	29.4
Open natural forest	7	1993	21.9
Shrubland	8	2332	25.6
Grassland	9	452	5.0
Road	10	105	1.2
Bare land (gravel/rocky surface)	11	77	0.8
River/main stream	12	266	2.9
Total		*9095*	*100.0*

Data source: calculation based on the land-use/cover grid.

Grasslands (P_{cover} = 9) in the study area are dominated by tall and coarse grass species, such as *Imperata cylindrica* Beauv., *Saccharum arundinaceum* Retz., *S. spontaneum* L., and *Thysanolaena latifdia* Honda (Gramineae). It is also distinguished between grass parcels and sheets. Grass parcels develop immediately after swidden fields are fallowed, normally at forest edges or gaps, and will transform to shrubland in the next 2.3 years. Grass sheets are large and continuous areas on hillsides, invaded by *Imperata* grasses (*Imperata cylindrica*) or broom grasses (*Thysanolaena latifdia*). There is almost no chance for naturally regenerating forest, unless reforestation is carried out.

Plantations in the area include two cover categories: plantations with no forest canopy, called young plantation (P_{cover} = 4), and plantations with forest canopy, called forest plantation (P_{cover} = 5). Young plantations were established in 2000, 2001 and 2002; *Acacia mangium* Willd.is the only tree species. Forest plantations were established during the years 1994, 1995, 1996, and 1997. Most forest plantations are single species stands, with either *Acacia auriculiformis* or *Acacia mangium* trees.

Agricultural land was classified into paddy rice (P_{cover} = 2), upland crop (P_{cover} = 1) and agroforestry (P_{cover} = 3). Paddy fields, with two crops of wet rice a year, are tiny patches located mainly in saturated zones along the Rao Nho river. Upland crop fields are found in either flat lands or hillsides, where upland rice, maize and cassava are the main crops. Agroforestry farms are often located along roads and mixed with residential areas, often in the form of home gardens. The tree component of agroforestry farms includes mainly jackfruit trees (*Artocarpus heterophyllus* Lam.) or *Acacia auriculiformis* trees, which creates approximately 10 – 30 % tree coverage on the farms. However, the main products of agroforestry farms are from fruit crops, e.g., bananas and pineapples, rather than from the tree component, thus the term fruit-based agroforestry farms.

5.4.2 Modeling the dynamics of agricultural yield responses

Descriptive statistics of variables used for agricultural year models

Descriptive statistics of variables used for agricultural yield models for the agricultural year 2002/03 are given in Table 5.4. Paddy rice yields were 6099 ± 705 kg ha^{-1} $year^{-1}$, nearly double the yields of the two other cultivation types. The high yield relates to the fact that most farmers in Hong Ha grow high-yielding varieties of rice (i.e., TH-30 and

Khang Dan) with two crops a year, and invest much larger amounts of chemical fertilizers than in the upland and fruit-based system. The paddy rice system is also the most labor demanding. During the production cycle of paddy rice, activities such as land preparation, transplanting and weeding usually require many man days. Most of the paddy fields are located on flat land, with slopes ≤ 5° (mostly slope class I). The upslope contributing area is not significantly different among plots of different agricultural types.

Table 5.4 Descriptive statistics of variables for the three agricultural yield models in the agricultural year 2002/03

Model	*Number of plots (n)*	*Mean*	*Standard deviation (SD)*	*Confidence interval at 95% level ($CI_{0.05}$)*
Paddy rice yield $P_{y\text{-}paddy}$ (kg rice ha^{-1} $year^{-1}$)	73	6099	3073	705
Agrochemical input I_{chem} (*1000 VND ha^{-1} $year^{-1}$*)	73	2463	3561	817
Labor input I_{labor} (*man day ha^{-1} $year^{-1}$*)	73	882	543	125
Slope P_{slope} (*degree*)	73	4	2	1
Upslope contributing area P_{as} (*m m^{-2}*)	73	1474219	3915630	898230
Cropping time P_t (*year*)	73	7	4	1
Upland crop yield $P_{y\text{-}upcrop}$ (kg rice ha^{-1} $year^{-1}$)	134	3637	2785	471
Agrochemical input I_{chem} (*1000 VND ha^{-1} $year^{-1}$*)	134	153	597	101
Labor input I_{labor} (*man day ha^{-1} $year^{-1}$*)	134	681	598	101
Slope P_{slope} (*degree*)	134	6	5	1
Upslope contributing area P_{as} (*m m^{-2}*)	134	1857094	5719372	968375
Cropping time P_t (*year*)	134	3	2	0
Fruit-based af yield $P_{y\text{-}af}$ (kg rice ha^{-1} $year^{-1}$)	47	3602	3031	867
Agrochemical input I_{chem} (*1000 VND ha^{-1} $year^{-1}$*)	47	78	220	63
Labor input I_{labor} (*man day ha^{-1} $year^{-1}$*)	47	448	413	118
Slope P_{slope} (*degree*)	47	6	4	1
Upslope contributing area P_{as} (*m m^{-2}*)	47	1533564	4611574	1318402
Cropping time P_t (*year*)	47	6	3	1

Data source: Plot-based survey during 2003. P_{As} data were calculated from the DEM as positions of the plots are known. VND is the Vietnamese currency unit.

The yield of annual upland crops is equal 3637 ± 471 kg rice ha^{-1} $year^{-1}$. In contrast to paddy rice, the upland crop system in Hong Ha is an extensive cropping system, with no or very low chemical fertilizer input. Although labor input is significantly lower than in the case of paddy rice, the upland crop system is still labor demanding. For the plots opened for the first time, slashing vegetation and removing burnt material is usually the heaviest work. For plots cultivated in later years, weeding is probably the most labor consuming work, as it needs to be done frequently to prevent the invasion of weeds, e.g., *Imperata* grasses. Most of the upland crop fields have been located on gentle slopes (5 - 7°), some on either flat lands or hillsides. The steepest slope observed is as 25°. The shorter cropping time of this cultivation type (about 3 years) relates to the rotation cropping-fallow of swidden cultivation. The cropping time of upland crops is shorter than the lowland cropping period, i.e., usually 4 years for upland crops on hillsides and about 7 years in the flat zone.

The yield of fruit-based agroforestry equals 3602 ± 867 kg rice ha^{-1} $year^{-1}$, which is more or less the same as that of upland crops. This cultivating system has the lowest annual fertilizer input and labor demand. Most of the agroforestry farms are located on gentle slopes (5 - 7°). The average cropping time of agroforestry farms reported here does not relate to any specific cultivation period. Interviews with the farmers showed that newly established agroforestry farms are likely to be continued in the long term.

Modeling agricultural yields

The results of log-linear regression analyses for yield models of the three main agricultural land-use types are reported in the Table 5.5.

Estimation of paddy rice yield

The significant value of F-statistic $p < 0.01$ indicates that the model is capable of explaining the change in paddy rice yield. Quantitatively, the R^2 of 0.66 means that 66% of the variation in the observed paddy yield is explained by the model. This indicates a good fit of the model to the observed data. The standard error (s) of the estimate of 0.311, i.e., $(0.311/8.589) \times 100\ \% = 4\ \%$ of the average predicted $ln(P_{y\text{-}padd})$, shows a good predictive precision of the model.

Table 5.5 Results of log-linear regressions for yields of three agricultural land-use types

Agriculture yield models	*Pearson correlation coefficient (r)*	*Unstandardized coefficient (yield elasticity) (β)*	*Standard error of β (σ_β)*	*Confidence interval at 95% level*
Ln of paddy rice yield $ln(P_{y\text{-}paddy})$ n=73; $mean(ln(P_{y\text{-}paddy}))$=8.589 R^2=0.660; s = 0.311; p= 0.000				
(constant)		5.418***	0.506	1.010
Ln of agrochemical input $ln(I_{chem.})$	0.570***	0.040***	0.012	0.025
Ln of labor input $ln(I_{labor})$	0.766***	0.470***	0.065	0.130
Ln of slope $ln(P_{slope})$	-0.300***	-0.125	0.083	0.166
Ln of upslope contributing area $ln(P_{as})$	0.044	0.007	0.013	0.026
Ln of cropping time $ln(P_t)$	-0.198**	-0.029	0.046	0.093
ln of upland crop yield $ln(p_{y\text{-}upcrop})$ n=134; $mean(ln(P_{y\text{-}upcrop}))$= 7.927 R^2=0.379; s = 0.619; p=0.000				
(constant)		6.130***	0.464	0.919
Ln of agrochemical input $ln(I_{chem.})$	0.164**	0.045**	0.023	0.046
Ln of labor input $ln(I_{labor})$	0.464***	0.368***	0.060	0.118
Ln of slope $ln(P_{slope})$	-0.231***	-0.271***	0.090	0.178
Ln of upslope contributing area $ln(P_{as})$	0.157**	0.021	0.019	0.038
Ln of cropping time $ln(P_t)$	-0.304***	-0.335***	0.077	0.152
ln of agroforestry yield $ln(P_{y\text{-}af})$ n=47; $mean(ln(P_{y\text{-}af}))$= 7.866 R^2=0.360, s = 0.710; p= 0.002				
(constant)		3.317***	1.104	2.230
Ln of agrochemical input $ln(I_{chem.})$	0.018	0.004	0.046	0.093
Ln of labor input $ln(I_{labor})$	0.364***	0.452***	0.140	0.283
Ln of slope $ln(P_{slope})$	-0.014	0.221	0.176	0.356
Ln of upslope contributing area $ln(P_{as})$	0.204*	0.054	0.040	0.080
Ln of cropping time $ln(P_t)$	0.402***	0.584***	0.174	0.351

Note: The symbols *, **, and *** indicate statistical significance at the confidence level of 90%, 95% and 99%, respectively. Data source: Plot-based interview during the summer 2003.

Explanatory variables that have significant effects on paddy yield are I_{chem} (+) and I_{labor} (+). The directions in which these variables operate support our hypotheses and justification of their roles in the yield response model. The non-significant effect of terrain variables on paddy yield relates to the fact that all paddy rice plots are located on flat land, resulting in very small variations in terrain conditions among paddy plots (see Table 5.4). Cropping time (P_t) has only a marginally significant effect on paddy yield. This probably relates to the fact that soil degradation over cropping years is not a serious problem for paddy rice production, as paddy rice fields are located on flat land and fertilizer is applied regulary.

The quantitative interpretation of the yield response patterns should be judged by the magnitude of the coefficient for yield elasticity (β). The β for I_{chem} and I_{labor} are less than 1, which means that the paddy yield curve is increasing but concave down along with the increment of agrochemical or labor inputs. This concave down pattern of the yield response indicates that the paddy yield increase is likely to be more marginal when farmers invest more agrochemicals or labor (see Figure 5.3). Moreover, the value $\beta = 0.040$ for I_{chem} means that if farmers double (i.e., 100 %) the agrochemical input, the paddy yield is likely to increase by only $0.040 \times 100\ \% = 4\ \%$, assuming that other variables remain unchanged. The low yield elasticity to agrochemical input in the context of high input and lower degradation risk (i.e., flat land) suggests that the average I_{chem} for paddy rice in Hong Ha (Table 5.4) may be close to the saturation point of demand.

A similar interpretation for I_{labor} shows that paddy yield is much more sensitive to the change of labor input ($\beta = 0.470$). This suggests that investing more labor in paddy rice field can be recommended to improve yield. However, it also indicates that paddy rice is a labor demanding cropping system.

Estimation of upland crop yield

The F-statistic test shows that the log-linear regression model is able to explain significantly the variation of the upland crop yield ($p < 0.01$) (Table 5.5). The R^2 of the model is only 0.379. However, Studenmund (1997) noted that for a cross-sectional dataset, which consists of observations crossing different types of explanatory variables for the same time period, an R^2 value of 0.50 would be considered at good fit. Because data of most variables were obtained through interviewing plot owners rather than through field measurement, and the yield data of different crops are converted to rice using monetary method[16], there may be considerable errors/distortion associated with either data acquisition or yield conversion. The standard error of the estimate is 0.619, i.e., $(0.619 / 7.927) \times 100\ \% = 8\ \%$ of the average predicted $ln(P_{y\text{-}upcrop})$, indicating an

[16] Yields of crops which are not rice (e.g., vegetables, peppers, pineapple, cassava, etc.) were converted to a monetary equivalent amount of rice using the local price in the summer 2003.

acceptably predictive capacity of the regression model. Moreover, most of the explanatory variables have significant effects on upland crop yield in the anticipated directions, which shows a good explanatory power of the model.

Explanatory variables significantly affecting upland crop yield are I_{chem}(+), I_{labor}(+), P_{slope}(-), and P_t(-). The directions in which these variables operate are as anticipated and justify their roles in deriving the yield response of upland crop patches. These results also shows that the estimated model is a reasonable representation of spatio-temporal dynamics of upland crop yield through the combination of land capability (viz. P_{slope}), management factors (viz. I_{chem} and I_{labor}), and time (P_t). P_{slope} varies greatly over space, I_{chem} and I_{labor} fluctuate highly along the diversity of human agents, P_t elapses regularly and stops stochastically. As a result, actual yield responses of landscape agents (i.e., land patches) are extremely heterogeneous.

Interpretation of the elasticity coefficient β show more explicit responses of upland crop yield to the changes in explanatory variables. The β values of I_{chem} and I_{labor} are less than 1, indicating that the yield increases of upland crop fields tended to be marginal with the increase of fertilizer and labor inputs. Similarly to the case of paddy rice, the yield elasticity of upland crop patches to agrochemical input is very low (β = 0.045). However, the underlying cause of the phenomenon this case is different. Small yield elasticity to fertilizers under the conditions of low fertilizer input and high soil erosion risk (i.e., higher P_{slope}) (see Table 5.4) suggests that upland crop fields in Hong Ha are generally close to a marginal status.

Upland crop yield responded quite elastically to the change of labor input (β = 0.368). The rapid invasion of weeds, especially *Imperata cylindrica* grasses, on upland fields inhibits crop yield through competion for mineral nutrients, lights, water, and living space. In order to prevent crop yield from declining dramatically, farmers usually have to weed 3 or 4 times per crop season, and this work is quite labor consuming. Field observations and interviews also showed that most plots with low or no yields were dominated by weeds, and there were less weeding activities there.

Upland crop yield is found to decrease considerably with land slope increase (β = -0.271). Assuming other variables remain constant, if the slope angle doubles, e.g., from the mid-point of the slope class II (9 - 10°) up to the mid-point of the slope class III (20°), the upland crop yield will likely be reduced by 0.271 × 100 % = 27.1 %. This finding illustrates a spatially explicit rule that generates heterogeneously spatial responses of upland crop yield.

Upland crop yield was also considerably reduced over the cropping year (β = 0.335). Assuming that other variables remain unchanged, after four years of cultivation, the upland crop yield would decrease to 57 % of the yield in the first year (see Figure 5.11). This finding is consistent with the observation that villagers often fallow upland crop fields on hillsides after about 4 years of cultivation.

Estimation of fruit-based agroforestry yield

The F-statistic test shows that the model explains significantly the change of fruit-based agroforestry yield ($p < 0.01$) (Table 5.5). As in the case of upland crop yield, the R^2 of 0.360 indicates that the model is reasonably good in fitting the observed cross-sectional data. The prediction standard error is 0.710, i.e., (0.710 / 7.866) × 100 % = 9 % of the average predicted $ln(P_{y-af})$, which shows a fair predictive capacity of the model.

Explanatory variables significantly affecting upland crop yield are I_{labor}(+) and P_t(+) with the directions in which these variables operate as expected. Non-significant effects of I_{chem} and P_{slope} relate to the fact that almost all farmers in Hong Ha have not used chemical fertilizers on their agroforestry farms, and that most of the farms of this type are located on flat land.

Yield elasticity with respect to labor input is similar to the case of paddy rice or upland crops (β = 0.452). However, in reality, labor input for agroforestry farms is probably lower than for the other two cropping systems, since home-gardening does not involve periods of labor concentration, and household members (including elders and children) can use some of their free time for gardening work.

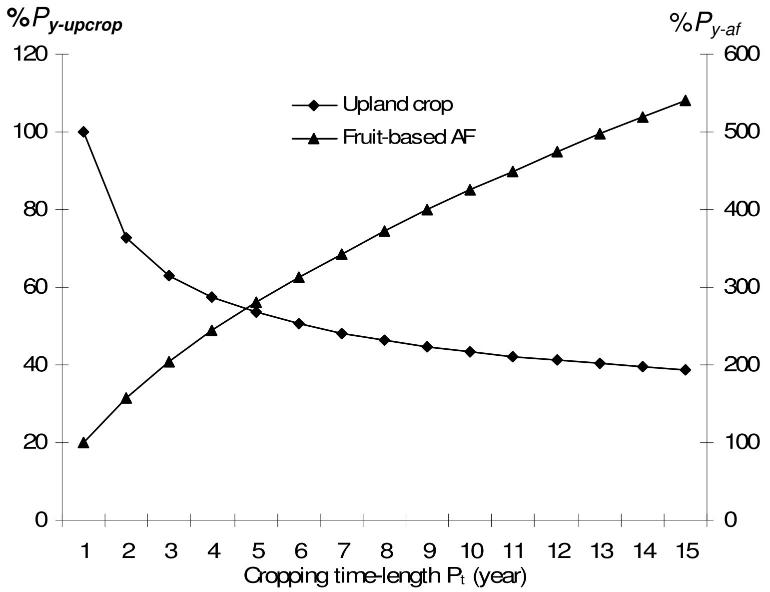


Figure 5.12 Simulated relative changes of upland crop yield (%$P_{y\text{-}upcrop}$) and fruit-based agroforestry (%$P_{y\text{-}af}$) over cropping time (P_t). Note: yield in the first year is regarded as 100% and used as the base for calculated relative yields in later years

In contrast to upland crops, yields of fruit-based agroforestry farms increase remarkably with cropping time (β = 0.584). Assuming other variables do not change, projecting relative $P_{y\text{-}af}$ along P_t shows that fruit-based agroforestry seems to be productive in both the short and long term (Figure 5.12). For example, in the fourth year the agroforestry yield increases to up to 2.5 times the yield of the first year. This finding agrees with the fact that pineapple and banana crops will be harvested for the first time two or three years after planting. Subsequently, the auto-vegetative propagation of bananas and pineapples increases the density of these crops and subsequently return higher yields. In later years, fruit-trees (e.g., lemon and jackfruit trees) and black peppers will probably increase overall annual yields, while some banana and pineapple crops will be replaced due to declining yields. Thus the annual yield will still increase but at a slower rate (see Figure 5.12).

5.4.3 Modeling the dynamics of stand basal area

Random-bounded extrapolation of initial basal area for forested patches

The descriptive statistics of the forest stock in 2002 in terms of stand basal area ($^{2002}P_{Gr}$) for three main forest cover types are shown in Table 5.6. The stand basal area of dense/rich natural forest, open/poor natural forest, and *Acacia* forest plantation are 32.94 ± 4.38 m^2 ha^{-1}, 18.28 ± 2.74 m^2 ha^{-1}, 8.37 ± 2.98 m^2 ha^{-1}, respectively. The stand basal areas of the two categories of natural forests in our case agree with the inventory results for natural forests in the Phong Dien Natural Reserve, which is located about 15 km north of the Hong Ha watershed (Le Trong Trai *et al.*, 2001).

Based on the above results and the raster image of the land cover 2002, the rule-based function for extrapolating forest basal area for all forested patches in 2002 was calibrated as follows:

$$ {}^{2002}P_{Gr(j)} = \begin{cases} 28.56 + random(8.76) & if \quad {}^{2002}P_{cover(j)} = 6 \\ 15.54 + random(5.48) & if \quad {}^{2002}P_{cover(j)} = 7 \\ 5.40 + random(5.96) & if \quad {}^{2002}P_{cover(j)} = 5 \\ 0 & if \quad {}^{2002}P_{cover} \notin \{5,6,7\} \end{cases} \tag{5.25} $$

where ${}^{2002}P_{Gr(j)}$ and ${}^{2002}P_{cover(j)}$ are stand basal area ($m^2.ha^{-1}$) and cover type (code) of a patch j^{th}, respectively.

Table 5.6 Descriptive statistics of stand basal area 2002 for the three main forest cover types in Hong Ha watershed

	Number of surveyed plots	Mean stand basal area [m^2 ha^{-1}] ($\overline{P_{Gr\,(i)}}$)	Standard deviation (SD_{gi})	Confidence interval at 95% level [m^2 ha^{-1}] (CI_{gi})	Confidence limits at 95% level	
					Lower bound [m^2 ha^{-1}]	Upper bound [m^2 ha^{-1}]
Dense/rich natural forest (p_{cover} = 6)	34	32.94	13.03	4.38	28.56	37.32
Open/poor natural forest (p_{cover} = 7)	28	18.28	7.39	2.74	15.54	21.02
Forest plantation[a, b] (p_{cover} = 5)	12	8.37	6.26	2.98	5.40	11.35
Young plantation[a, c] (p_{cover} = 4)	4	0.00	0.00	0.00	0.00	0.00

[a] Data from FOF inventory in October 2002.
[b] Age of forest plantation plots ranges from 5 to 8 years, average age is 5.9 years.
[c] Age of young plantation plots ranges from 1 to 2 years.

The generation of the grid [${}^{2002}P_{Gr(j)}$] using equation 5.25 was done within the initialization procedure of the VN-LUDAS on the *NetLogo* platform. Through the random-bounded rules, the generated distribution of the stand basal area over all patches within each forest-cover type is shown in form of a coarse texture (see Figure 5.13).

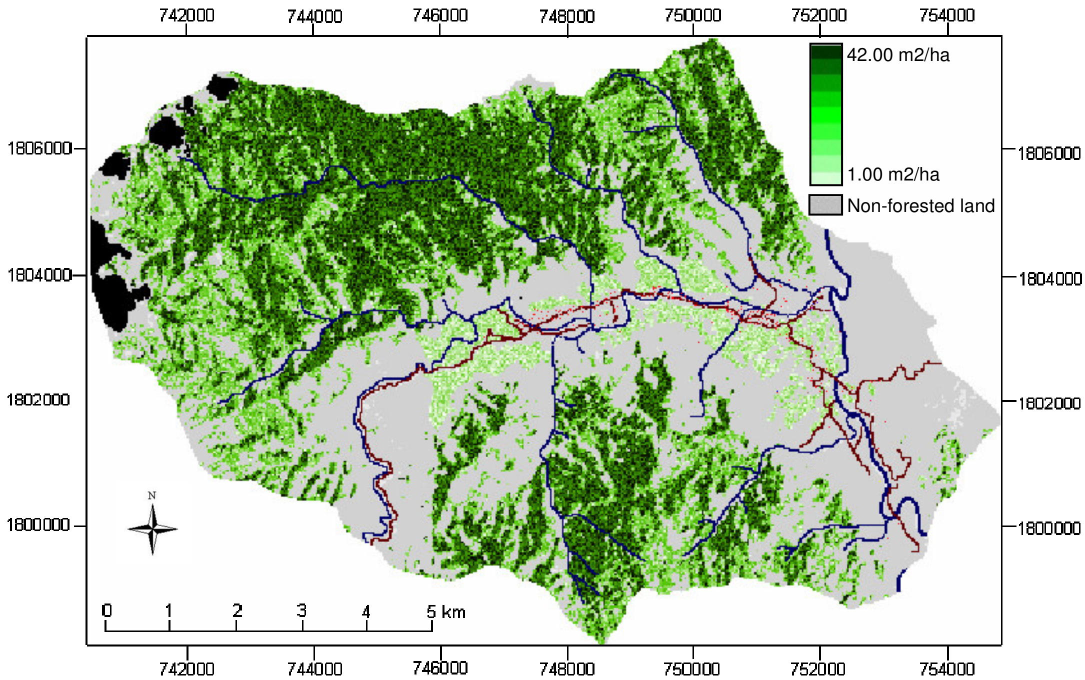


Figure 5.13 Maps showing the spatial distribution of the stand basal area in forested area in 2002 (computed in the *NetLogo* platform using the random-bounded rules as in equation 5.25). Map coordinate system: UTM, Zone 48 North, Datum WGS84

Projecting the dynamics of forest patches: results of *ForestYieldDynamics* sub-model

Yield dynamics of natural forest

The upper confidence limit (95%) of ${}^{2002}P_{Gr}$ for the dense/rich forest in the study area is 37 - 32 m^2 ha^{-1} (see Table 5.6), which does not differ much from the limit 35 - 36 m^{2}ha^{-1} given by Alder (1996, 1998). Therefore, we set the parameter ${}^{equil}P_{Gr}$ of 38 m^2 ha^{-1}. The set value of ${}^{equil}P_{Gr}$ is a little higher than the limit given by Alder (1996a, 1998) and the upper confidence limit of our measurement. Thus this set value is treated as a *theoretical threshold* of the yield curve, which the yield curve can approach but has never mathematically met. In this study, the actual projected yield curve was almost stable when it reached the value 35 – 36 m^2 ha^{-1}.

Interviews with local farmers revealed that the size of logged trees ranged from 60 to 90 cm *dbh*, i.e., 0.283 – 0.636 m^2 of the basal area. Hence, the parameter g_{logged} is empirically random-bounded as followed:

$$g_{logged} = 0.2827 + random(0.2827\text{-}0.6361) = 0.2827 + random(0.3534)$$

The *ForestYieldDynamics* sub-model was built into the VN-LUDAS model, in form of a procedure programmed in the NetLogo platform. The fully spatio-temporal dynamics in stand basal area of all forested patches and feedbacks is concurrent with the dynamics of other sub-systems of the VN-LUDAS. To illustrate the yield dynamics of a forested patch through the functioning of the *ForestYieldDynamics* sub-model, we temporally examined the behavior of a single forested patch as shown in Figure 5.14.

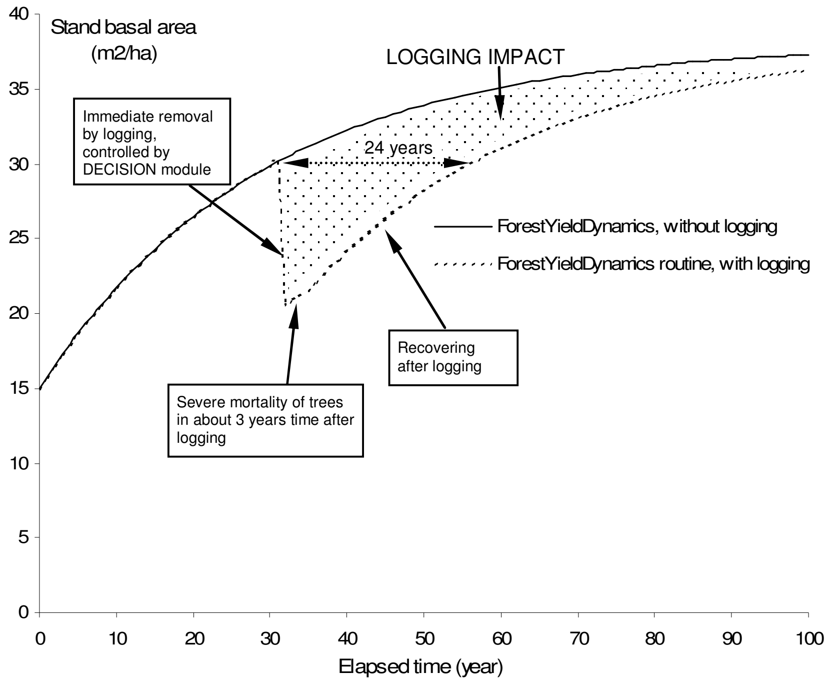


Figure 5.14 Projections for basal area dynamics of a typical open/poor forest stand using *ForestYieldDynamics* sub-model, with and without logging

Figure 5.14 shows the projection for stand basal area P_{Gr} of a typical open/poor forest patch, with the initial stock ${}^{2002}P_{Gr}$ = 15 m^2 ha^{-1}, using the *ForestYieldDynamics* sub-model. If there is no disturbance (e.g., no logging), stand basal area of the patch increases over time. After about 20-25 years, the forest has recovered with a relatively high growth rate. With time advancing, the growth rate gradually decreases. After the 50*th* year, the forest stand approaches the equilibrium state, and the stand basal area has become stable, i.e., it approaches the upper threshold of 36.37 m^2 ha^{-1}.

In the case that a logging event occurred, there is a clear impact on the growth of the forest stand (Figure 5.14). Given a forested patch following the logging of a basal area of 5.8 m^2 ha^{-1}, in the 32*th* year, the growing stock is actually reduced by 9.8 m^2 ha^{-1} due to an additional immediate damage of about 4.0 m^2 ha^{-1}. Then, severe tree mortality induced by logging occurs from year 32 to year 35, creating a slight concave up ward bend in the yield curve for this short period. Later years are the recovery period of the forest stand. Due to the impacts of the logging event, the growth of the forest stand is set back by 24 years compared to the non-disturbed growth. The overall impact of logging is visualized as the dotted area in Figure 5.14.

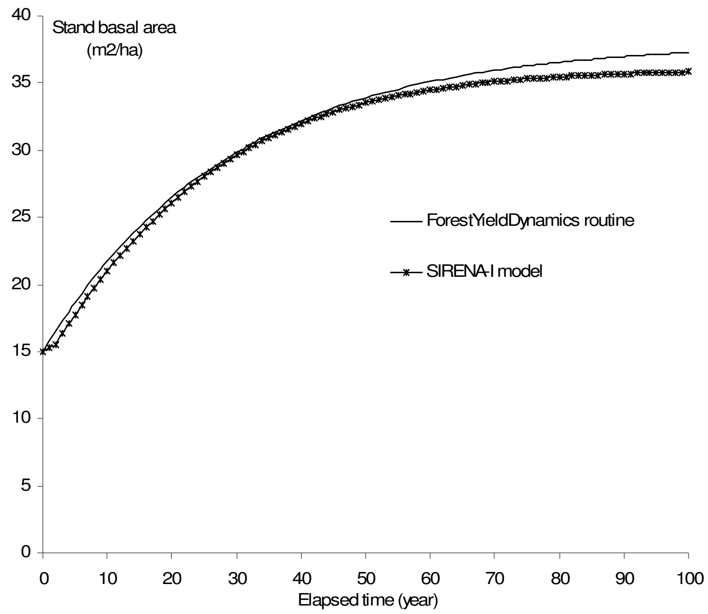


Figure 5.15 Simulated results for natural forests using *ForestYieldDynamics* and SIRENA-I, both without logging. Note: Functions of SIRENA-I model taken from Alder (1996b)

The comparison of the simulated results using the *ForestYieldDynamics* sub-model and the SIRENA-I (Alder, 1996b) model is shown in Figure 5.15. It is interesting that the two models, which have been developed using different approaches of forest growth modeling and are presented with different functional forms, give the same patterns of stand basal area over time. Since the SIRENA-I was built on growth increment data measured periodically from permanent plots (see Alder, 1996b), the good fit showed in Figure 5.15 validated our *ForestYieldDynamics* sub-model and its specified parameters to a certain extent.

Yield dynamics of forest plantation

Figure 5.16 shows the projection curve for the stand basal area P_{Gr} of *Acacia* forest plantations in the Hong Ha commune using the *ForestYieldDynamics* sub-model. Because the range of stand age of all 17 surveyed forest plantation plots is too narrow (see Figure 5.16), it is not possible to validate the projected plantation yields against time-series observed data through regression or correlation analysis. Alternatively, we compare the projected stand basal area at the average age of the surveyed plots to the average stand basal area.

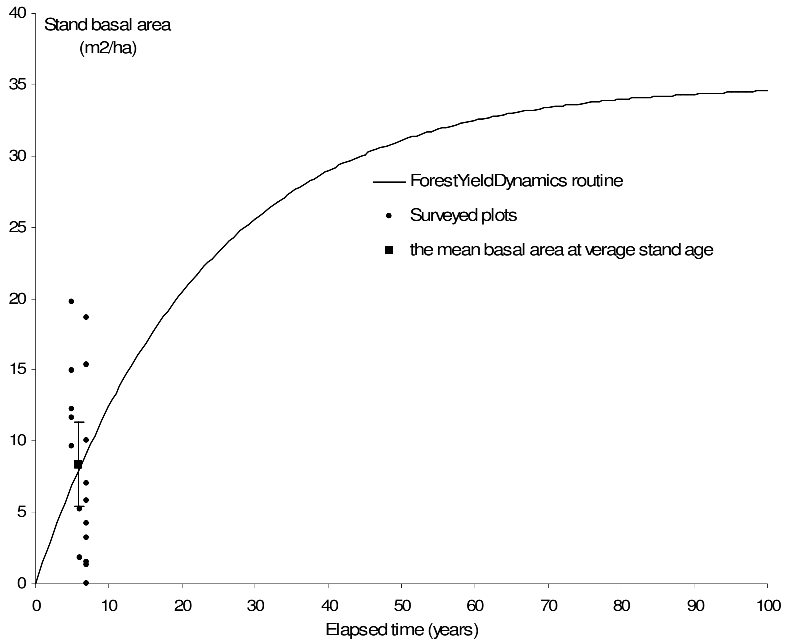


Figure 5.16 Projections for basal area dynamics of *Acacia* forest plantations in Hong Ha, using *ForestYieldDynamics* sub-model, compared with observed stand basal area. Note: The vertical error bar of average observed stand basal area is measured by confidence interval of the mean at 95% level ($CI_{0.05}$). Data source: FOF inventory 2002

The average stand basal area is 8.37 ± 2.93 m^2 ha^{-1}, and the average age of these plots is 5.88 ≈ 6 years-old (see Table 5.6). The basal area projected by the *ForestYieldDynamics* sub-model in the 6*th* year is 8.09 m^2 ha^{-1}, which is very close to the mean of the observed basal area at this time point. This fact gives only a modest validation of the *ForestYieldDynamics* sub-model and its parameters specified for the *Acacia* forest plantations.

5.4.4 Calibration of the *NaturalTransition* sub-model

The threshold value of stand basal area (P_{Gr}) for the transitional rule *N1* (see Figure 5.6), i.e., transitions from the dense/rich natural forest (P_{cover} = 6) to open/poor natural forest (P_{cover} = 7) and vice versa, is specified by the mid-point of the two mean values of the stand basal area of the two forest types (see Table 5.6):

$$\theta_{rich\text{-}forest} = (32.94 + 18.28)/2 = 25.61 \qquad (m^2\ ha^{-1})$$

Interviews with local farmers and forestry officers confirmed that if the shrubland patch is not farther than about 20 - 30 m, from the edge of the natural forest and there is no disturbance, it takes about 7-9 years to become open/poor natural forest. Thus, the threshold parameters of the transition rule N2 are specified as follows:

$$\theta_{d\text{-}forest} = 1 \qquad \text{(pixel lengths)}$$

$$\theta_{t\text{-}forest} = 7 + random_{int}(2) \qquad \text{(years)}$$

where the function $random_{inf}$ (2) returns randomly an integer number within [0,2], i.e., 0, or 1, or 2.

Similarly, if a grassland patch is not farther than about 20-30 m from the natural forest edge, it takes about 1–3 years to recover to shrublands. Hence, the specified threshold parameters of the transition rule *N3* are:

$$\theta_{d\text{-}forest} = 1 \qquad \text{(pixel lengths)}$$

$$\theta_{t\text{-}shrub} = 1 + random_{int}(2) \qquad \text{(years)}$$

The FOF inventory data of 21 survey plots for *Acacia* plantations in Hong Ha show that a young *Acacia* plantation takes about 3 - 4 years to become a forest plantation (see Table 5.6). Thus, the parameter of the transitional rule *N4* is:

$$\theta_{t\text{-}plantation} = 1 + random_{int}(1) \qquad \text{(years)}$$

The *NaturalTransition* routine was built into the architecture of VN-LUDAS on the NetLogo platform, and functions in concert with the functioning of other routines that influence or interrupt the natural transitions of cover types. The performance of this routine alone, i.e., without any disturbance/intervention through human agents, can be visualized using the VN-LUDAS model and switching off the functions of all human agents.

5.5 Conclusions

The landscape characterization at patch level delineates spatial patterns of landscape variables that are relevant to human-environment interaction studies, including proximities for soil-water distribution, spatial accessibility to transportation, and water bodies and land cover. The spatial depiction of topographic variables (i.e., slope gradient, upslope contributing area, and wetness index) show a heterogeneous spatial pattern of topography in the study area. Since topography is considered a fundamental driving force that regulates landscape distributions of soil and water, this terrain heterogeneity creates a patchy environment of natural land-use suitability. In addition, spatial heterogeneities are also the inequalities among locations in accessing to road and river/stream networks. Since human agents make use of these spatial variables in arriving at their land-use decision, these spatial heterogeneities, in concert with the diversities of the human agent profile and categories, result in the complexity of LUCC.

Land-cover classification has revealed cover categories that are similar to many remote sensing-based classifications of land cover in the Vietnam uplands (see Fox *et al*., 2000; Sadoulet *et al*., 2002; Castella *et al*., 2002c; Fatoux *et al*., 2002; Alther *et al*., 2002; and Zingerli *et al*., 2002). The delineated spatial pattern of land covers over the study area shows a recognizable spectrum of declining forest successions along transects across village centers: dense forests → degraded forests → shrubland →

grassland → cultivation land/plantation. The remaining natural forests are degraded and highly fragmented. In a non-linear dynamic system such as the VN-LUDAS model, land cover at a particular time is treated not only as a response variable to the variation of a range of drivers, but also as an explanatory variable affecting land-use decision or land-cover status in subsequent time points. This path-dependency nature of the changes implies that careful classification of initial land cover, such as the land-cover map 2002 in our case, could be an important calibration for bringing the realistic landscape into the MAS system.

The *AgriculturalYieldDynamics*, a sub-model built into landscape agents, which were developed in the form of empirical bio-economic yield models, are able to perform the non-linear combinational responses of landscape agents along three dimensions of heterogeneity: i) spatial heterogeneity, ii) temporal dynamics, and iii) human agent diversity. Spatial heterogeneity of agricultural yield responses is represented by variation of site productivity, which is abstracted in terms of slope and upslope contributing area, following an indirect geocentric view (Leary, 1985). In contrast to paddy rice, the yields of upland crop and fruit-based agroforestry are highly variable along the patchy pattern of the slope gradient and thus heterogeneous over space. Temporal dynamics of agricultural yield are an indirect representation of the trend either of soil fertility decline in the case of upland crop fields, or of fruit crop/tree growth and development such as in the case of fruit-based agroforestry farms. Heterogeneities of agricultural yields along the household agent diversity is represented by agrochemical and labor inputs, which are specific for every human agent. The combination of these three-dimensional heterogeneities with the yield function results in extremely complex patterns of actual yield responses in time and space.

In general, all selected explanatory variables, except upslope contributing area, (i.e., labor and agrochemical inputs, slope gradient, and cropping time) influence agricultural yields in directions as theoretically expected. The unexpected effect of upslope contribution area on agricultural yield may be due to the fact that either the coarse spatial resolution of its spatial dataset (30 m × 30 m) may not reflect finer variations of this variable among plots, or there may be errors associated with the topographic map. However, the patterns of yield responses and the set of significant drivers are different for different farming types, depending upon the nature of cropping systems. Cultivated on flat land less prone to soil degradation, paddy rice yield is mainly determined by labor and agrochemical inputs, thus the yield dynamics are mainly influenced by the intensification decisions made by the household agents. Fruit-based agroforestry yield is sensitive to labor input, and increase with the growth of the fruit crop/tree components. Thus the yield dynamics are affected by human agent's decision that probably have a long-term perspective. The response pattern of upland crop yield has been found to be the most complex, as it varies across the three dimensions: heterogeneous space (viz. slope gradient), diverse human behavior (viz. household's decisions on labor and agrochemical inputs), and elapsed time (viz. cropping time). Thus, the estimated function of upland crop is one of a typical representation for heterogeneous landscape dynamics interacting with the human system.

Besides representing in part heterogeneities of landscape dynamics, the sensitivity analyses for agricultural yield responses give a number of useful suggestions for agricultural production in the study area. For upland crop farms, evidence of low yield elasticity with respect to agrochemical input, and rapid yield decline associated with decreasing yield with slope increment support the statement that the mountainous hillsides are rather marginal in terms of potential for food production (Castella *et al.*, 2002b; Gomiero and Giampietro, 2001). This implies the need for more efficient management for crop production on the hillsides in the study area. The marginal response of paddy rice yield to the increment of agrochemical input in conditions of lower environmental risks indicates the limits of intensification through increasing agrochemical input. This finding suggests that other management alternatives, e.g., new varieties, are probably needed. The fact that labor input is a major constraint for the yields of all three farming types suggests that labor allocation strategies could play an important role in maximizing agricultural benefits.

The *ForestYieldDynamics*, a sub-model built into landscape agents, was developed to perform assessment of forest yield dynamics in response to the vegetative condition of the site (viz. previous stand basal area) and human disturbance (viz. logging activities), and thus links natural and human system dynamics. In the model, the site factor is indirectly represented using stand basal area at the previous time point, thus following a phytocentric view in site evaluation (Leary, 1985). Although the natural basal area is conceptually an expression of site productivity (Vanclay, 1994), it is still considered here as a constant parameter rather than as a spatial variable of the model. However, although the sub-model is theoretically developed, the use of the model with careful setting of a few model parameters on the basis of literature review or extensive inventory data can return acceptable results comparable to an empirical model. The model is also able to show the impacts of logging activities. As the sub-model is theoretically based and simple with a few input parameters, the main advantages of the model are: i) applicable with even poor forest growth data, ii) capable of coupling regular natural growth dynamics with intervention by human agents, and iii) probably more reliable for predictions that involve extrapolations beyond the range of empirical data.

The *NaturalTransition*, a sub-model built into landscape agents, performs annual natural vegetation succession in ways that are beyond the control of human agents. Through this routine, vegetation covers can evolve following natural succession rules without intervention by human agents. Transitions among natural forest are made based on the decision whether the forest yield (i.e., stand basal area) of the landscape agent exceeds the thresholds, which were calibrated based on forest inventory data. In other words, the routine performs both the modification of forest covers (i.e., gradually progressive recovery) and conversions (i.e., discrete transitions among forest cover types). Transitions among non-forest vegetative covers take place based on the evaluations of the life span of the cover and the neighboring natural forests, and are thus both path and neighbor dependences. Therefore, this transition mechanism is also a conversion on the basis of gradual vegetation growth. Moreover, the progressive natural recovery processes, either natural growth of existing forests or regeneration of non-forest vegetation types, can be disrupted by human agent and quickly converted to another state.

Building the two sub-models *ForestYieldDynamics* and *NaturalTransition* into the landscape module make the VN-LUDAS able to capture both modification and conversion in LUCC. Many previous LUCC models assumed that land-cover changes consisted of mainly conversion of pristine forest to agricultural uses (deforestation) or destruction of natural vegetation that led to desert conditions (desertification) (Lambin *et al*., 2003). Our consideration of vegetation growth in modeling LUCC also coincides with the new shift in the understanding of LUCC processes from a reactive view, which critisizes human impacts on the environment as mostly leading to a deterioration of the earth system's processes, to a proactive view, which emphasize the potential for ecological restoration through management (Victor and Ausubel, 2000). In general, if the forest ecosystem is disturbed/harvested within its resilience limit, which is normally associated with an appropriate forest management, the ecosystem can maintain its structure and functions to ensure its goods and services for human communities.

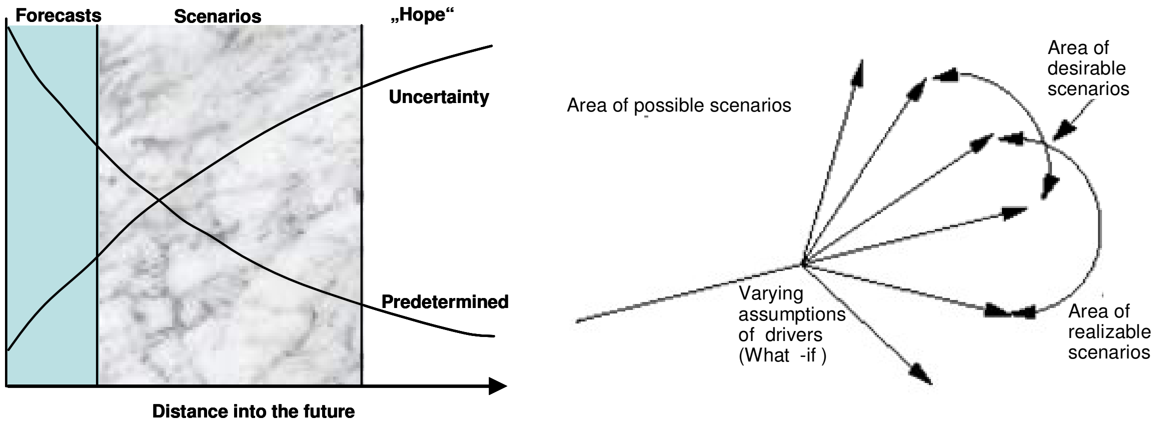
In sum, through building the three sub-models as specified and calibrated above into landscape agents, we have represented the landscape environment in a dynamic, adaptive, and realistic manner. The dynamic environment is the landscape agent, which has natural processes operating in it, changes in ways beyond the human agent's control (Woodridge, 1999). The adaptive environment is one where the constituent units (i.e., landscape agents) have specific capabilities to interact with and respond to the changes of the surrounding environment, including human agents. The realistic environment is where the state and behavior parameters of the landscape agents are empirically grounded on real environmental data.

6 INTERGRATED SCENARIOS OF LAND-USE/COVER CHANGES AND IMPACT ASSESSMENT OF LAND-USE POLICIES IN HONG HA WATERSHED

6.1 Introduction

To provide a knowledge base for effective discussions and informed decision-making in proactive land management and planning, recognition of the wide range of future outlooks of the coupled human-environment system is a key issue (Raskin *et al.*, 2004). These future outlooks can be derived through scientific experimentation, careful observation and feedback (Wollenberg *et al*., 2000). If the system under consideration is simple and causal interactions are predetermined, the future performance of the system could be a straightforward prediction, which presumes that the future can be derived from the monitoring and analyses of actual histories. When derived from past data, a prediction *totally* depends on *structural inertia* of the actual history, and thus providing a *single-line future* in the short-term. Unfortunately, because the dynamics of the coupled human-environment system are inherently complex and uncertain in time and space (see Chapter 1), such a deterministic prediction becomes problematic. Moreover, as sustainable land management requires long-term perspectives, the associated risks and uncertainty become rather high, thus reducing the prediction power (Raskin *et al*., 2004; O'Brien, 2001; Wollenberg *et al*., 2000) (see Figure 6.1a). Also, as single-line prediction does not introduce alternatives, the approach does not encourage creative thinking and rational choices in land management and planning.

Alternatively, a scenario-based approach has been recognized as a natural and powerful way for advancing the problem of viewing the system's future in the face of high complexity and uncertainty. By definition, a scenario is a description of a *hypothetical future situation* and the *course of events*, which allows one to move forward from the original situation to the future situation, with the purpose of focusing attention on *decision points* (see Godet and Roubelat, 1996; Kahn and Wien, 1967 cf. Raskin *et al*., 2004). Unlike classical predictions, a scenario is not necessarily an accurate forecast of a likely single future drawn on past data. Instead, scenarios are *multiple possible future pathways* of the system evolution under a spectrum of conditions that are hypothesized as sources of risks and drivers of change (Maack, 2001; O'Brien, 2001). Through generating several possible scenarios in consistent and plausible ways, the scenario-based approach enables people to strategically navigate and communicate different visions and to provide a focus for discussion about the course of society (Figure 6.1b). The approach, therefore, opens up possibilities for a more critical understanding of social, economic and environmental impacts of human actions and creative thoughts of policy alternatives, thus supporting stakeholders through more informed and rational decision-making (Wollenberg *et al*., 2000; Raskin *et al*., 2004).



(a) (b)

Figure 6.1 (a) Forecasts and scenarios have different arenas, depending on the level of uncertainty. Sources: after Van der Heijden (1998) and O' Brien (2001)
(b) Scenarios as tools for scanning the uncertain future of a complex system. Source: modified from Godet and Roubelat (1996)

Sustainable management of land resources requires *integrated scenarios*, which characterize possible *co-evolution pathways of combined human and environmental systems* with a balance between the narrative and quantitative forms/methods (Kemp-Benedict, 2004a; Raskin *et al*., 2004). The scientific insights for such integrated scenarios depend on how well the human-environment interactions are understood and used for the development of the scenarios. For people seeking to use scenarios in land-use management and planning, the range of methods available has been limited due to the inadequate representation of the human-environment interrelationships in the methods. At one extreme, purely LUCC scenarios built on spatial predictive land-use models developed by natural scientists often ignore the explicit roles of human actors in the changing of landscapes, thus being very weak in linking and transforming environmental anticipations into human actions (Huigen, 2004; Veldkamp and Verburg, 2004). At another extreme, scenarios developed by many bio-economic models tend to treat the biophysical world as consistent drivers only, thus being weak in assessing environmental impacts of human actions at landscape or larger scales (Verburg *et al*., 2002; Jansen and Stoorvogel, 1998; Kaimowitz and Angelsen, 1998). Multi-agent simulation (MAS) modeling has been recognised to be well suited to exhibit the co-evolution of the human and landscape systems based on the interactions between human actors and their environment. However, the state-of-art of MAS is still too crude for applied work in terms of its capabilities to represent the real human-landscape system at an operational level (see Chapter 1).

The qualification of the scenario-based studies lies in the *forms* that describe the scenarios and the *processes* that generate them. Scenarios can take either narrative forms, which are generated using the shared mental models of their authors, or quantitative forms, which is developed using quantitative formal models. Each scenario form and its generating method have its own merits: the narrative provides texture, richness and insight description (i.e., descriptive tradition), while the quantitative offers structure, rigor, as well as replicability, and transferability of interpretation (i.e., interpretive tradition) (Raskin *et al*., 2004). An efficient scenario study should, therefore, offer ways to integrate the narrative and interpretive traditions in a particular balance (Kemp-Benedict, 2004a). From a learning perspective, real limitations in scenario studies are the imagination of the people using scenarios and their interest to participation in creating them (Wollenberg *et al.*, 2000).

To construct integrated scenarios satisfying the requirements above, support is needed from integrating computer tools, i.e., so-called decision support systems (DSS). A DSS should have at least three major abilities, namely: i) *simulation modeling* (i.e., the scientific reasoning capability to represent the system concerned and to generate integrated scenarios), ii) *visualization* (i.e., the capability to illustrate scenarios), and iii) *communication* (i.e., the capability to share/transfer scenarios information and enable stakeholders to jointly improve learning about the consequences of actions) (see Orland *et al*., 2001). Multi-agent simulation models have been recognized to be well suited for representing the coupled human-environment system and anticipating its future (see Chapter 1). *Static visual formats* - such as temporal calibrated maps, time-series graphs

- obviously combine the strengths of both narrative and quantitative forms, thus being more powerful and efficient tools for communication than a strictly verbal or numerical format. More powerful than static visualization, the *calibrated virtual landscapes* may greatly extend the ecological and social validity of system representations and provide more meaningful and insightful feedbacks of the real world based on human experience (Orland *et al*., 2001). Unfortunately, as far as we know, no operational agent-based DSS for tropical forest margins in developing countries exists that satisfy the desired characteristics.

Given the VN-LUDAS theoretical framework specified in chapter 3, empirical case studies for the human system (in chapter 4) and the natural landscape system (in chapter 5), in this chapter we aim to develop and implement an *operational* VN-LUDAS with functionalities of a modest DSS for land-use policy decisions in the Hong Ha watershed or other similar areas. The chapter has three specific objectives:

i) to identify puzzle decision points of particular land-use policies in Hong Ha and policy interventions that should be tested for their likely impacts,
ii) to develop an operational VN-LUDAS model with functionalities of a DSS to support impact assessments of the tested policies, and
iii) to simulate integrated scenarios of the coupled human-landscape system using VN-LUDAS, thereby assessing the impacts of policy changes and identifying potential policy alternatives.

From a methodological point of view, the chapter is expected to illustrate how a DSS can be built based on a calibrated multi-agent simulation model and recent advances in computer sciences, as well as how the agent-based DSS is used for generating integrated scenarios of the coupled human-landscape system to support informed land-use policy decisions.

6.2 Land-use policies in Hong Ha: overall setting and puzzle decision points

Like many upland watersheds in central Vietnam, the Hong Ha watershed has been subject to both national land-use polices: protection of remaining forest resources and promotion of agricultural production. Being located on the head-water of the Bo River, i.e., important water body in the northern Thua Thien-Hue province (see Figure 4.1, Chapter 4), in 1991 the watershed was declared a protected area. According to the introduced protection policy, natural forest lands are protected from deforestation activities, and sloppy lands are designated for planting protective forest. Concurrently, policies supporting agriculture development, such as agricultural extension and agrochemical subsidy schemes, have been adopted to increase agricultural production and compensate for the loss of forest-based income. In general, these land-use policies are *expected* to encourage a structural change in upland production systems from forest-based and extensive to an agriculture-based and intensive system, thus increasing household income while reducing the population pressure on forest resources.

However, the expected transition is harmed by an ambiguous and uncertain knowledge base about human-environment interrelationships in highly diverse socio-ecological conditions. While most government organizations consider the uplands to still have a great potential for agricultural production and a new frontier of agricultural development, many environmental scientists perceive that the upland ecosystems are fragile, already degraded and marginal for agricultural investment (Jamieson *et al*., 1998; Rambo, 1995). There also are still disagreements about deforestation causes in Vietnam (Morrison and Dubois, 1998). The trade-off effects of the changes in agricultural development or forest protection policies are also largely unknown (Castella *et al*., 2002d). From a methodological view point, the extreme diversity of the mountain regions and the lack of empirical time-series data cause great difficulties for policy impact assessment using traditional comparative analysis with different case studies.

Also, the formulation of operational land-use policies in Hong Ha is facing great difficulties due to the different values of stakeholders regarding land use and management in Hong Ha. Stakeholder analyses and focused group discussions[17] show a range of different mandates, contrasting interests and expectations of main stakeholders in the area. Governmental bodies for forest protection, i.e., Bo River Forest Protection and A-Luoi Forest Inspection Division, claim all sloped land for forest protection excluding local community uses. Agricultural development organizations, i.e., A-Luoi Agricultural and Rural Development Division and Provincial Department of Resettlement, and communities want to keep the hill sides for agricultural production.

[17] Stakeholder analyses and focus group discussions were done as parts of participatory processes under a project titled "Community-based upland natural resource management", 1998-2001, of which the author was a project research member, as part of the field research for this doctoral study (September 2002 to July 2003).

Moreover, while the local communities expect to be subsidized as much as possible, the agricultural development organizations seek to optimize subsidies and extension for agriculture.

Viable policies that resolve these value conflicts must be developed by interpretation and applications of scientific information in light of the needs and priorities of the stakeholders. There is no doubt that scientific information alone cannot provide all answers to stakeholders' questions as science necessarily is silent on the human value underlying the decision stakeholders make. However, scientific tools and information are very useful for helping people to develop options and to understand and evaluate consequences for policy actions, thus creating a basis for discussion to reach consensus. Although land users/managers in Hong Ha initially negotiate to improve the management of the watershed, a key obstacle for efficient multi-stakeholder negotiations is the lack of a feedback tool to scientifically visualize integrated scenarios of policy choices.

It is impractical to test all possible policy scenarios. Thus, it is necessary focus on particular policy issues of local concerns (i.e., use cases). Through interviewing local key informants and organizations, three policy issues for scenario development were identified: i) forest protection zoning, ii) the spreading of agricultural extension services, and iii) the extent of agrochemical subsidy in the subject communities.

6.2.1 Forest protection zoning

Chapter 2 of the National Technical Codes for Watershed Protection Planning (QPN-13.91), after the Decision No. 134-QD/KT 1991 issued by the former Ministry of Forestry (now merged into the Ministry of Agriculture and Rural Development – MARD), defines principles and criteria for defining watershed forest protection zones. The regulation shapes a watershed into zones for different land-use management based on *critical* levels in watershed protection: Level I: highly critical for watershed protection, Level II: critical, Level III: less critical. Land-use planning and management in the protected watersheds, regulated by laws, are different for each critical level. In most cases, areas classified into Level I or II are protected from logging or vegetation clearance, while collection of non- timber forest products (NTFPs) are allowed. Areas categorized into Level III can be used for agricultural production or other non-forest land uses. Based on that legal framework, the Forest Inventory and Planning Institute of Vietnam (FIPI) was assigned to develop a technical scoring system as a technological basis to define the critical levels.

Table 6.1 Look-up table for calculating the watershed protection score proposed by the Institute for Forest Inventory and Planning of Vietnam (FIPI). Source: Le Sau and Tran Xuan Thiep (1997)

Factor	*Potential contribution to land degradation*	*Characteristics*	*Score*
Annual rainfall (M)	M_1 (high)	• 2000 mm/year, or • 1500 – 2000 mm/year with an uneven distribution	6
	M_2 (medium)	• 1500 – 2000 mm/year, or • 1000 – 1500 mm/year with an uneven distribution	4
	M_3 (low)	< 1000 mm/year	2
Slope factor (α)	α_1 (high)	> 35°	6
	α_2 (medium)	25° - 35°	4
	$\alpha_{3.1}$ (low)	15° - 25°	2
	$\alpha_{3.2}$ (very low)	8° - 15°	1
	$\alpha_{3.3}$	< 8°	0.5
Soil factor (D)	D_1 (high)	• Sandy soil, < 80 cm depth, or • Sandy loam or silty loam, < 30 cm depth	3
	D_2 (medium)	• Sandy soil, > 80 cm depth, or • Sandy loam or silty loam, 30 – 80 cm depth, or • Clay loam or clay, < 30 cm depth	2
	D_3 (low)	• Clay loam or clay, > 30 cm depth, or • Sandy loam or silty loam, > 80 cm depth	1
Elevation factor (C)	C_1 (high)	2/3 H_{max} [(a)]	3
	C_2 (medium)	1/3 – 2/3 H_{max}	2
	C_3 (low)	< 1/3 H_{max}	1

[(a)] H_{max} is the relative altitude of the highest point in the study watershed.

The FIPI's score criteria include four main physical factors: rainfall (3 classes), slope (5 classes), relative elevation (3 classes), and soil physical conditions (3 types). Details of the scoring system are given in Table 6.1. The scoring system based on these criteria assigns a land patch in a watershed a sum of scores that ranges from 4.5 (minimum) to 18.0 (maximum).

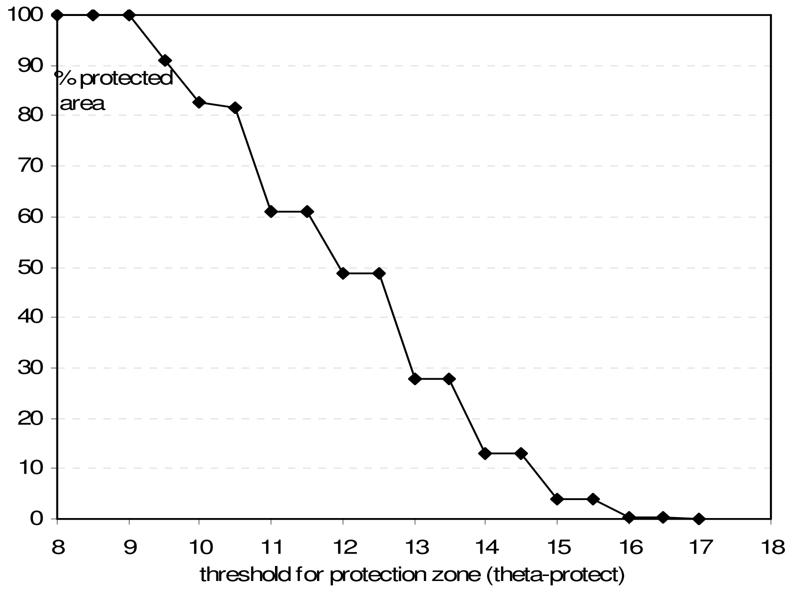


Figure 6.2 Percentage of protected area in Hong Ha watershed vs. zoning threshold ($\theta_{protect}$). Data source: calculated from VN-LUDAS

By tuning the zoning score threshold for land protection, planners can control the extent of the protection zone (Figure 6.2). The shifting of these thresholds by a few units can create substantial land-use changes at the landscape level and subsequently affects livelihoods of local communities, who are land and forest dependent (see Figure 6.2). Govermental organizations who are responsible for watershed protection always tend to reduce the zoning threshold to include a larger area for protection, whereas local communities want to increase the threshold to have more space for agricultural production and harvesting timber. Therefore, the critical debating point of zoning policy is: what should the *classification threshold* be in order to define the appropriate extent of protected area.

In prospective watershed planning and management, the identification of these thresholds must be done through participatory processes that involve many stakeholders. During negotiation processes, stakeholder groups may like to see, spatially explicit and in "*real-time*", how the changes of protection zoning rules affect the dynamics of the environment and local livelihoods. These explorative trajectories provide a basis to assist participatory processes to obtain a consensus. Thus, it is important to include the zoning factor in the global module of the model.

In VN-LUDAS, two global parameters were needed for quantitatively representing the zoning policy factor: i) the score threshold ($\theta_{protect}$) for deciding whether a land patch is protected or not, and ii) the enforcement coefficient ($\theta_{enforce}$) which reflects the probability that the introduced protection rule is enforced. The parameterization of this policy factor within VN-LUDAS modeling framework was described in details in Chapter 3.

6.2.2 Agricultural extension

Like many rural communes, Hong Ha has been subject to many agricultural extension projects/programs/schemes. Agricultural extension services in Hong Ha mainly include the provision of technical guides for livestock and crop production through household field visits and on-farm training. Access to agricultural extension services directly affects the land-use decisions of upland farmers (see Chapter 4). Farmers with access to extension schemes may change their attitude in adopting agricultural land-uses or may have a better opportunity for intensifying land use. Thus, their land-use decision space may be changed accordingly.

The question for extension policy is how the change in the extent of agricultural extension affects land use/cover and household economy. Focused discussion with extensionists in the area revealed that they want to know the optimal fraction of the community that can be reached by agricultural extension given their available human resources. The answer to this question will enable extension centers to *balance the quantity and the quality* of their extension services. Given the human resources constraint, if the extensionists have to cover a large proportion of the population, they will not have enough time for quality technical guidance for each target household. In contrast, limiting extension services to a small proportion of the population will enhance the quality of the guidance, but the overall positive effects on the whole community may be not high. Also, in the latter case, the inequality in access to agricultural extension services will be higher.

In VN-LUDAS, the extension policy factor is approximated by the percentage of farming households that have access to extension services (θ_{exten}) as expected or planned by agricultural extension schemes. The parameterization of this policy factor within the VN-LUDAS modeling framework is described in detail in Chapter 3.

6.2.3 Access to agrochemical subsidy

In the development policy for the Vietnam uplands, the government subsidizes industrial fertilizers and pesticides to encourage poor upland farmers to intensify their agricultural production. The aim is to stabilize the upland livelihoods and to reduce the pressure on forest resources. In general, empirical analyses for the study area show that poor farmers all expect to be supplied with subsidized agrochemicals for crop production (see chapter 4). However, the effects of subsidies on agricultural productivity on a larger scale, on household income and its distribution are largely unknown. Very likely, many rural developers and scientists in the region want to see the potential trade-off impacts of a change in subsidy access on land use in the forest margin zone, since this relates directly to the policy assumption of many governmental organizations. Thus, the VN-LUDAS needs to explore how a change in the subsidy factor affects agricultural productivity, household income and income distribution over the community, as well as land-use and land-cover change.

The agrochemical subsidy factor is approximately represented by two parameters: the percentage of farming households who have received the subsidy ($\theta_{subsidy}$), and the subsidy amount ($W_{subsidy}$) as expected or planned by the governmental subsidy program. The parameterization of this policy factor within the VN-LUDAS modeling framework was described in details in chapter 3.

6.3 Methodology

6.3.1 Defining tested land-use policy interventions

Given the policy factors and their parameters, policy interventions for simulations were systematically defined as follows:

- The current policy setting (as in 2002) is used for developing a *baseline* scenario, or *likely current trend*. This scenario will be used as a baseline for evaluating the impacts of changes in land-use policies.
- Given the current policy setting, each single policy factor will be shifted from the baseline to form a *scenario spectrum* of the considered policy

factor. Other non-experimental policy factors are kept the same as that of the current setting. To avoid an overabundance of tested scenarios, each scenario spectrum, including the baseline, consists of 3 or 4 scenarios only.

- Based on the simulation results for single scenarios defined in the previous step, about 1 or 2 *expected combinational scenarios* will be developed by combining promising single interventions recognized in the previous steps.

The different policy interventions of each scenario spectrum are briefly described below:

Baseline/current scenario: The current trend

The current/baseline scenario (S0) has the policy setting as in the situation 2002. The current policy for forest protection zoning in Hong Ha basically follows the zoning rule proposed by FIPI. The zoning rule is that a land patch with a zoning score > 9.5 will be protected. The total protected area in these scenarios occupies 90 % of the total area. Interviewed local officers and inhabitants estimated that the enforcement degree of the protection regulation is around 50 %. Descriptive statistics of the sampled households (69 households) show that 67 % of the total households were reached by extension services, 23 % households received subsidized agrochemicals amounting to 260 VND household^{-1} year^{-1} (i.e., 16 USD household^{-1} year^{-1}) (see Table 6.2). It should be recognized that Hong Ha is a community receiving ample supports from governments and projects compared to other upland communities in the Thua Thien – Hue province.

Scenarios for assessing the impacts of changes in forest protection zoning

The scenario spectrum for scanning the effects of forest protection zoning policy includes the three following scenarios. Scenario S0 (*baseline*) as described as above reflects the current protection status in Hong Ha: the governmental forestry agencies claim a very large area for protection, but lack the necessary resources and capacities to ensure the allocated tasks, resulting in poor enforcement of the protection rule (only about 50 %). Scenario S-Pro0 (*no protection*) assumes there are no restrictions on forest use by setting the zoning threshold to a score of 16.5. Scenario S-Pro3 (*strict forest protection*) assumes the government invests more man power and other resources to ensure good enforcement of the protection rule (about 80 %); the zoning threshold is still the same as for the baseline case.

Scenarios for assessing the impacts of changes in agricultural extension

The scenario spectrum for exploring the impacts of agricultural extension level includes the 3 following scenarios: S-Ext0 (“*no/little extension*”), S-Ext1 (“*low extension*”) and S0 (baseline – “*high extension*”). This translates into an extent of agricultural extension of 5 %, 35 % and 67 %, respectively, with an interval of change in extension expansion level of about 30 % of the population.

Scenarios for assessing the impact of changes in agrochemical subsidy

Because the land-use choice behaviour of households in Hong Ha is quite sensitive to changes in agrochemical subsidies (see Chapter 4), the scenario spectrum for assessing the impacts of the subsidy policy was 4 scenarios, with a finer interval of subsidy level (i.e., about 25 % of the population is subsidized). The spectrum of scenarios S-Sub0 (*no subsidy*), S0 (*baseline - low subsidy*), S-Sub1 (*medium subsidy*), and S-Sub2 (*high subsidy*) expresses a gradual change in the subsidy coverage level from 0 to 23, 50, and 75 % of the population, respectively.

Expected combinational scenarios

As the policy factors are parameterized using continuous scales, the combinations of such parameterized factors will result in a large number of combination sets, which are unnecessarily complicated. To avoid scenario abundance, advantages of scenario results for single policy issue was taken in the first step. Through comparative analyses of scenarios within a scenario spectrum of a single policy factor, the most positive scenarios with respect to the improvement of environmental quality and household income are selected. Then, such positive policy scenarios, possibly with some additional adjustments as needed, are combined to form 1 or 2 combinational scenario(s) for assessing the impacts of multiple changes in policies (see Table 6.2).

Table 6.2 Policy settings for developing integrated land-use scenarios

Scenario	Quantitative setting of policy factors: Protection zoning – Protection zoning threshold ($\theta_{protect}$)	Protection zoning – Enforcement degree ($\theta_{enforce}$)	Agricultural extension – Percentage of the population reached (θ_{exten})	Agrochemical subsidy – Percentage of the population subsidized ($\theta_{subsidy}$)	Agrochemical subsidy – Subsidy amount ($W_{subsidy}$)	Illiteracy eradica- -tion rate (θ_{edu})
Unit	FIPI score	%	%	%	1000 VND /household /yr	%
Current scenario (as policy setting in 2002):						
S0 (current trend)	9.5	50	67	23	260	10
Scenarios for exploring the impacts of changes in forest protection zoning:						
S-Pro0 (no protection)	**16.5**	50	67	23	260	10
S-Pro3 (strict protection)	9.5	**80**	67	23	260	10
Scenarios for exploring the impacts of changes in agrochemical subsidy:						
S-Sub0 (no subsidy)	9.5	50	67	**0**	**0**	10
S-Sub1 (medium subsidy)	9.5	50	67	**50**	260	10
S-Sub2 (high subsidy)	9.5	50	67	**75**	260	10
Scenarios for exploring the impacts of changes in agricultural extension:						
S-Ext0 (minor extension)	9.5	50	**5**	23	260	10
S-Ext1 (medium extension)	9.5	50	**35**	23	260	10
Scenarios for exploring the impacts of combinational policy changes:						
S-COM1 (combination 1)	**12.0**	**80**	**75**	**5**	260	10
S-COM2 (combination 2)	**12.0**	**80**	**35**	**5**	260	10

6.3.2 Developing an operational VN-LUDAS for policy decision purposes

We have developed an *operational* VN-LUDAS that has the basic functionalities of a DSS as shown in Figure 6.2. The figure shows the basic flows of information in scenario studies with a decision support system.

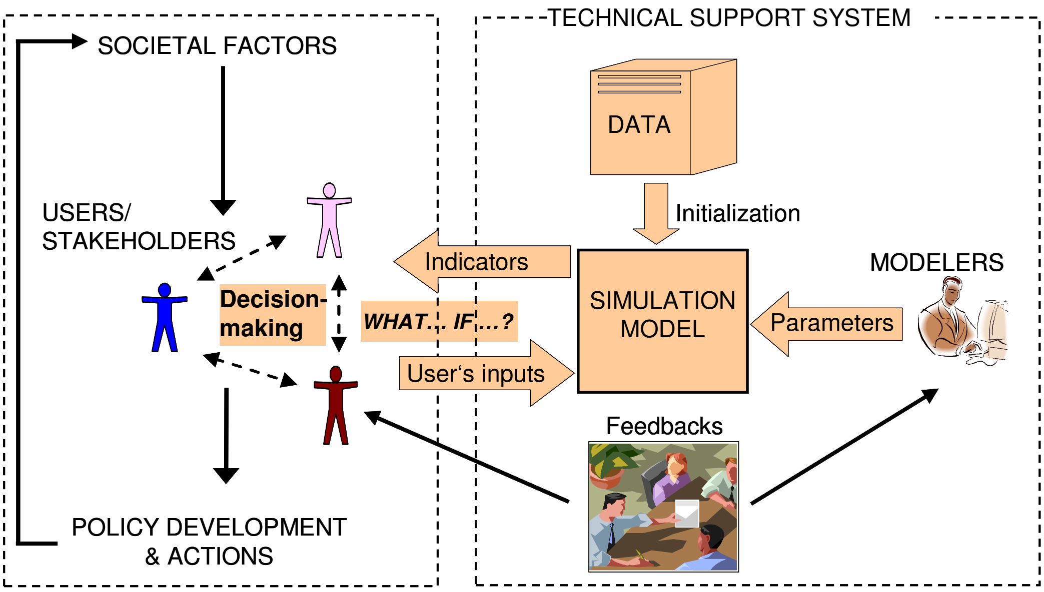


Figure 6.2 Main information flow in a scenario exercises and technical support system for policy decision-making processes. Sources: synthesized from Bernabo (1998) and Kemp-Benedict (2004)

VN-LUDAS computer program

The core VN-LUDAS framework, which was theoretically specified in Chapter 3, was programmed in the *NetLogo* 2.1 package. *NetLogo* (Wilenski, 1999) is a multi-agent modeling environment that allows programming intelligent agents, their interactions, and monitoring the connections between micro-level behavior of agents and macro-level patterns of the whole system. We programmed the theoretical VN-LUDAS framework in *NetLogo*, using object-oriented programming techniques. The program includes the main sub-models/procedures briefly described in Table 6.3. As all human and landscape agents and their built-in sub-models were programmed using a united language and platform that allow concurrent interactions, from a view point of modeling coupling the VN-LUDAS itself can be seen as a tight (close) coupling system of many ecological and bio-economic sub-models.

We iteratively verified the VN-LUDAS computer program at two levels: verifying every single sub-program/procedure and verifying the whole simulation program/protocol. Verification is the checking of the performance of the computer program/sub-programs, detecting bugs, syntax errors and improving the codes for better performance.

Table 6.3 Main sub-models/procedures[a] of VN-LUDAS programmed entirely with *NetLogo*

Name	Brief functionalities/tasks	Involved agent
Initialization	Import GIS data and sampled household data, generate the remain population, create household-pixel links (complex [b] procedure)	household patch
SetLaborBudget	Annually set the labor list of the household	household
FarmlandChoice	Perform agricultural land-use choices, including bounded rational choice, many rule-based decision algorithms (complex[b] procedure)	household patch
ForestChoice	Perform forest-use choices, mainly rule-based algorithms (complex procedure)	household patch
GenerateOtherIncome	Generate non-crop and non-timber incomes	household
UpdateHouseholdState	Annually update changes in household profile	household
AgentCategorizer	Annually categorize household into the most similar group	household
GenerateHouseholdCoefficients	Generate behaviour coefficients of household, allow variants within groups, but stabilize behavior structure of the group	household
ForestYieldDynamics	Calculate forest stand basal area in response to human interventions (logging)	patch
NaturalTransition	Perform natural transition among vegetation types based on accumulated vegetation growth and ecological edge effects	patch
CreateNewHousehold	Create a young new household, controlled by an empirical function of population growth	household
DrawGrapths	Draw different graphs of system performance indicators	household patch

[a] All computer program codes of these main sub-models/procedures as well as other auxiliary routines are written by the author, using *NetLogo* language. They are available from the author.
[b] A complex procedure here means that it contains one or more procedures.

Inputs

Inputs for simulations with VN-LUDAS include two types: calibrated data and parameters. Calibrated data includes spatial data (GIS format) and household data. Two sub-types of parameters can also be distinguished: modeler's input parameters and user's input parameters.

Calibrated data

Data for initializing the coupled human-landscape with VN-LUDAS include GIS data in forms of text files, and household data as worksheets. There are three important points related to the data issue with VN-LUDAS. First, because good-quality data are used to validate in part the MAS model, all data used by the VN-LUDAS must be calibrated and/or processed outside the model to adequately represent the reality of the coupled human landscape system. Second, the linkages between the spatial and household datasets are crucial. Third, although the more detailed and accurate the data, the more creditable the results; however, the relevance of the dataset and the data availability are also important. Methodologies for processing/calibrating/classifying data from different sources, organizing the nested household-pixel dataset, and scientific approximation of relevant data for use in the VN-LUDAS were discussed in detail in Chapter 4 and 5.

Parameters

Within the VN-LUDAS, two types of parameters are distinguished. *Modeler's input parameters* are inputs in the model that modelers do not expose to users or that are built into the computer simulation program/protocol. This type of parameter includes most of the technical coefficients that were extracted from quantitative analyses in the case studies in Chapters 4 and 5. *User's input parameters* are mainly policy parameters, which enable users to set their own policy options for scenario development. Some of the technical parameters that are uncertain to some degree (e.g. "*vision*" of household agents) and global parameters (e.g., annual rainfall) are handled as user's input parameters to take advantage of so-called *expert knowledge* from expert users. The distinction between these two parameter types is more or less blurry, dependent on how client-oriented we let the VN-LUDAS be.

Outputs

The strength of the MAS-LUCC in general and VN-LUDAS in particular is that it gives a very informative set of outputs. When running a simulation, at any point in time and space, the VN-LUDAS will give three main types of outputs: simulated world, predefined indicators, and graphics.

Simulated world

The simulated world produced by the model at a point in future is the world file (worksheet format) containing all numeric information of the whole system. With the helps of functionalities "*export*" and "*save*" world-files of the NetLogo platform, we can calculate any numeric indicator of the system, sub-system and individual agents as needed. Therefore, the simulated world files always enable experts to conduct sophisticated interpretations of the simulation outcomes.

Predefined indicators

Predefined indicators are some common and popular indicators of system/sub-system performance, which are calculated within simulation protocol to expose their "real-time" responses on the built-in digital maps and graphs using the graphic-user interface (GUI) of the VN-LUDAS. If necessary, it is not so complicated to change, add and modify the indicators.

Digital image and graphs

A digital map interface was designed that enables users to navigate different landscape attributes at any point in time, through click buttons. This allows users to visually link changes in land use, forest yields and household mobility to important landscape attributes such as elevation, slope, distance to road/river, protected area, village territories, etc.

Graphs show "*real-time*" changes in predefined indicators as referred to above. Data underlying the graphs can be exported in worksheet format files at any point in time for further interpretation and documentation.

6.4 Results

6.4.1 VN-LUDAS as a tool for visualizing and testing the impacts of land-use policy interventions

The graphic user interface (GUI) of the VN-LUDAS for the Hong Ha watershed is shown in Figure 6.3. The GUI includes the following graphic components:

- User's input parameters and landscape navigation (parts (1) and (2) in Figure 6.3)
- Canvas of "*real-time*" maps of land-use/cover and forest stand basal area (part (3) in Figure 6.3, and Figure 6.4).
- Time-series graphs of the predefined indicators of the development path of the coupled human-landscape system (parts (4), (5), (6), (7), (8), (9), (10), (11), (12), (13), and (14) in Figure 6.3)

The user's input parameters are designed as numeric sliders, thus enabling users to continuously adjust the parameter values as needed (part (1) in Figure 6.3). They include global parameters and policy parameters. The global parameters include the size of the initial population that users want to generate, the spatial vision of households as a "*sphere of influence*" for their land-use choice, and the annual rainfall that will determine the calculation of the zoning protection score (see Table 6.1). The policy parameters are the five parameters that characterize the three policy factors as described above. The number of user's input parameters is kept to a minimum to reduce the cognitive overhead for end-users.

The digital landscape (part (3) in Figure 6.3) is associated with the system of graphic control buttons (part (2) in Figure 6.3), which enables users to highlight different attributes of the landscape with effective color schemes (see Figure 6.4). It also encourages users to visually correlate the landscape change, if any, with others landscape attributes.

The time-series graphs include two blocks. The first block includes graphs of indicators for the performance of the biophysical landscape system (see parts (4), (5), (6), (7), (8) and (9) in Figure 6.3). Graph (4) monitors changes in coverage of the 5 main land-use/cover types. Graph (5) monitors changes in the coverage of the dense/rich forest only, but calculates this for different land classes defined by from the distance from the road system. Proximity of roads is often associated with *hot spots* in land-cover changes. Graph (5) in Figure 6.3 can be linked to the digital map (f) in Figure 6.4 to obtain a visual image of the spatial extent of such hotspots. The same analogous approach can be applied to monitor the changes in the coverage of the other land-cover types.

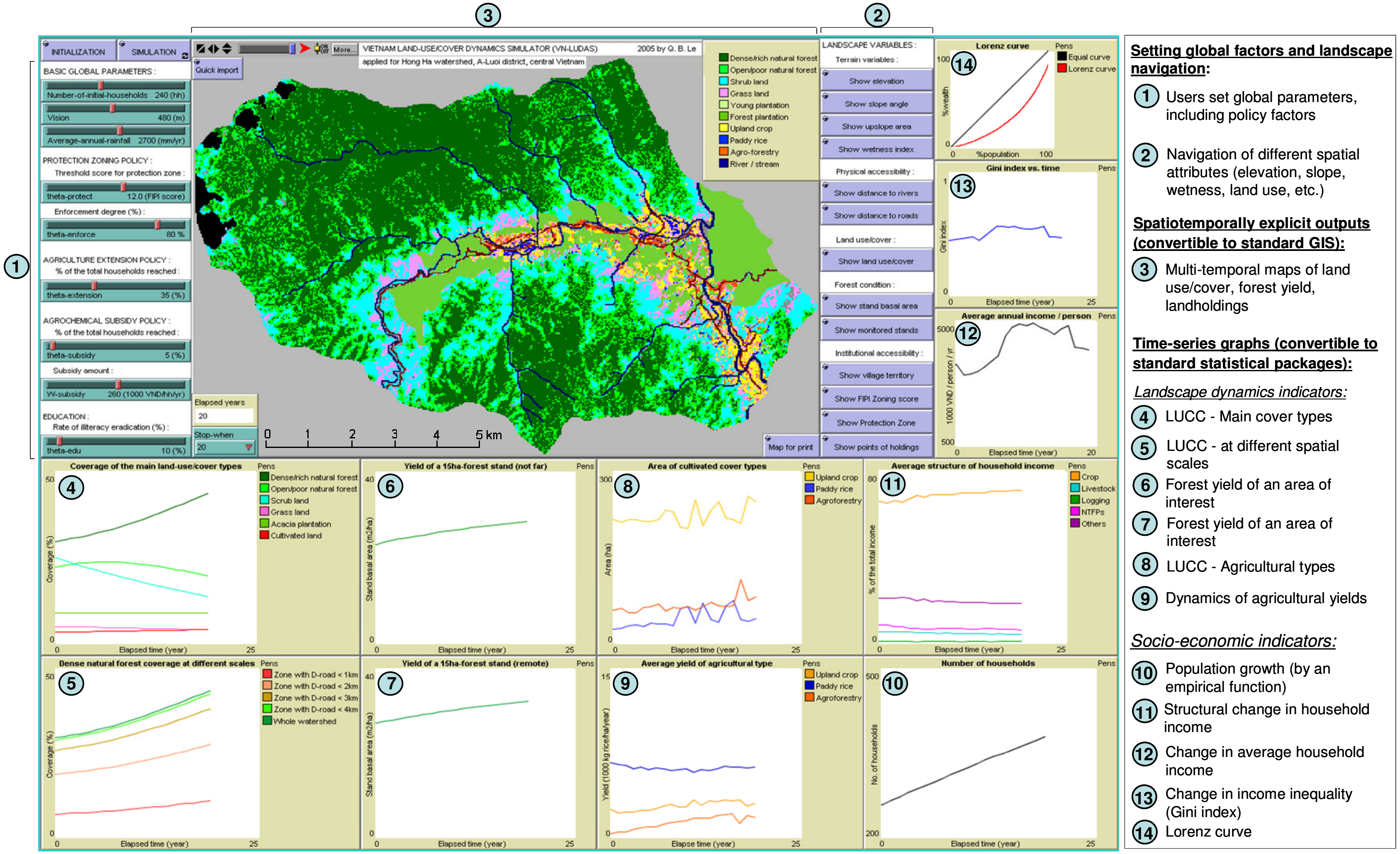


Figure 6.3 The VN-LUDAS's graphic-user interface enables users to visualize and test impacts of land-use policy choices

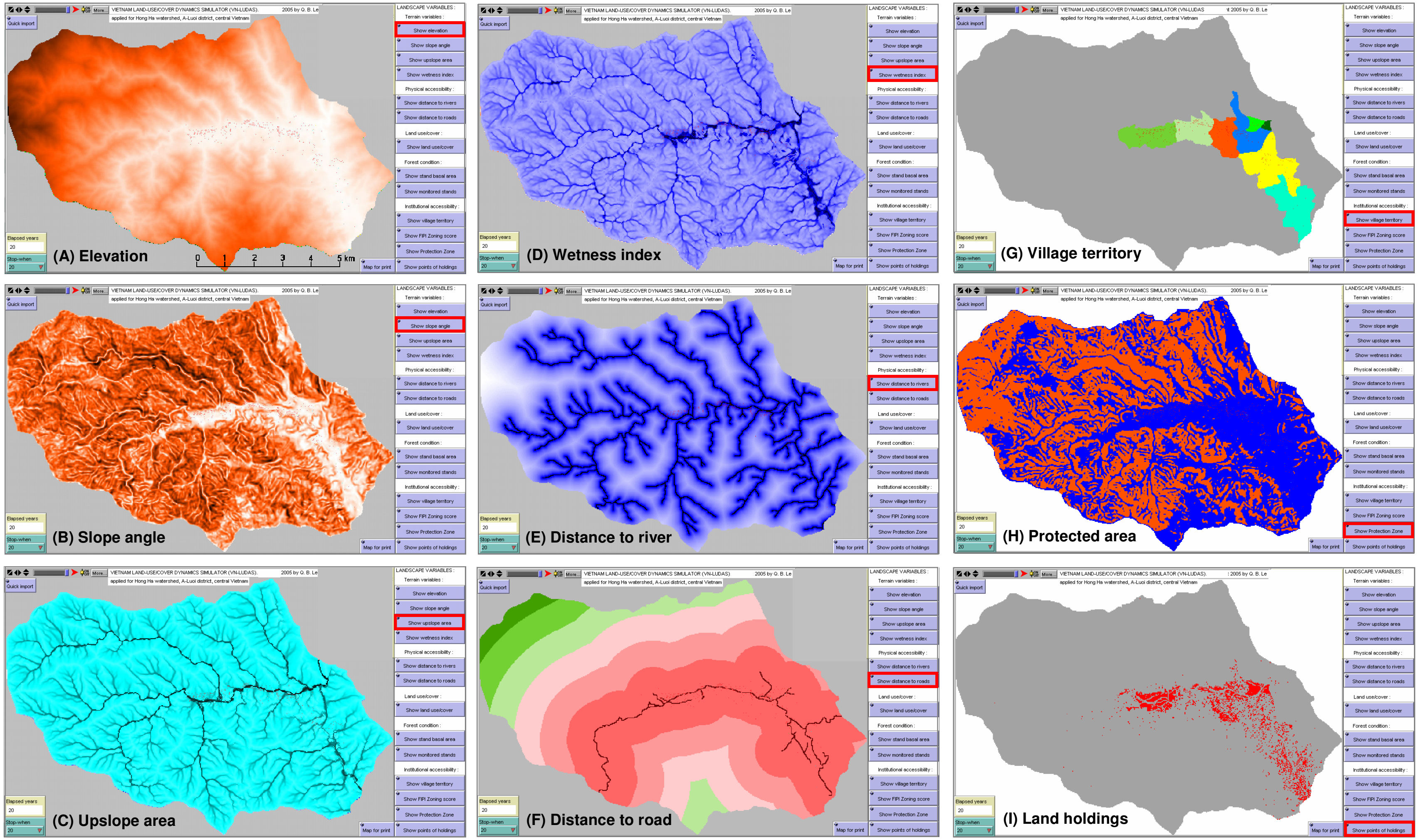


Figure 6.4 The VN-LUDAS's graphic-user interface enables users to create realistic landscapes with relevant spatial attributes. Note: button with the red border was activated to show the corresponding landscape attribute

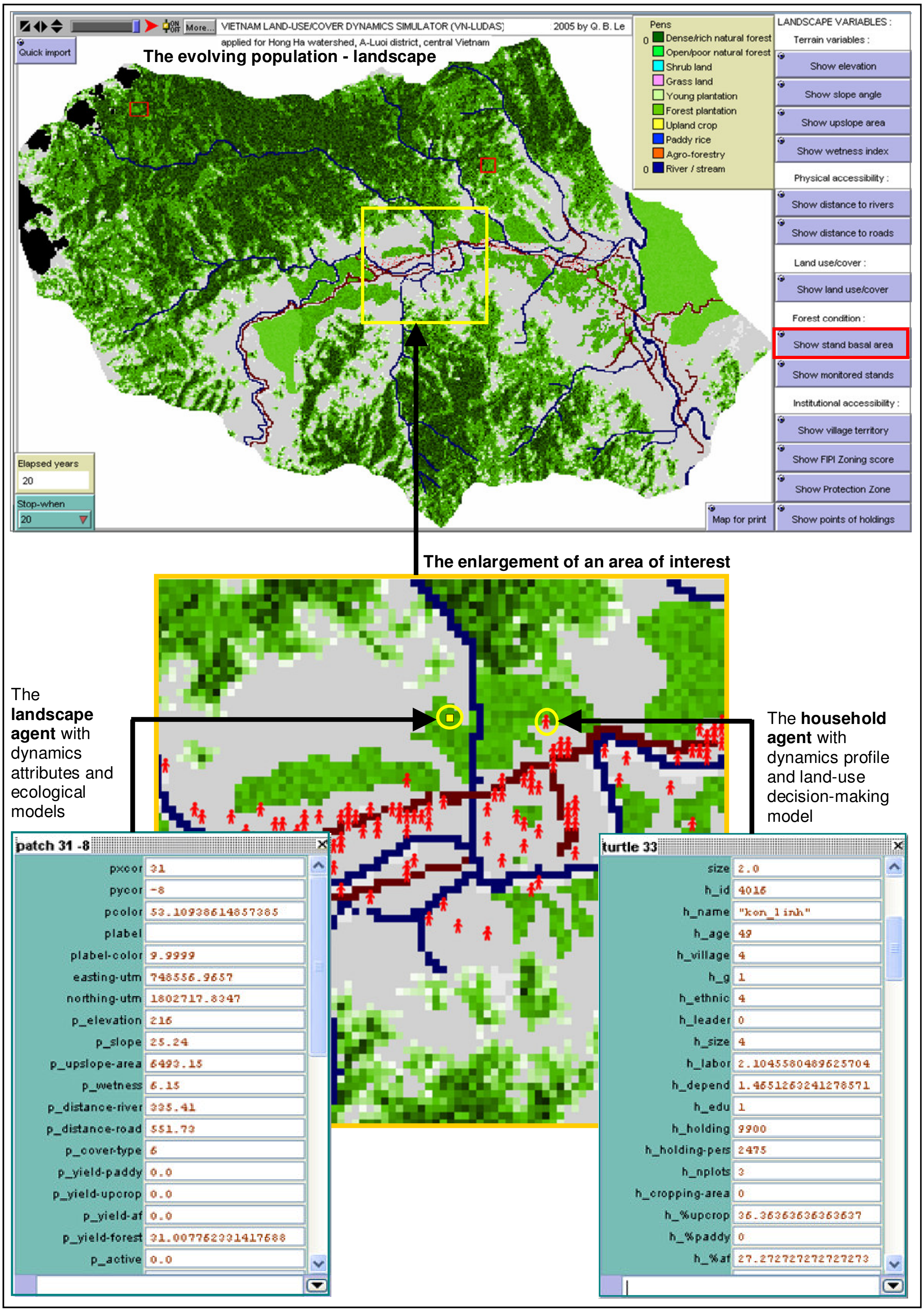


Figure 6.5 Real-time and spatially explicit monitoring of the evolution of the coupled human-landscape system at different aggregate levels in VN-LUDAS

Graphs (6) and (7) in Figure 6.3 illustrate that with the VN-LUDAS the spatial changes of any attribute of interest of the Hong Ha landscape can be monitored. These two graphs monitor the changes in forest stand basal area of the two rectangular plots of 15 ha in different positions (see the two red rectangular plots in the forest map of Figure 6.5). Similarly, it is possible to create a graph to monitor land-use change within a village territory that may be of interest to the head of such a village.

Graphs (8) and (9) in Figure 6.3 monitor changes in the 3 main agricultural land-use types, both area/extent (graph (8)) and productivity terms (graph (9)). As such, graph (8) monitors the area of each agricultural land-use type and graph (9) measures the yield of each land-use type. Based on that one can easily calculate the total production for each land-use type over time.

The second time-series block of graphs includes the graphs for monitoring changes in the human system, in parallel to the changes in the natural landscape system (see graphs (10), (11), (12), (13), and (14) in Figure 6.3). Graph (10) shows population trends for which the growth rates are calculated from a historical dataset of the Hong Ha population. Graph (11) shows changes, if any, in average income structure of households. Graph (12) monitors changes in the average income of a household, and graph (13) shows the inequality of income distribution over the community in term of the Gini index. The Gini index is calculated based on the Lorenz curve in graph (14). Using the same principles for monitoring changes in the landscape system, it is thus possible to create graphs to monitor socio-economic changes of any defined subset of the community.

Figure 6.5 illustrates that with the VN-LUDAS, co-evolution paths of the natural landscape system and human system can be monitored in a spatially explicit and "*real-time*" manner, at all aggregated levels of household/landscape agents.

By setting the policy parameters as needed and observing the scenario development on the GUI, an *interaction loop of input-indicator* develops between the model and its users. Through this interaction loop, learning of users about environmental consequences of human actions will be improved. The calibrated GUI of the VN-LUDAS also enables in-depth scenario studies through analysis and comparison of the simulated data.

6.4.2 Impacts of protection zoning policy on land use/cover and socio-economic status

Three scenarios S-Pro0 (*no protection*), S0 (*baseline*), and S-Pro3 (*strict protection*) were run to assess the impacts of change in forest protection zoning policy. The simulated spatiotemporal land-use/cover maps for the three tested scenarios S-Pro0, S0, S-Pro3 are shown in Figure 6.6 (S-Pro0 *vs.* S0) and Figure 6.7 (S-Pro3 *vs.* S0). Two main characteristics of the change in natural forest cover types are apparent with the increase of protection level. First, the decrease of protection level likely leads to a significant conversion of dense natural forest to open natural forest in the areas near the road system or the settlement areas, thus suggesting a scale-dependent forest degradation[18] trend. Second, overall natural forest coverage (including dense and open natural forest) is not likely to change much due to the change in protection policy, thus suggesting no significant deforestation[19].

Figures 6.7a and 6.9a provide a more accurate picture of the overall change in the natural forest cover types. The two graphs clearly show that if the protection policy is dismissed (Figure 6.7a), or the enforcement of the protection rules is poor, more dense/rich natural forest is converted to open forest area, indicating a significant increase in forest degradation. Both graphs also show that the non-forest-cover types (i.e., shrub land, grass land, forest plantation, and agricultural land) do not increase in their coverage.

Figures 6.7b and 6.9b reflect a significant effect of infrastructure on the forest degradation caused by halting forest protection. The graphs show a significant decrease of dense forest coverage within zones closer to road/settlement areas (especially within the zone with distance to road < 2 km), whereas the degradation becomes less if the spatial scale of the coverage calculation increases. Apparently, the labor constraints and physical inaccessibility to the remote, high and steep forested mountain areas are protection in themselves.

[18] Forest degradation refers to changes in forest quality, while the forest canopy still maintain.
[19] Deforestation refers to conversion of forest cover types to non-forest cover types.

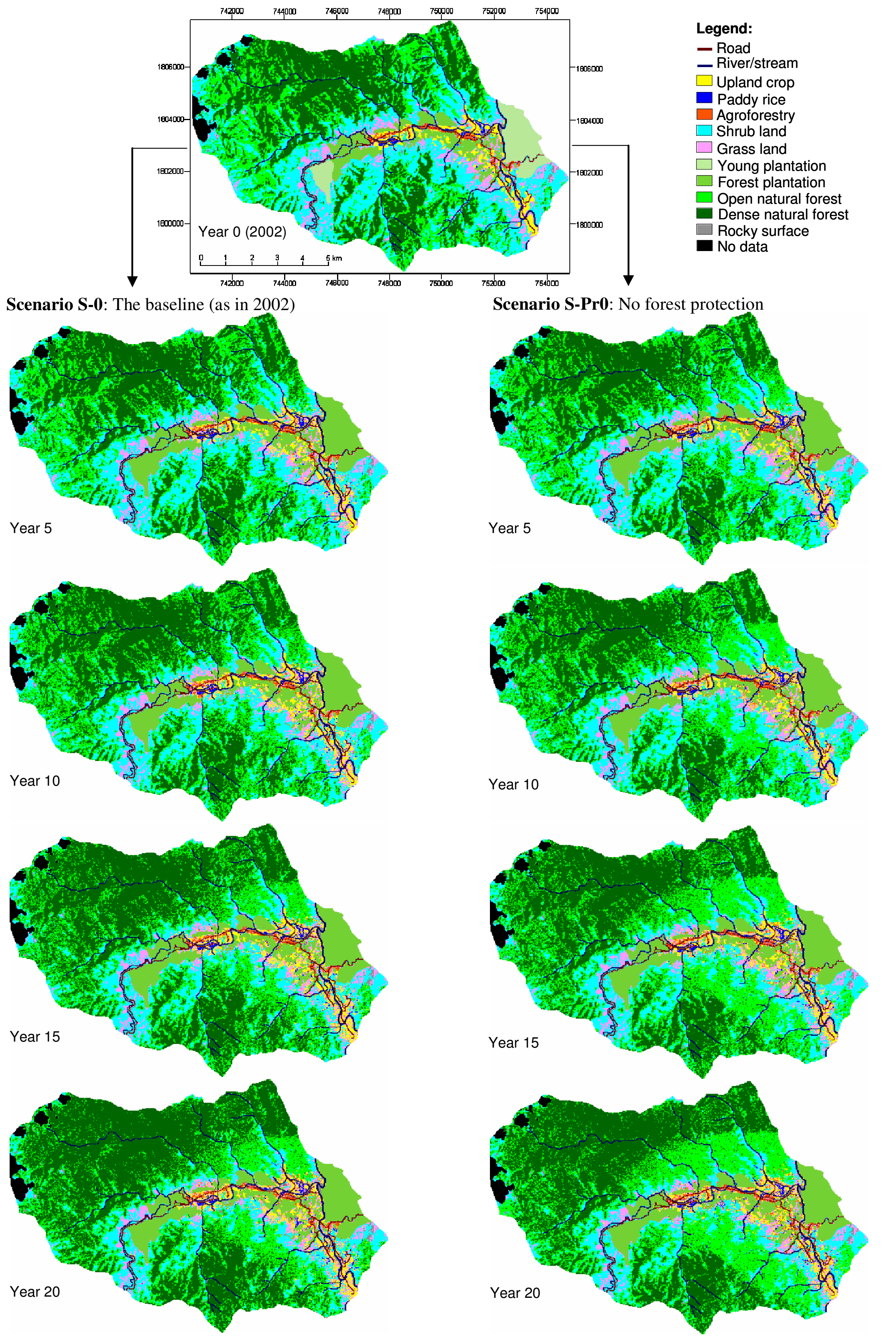
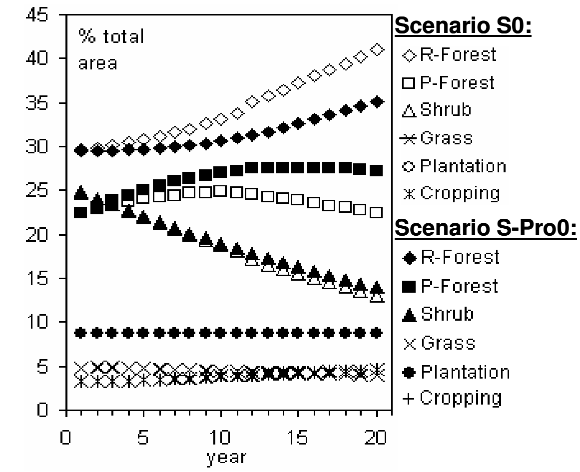
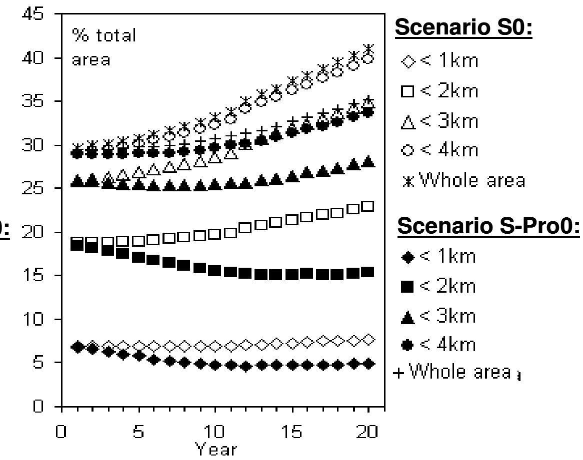


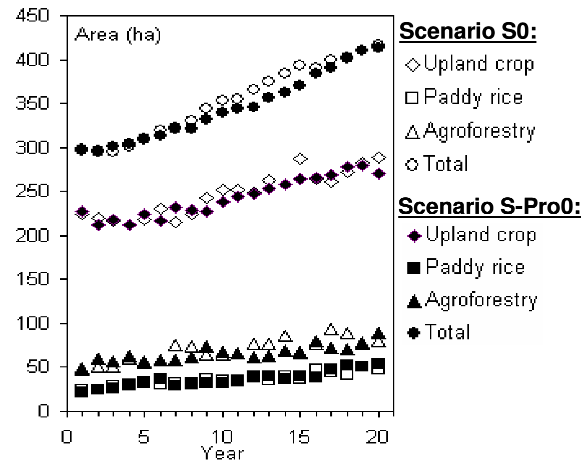
Figure 6.6 Simulated spatiotemporal land use/cover for policy scenario S-Pro0 (no forest protection) in comparison to scenario S0 (current trend). Source: data exported from VN-LUDAS's spatial outputs



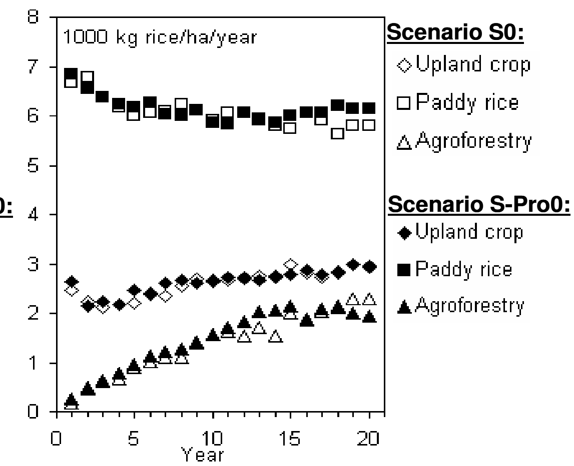
a) Coverages of 5 main land-use/cover types



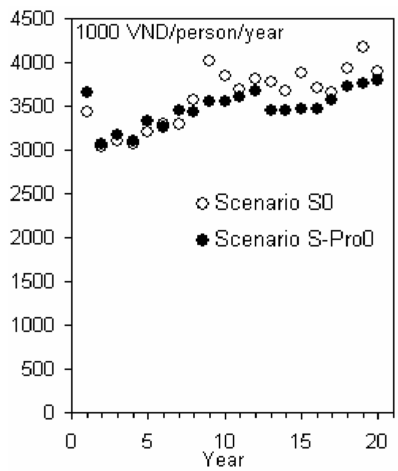
b) Coverages of dense forest *vs.* road proximity



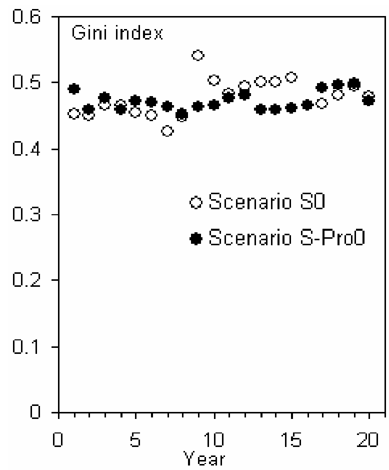
c) Areas of 3 main agricultural land-use types



d) Yields of 3 main agricultural land-use types



e) Per capita annual gross income



f) Gini indexes

Figure 6.7 Simulated land-use/cover and socio-economic changes for policy scenario S-Pro0 (no forest protection) in comparison to scenario S0 (the current trend). Note: R-forest: rich/dense natural forest, P-Forest: poor/open natural forest. Source: data exported from VN-LUDAS's temporal outputs

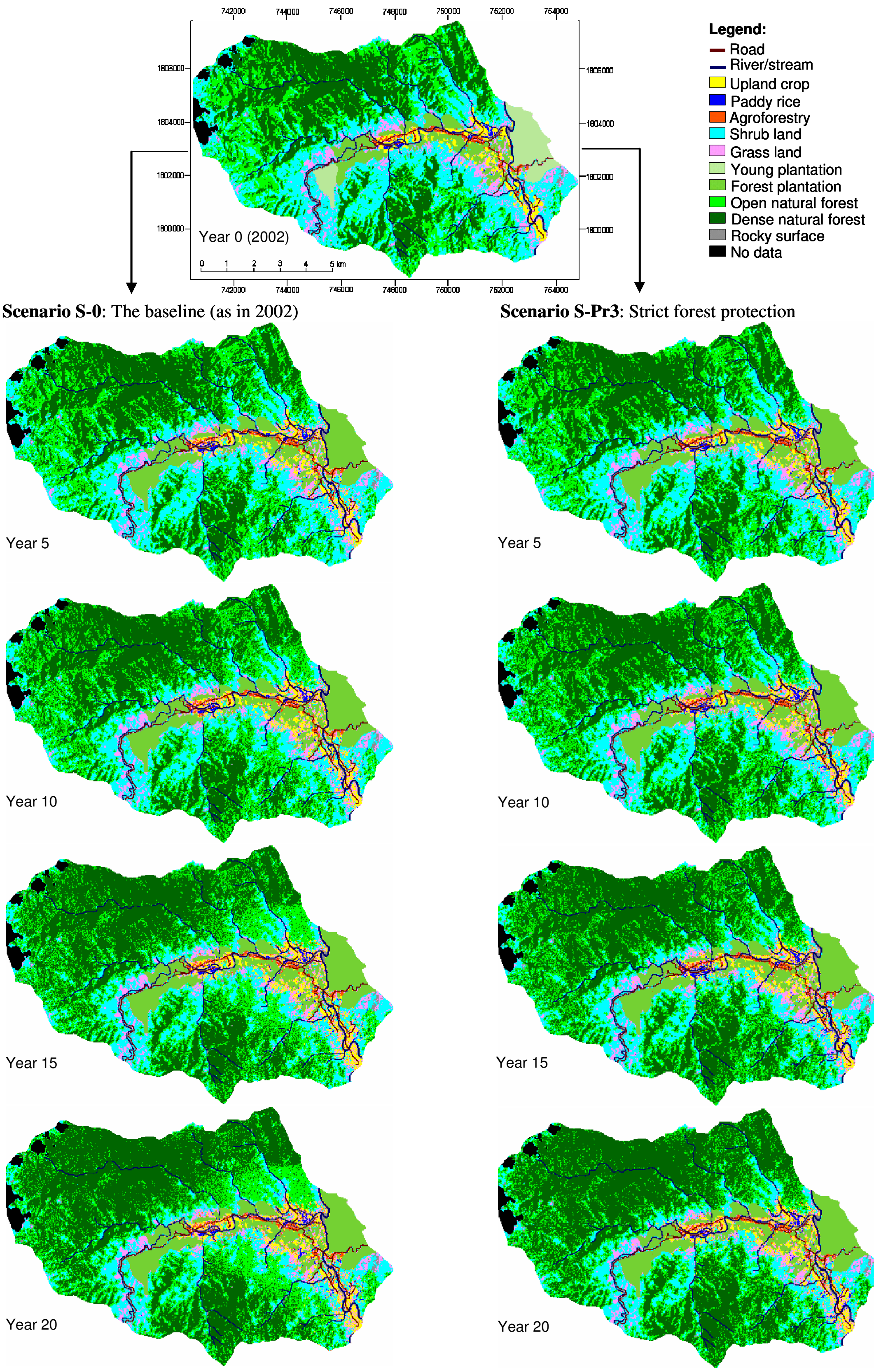
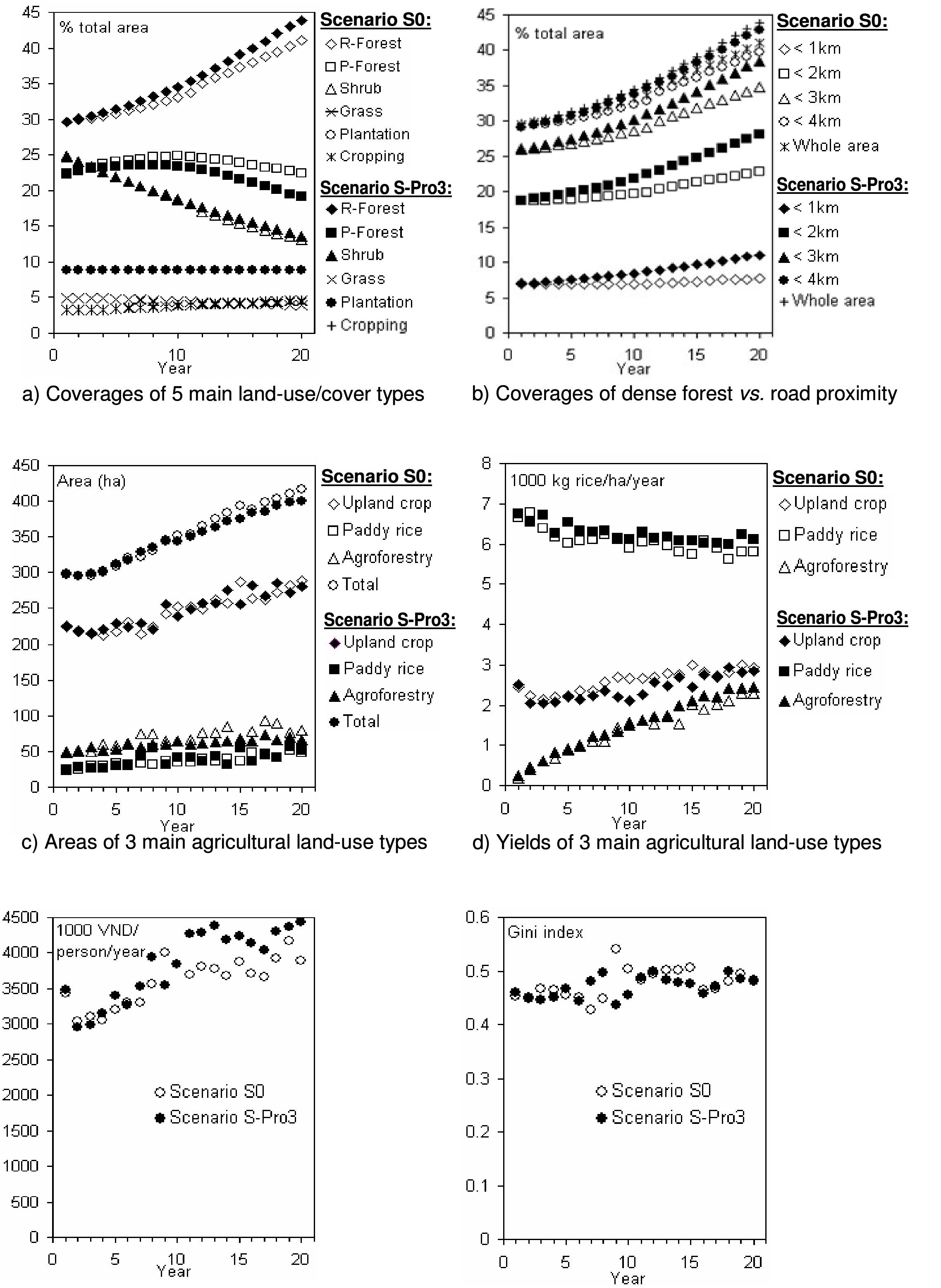


Figure 6.8 Simulated spatiotemporal land use/cover for policy scenario S-Pro3 (strict forest protection) in comparison to scenario S0 (the current trend). Source: data exported from VN-LUDAS's spatial outputs



a) Coverages of 5 main land-use/cover types

b) Coverages of dense forest *vs.* road proximity

c) Areas of 3 main agricultural land-use types

d) Yields of 3 main agricultural land-use types

e) Per capita annual gross income

f) Gini indexes

Figure 6.9 Simulated land-use/cover and socio-economic changes for policy scenario S-Pro3 (strict forest protection) in comparison to scenario S0 (current trend). Note: R-forest: rich/dense natural forest, P-Forest: poor/open natural forest. Source: data exported from VN-LUDAS's temporal outputs

The simulated changes in agricultural land-use type in Figures 6.7c and 6.9c indicate that there is likely no significant change in agricultural production caused by the change of protection policy. However, Figures 6.7e and 6.9e show that household income likely increases with the increase in forest protection enforcement. This is surprising, as many rural developers in Hong Ha often think that blocking farmer's access to forest tree resources will lead to a decrease in household income. Further interpretation would require conducting an analysis of the structural changes in household economies and transitions in household typology. With the simulated world data produced by the VN-LUDAS, such an analysis is principally feasible but certainly requires a high volume of statistical analyses as well as a fine micro-economic knowledge, and thus is beyond the scope of this chapter.

6.4.3 Impacts of agricultural extension on LUCC and community dynamics

The scenarios S-Ext0 (*no/minor extension*), S-Ext1 (*medium extension*) and S0 (*baseline – high extension*) were generated to explore the impacts of the change in the agricultural extension services. The simulated spatiotemporal land-use/cover maps in Figures 6.10 and 6.12, as well as the time-series graphs in Figures 6.11a, 6.11b, 6.13a, and 6.13b show that the change in the extension practice likely has no effect on land use/cover for any observed spatial scale.

Simulated extent of agricultural types (Figures 6.11c and 6.13c) and agricultural productivity (Figures 6.11d and 6.13d) indicate that reduced agricultural extension leads to a slower increase of agricultural land, especially the area of upland crops in one hand, but a higher agricultural productivity on the other hand. This suggests that current extension services in Hong Ha do not encourage intensification of crop production. Based on narrative information gathered during the field survey, there may be that farmers having more access to extension services often give more attention to the use of hillsides for crop production. However, due to considerable natural constraints, the hillside areas are generally marginal for crop production (see Chapter 5). The low crop yields of new upland fields on hillsides may reduce the overall crop yield.

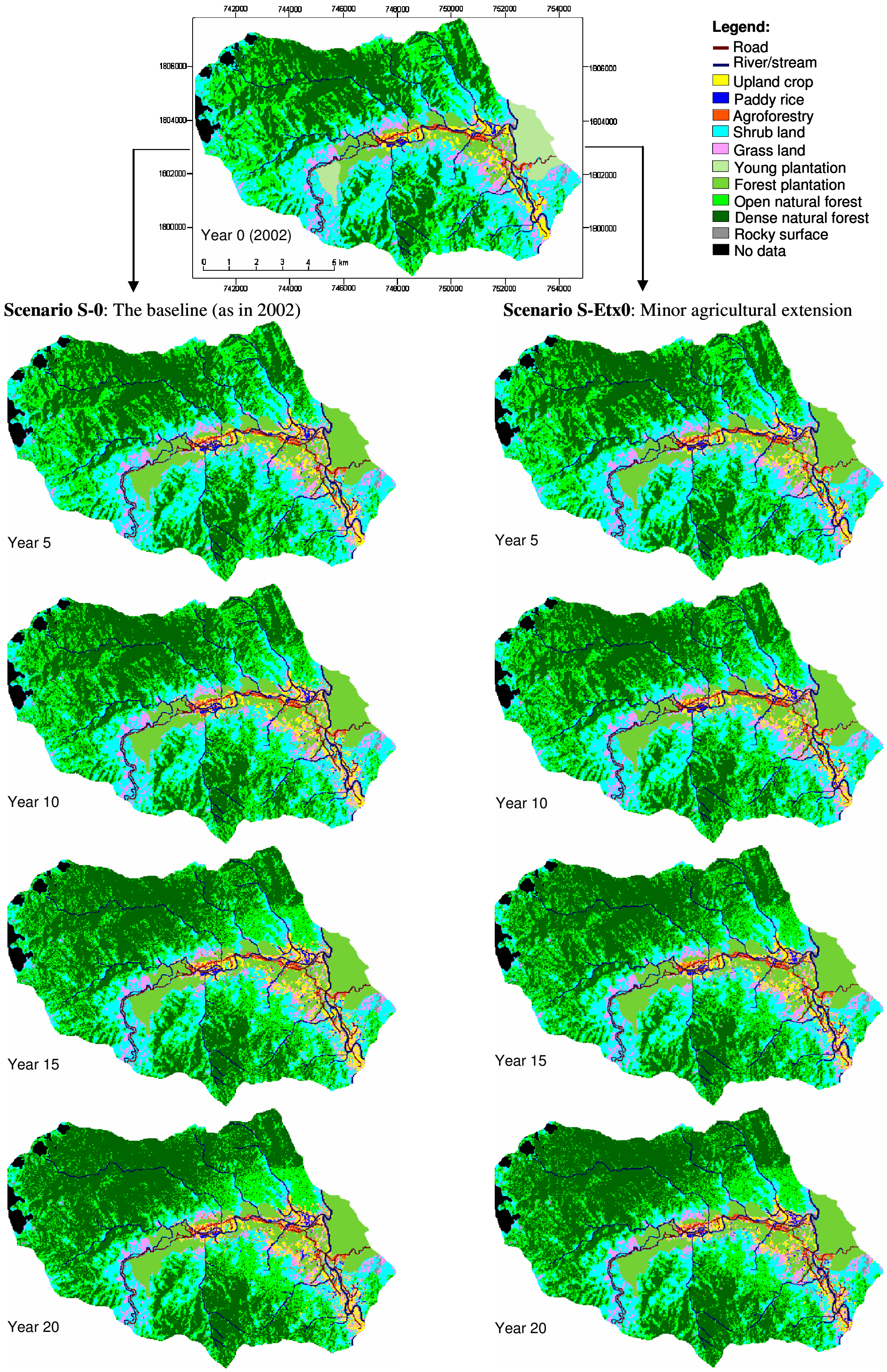


Figure 6.10 Simulated spatiotemporal land use/cover for policy scenario S-Ext0 (no/minor extension) in comparison to scenario S0 (the current trend). Source: data exported from VN-LUDAS's spatial outputs

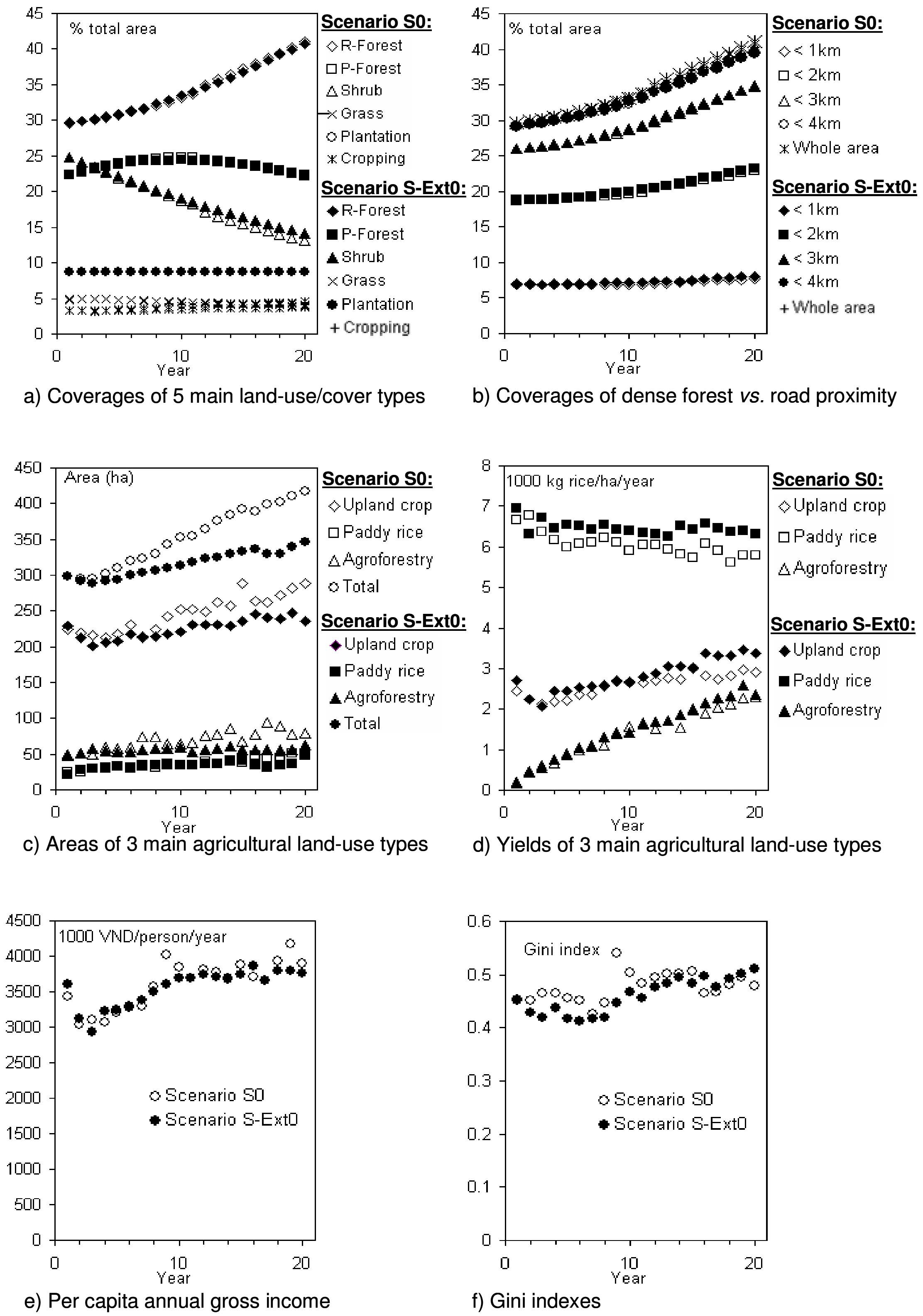


Figure 6.11 Simulated land-use/cover and socio-economic changes for policy scenario S-Ext0 (minor/no agricultural extension) in comparison to scenario S0 (current trend – high agricultural extension). Note: R-forest: rich/dense natural forest, P-Forest: poor/open natural forest. Source: data exported from VN-LUDAS's temporal outputs

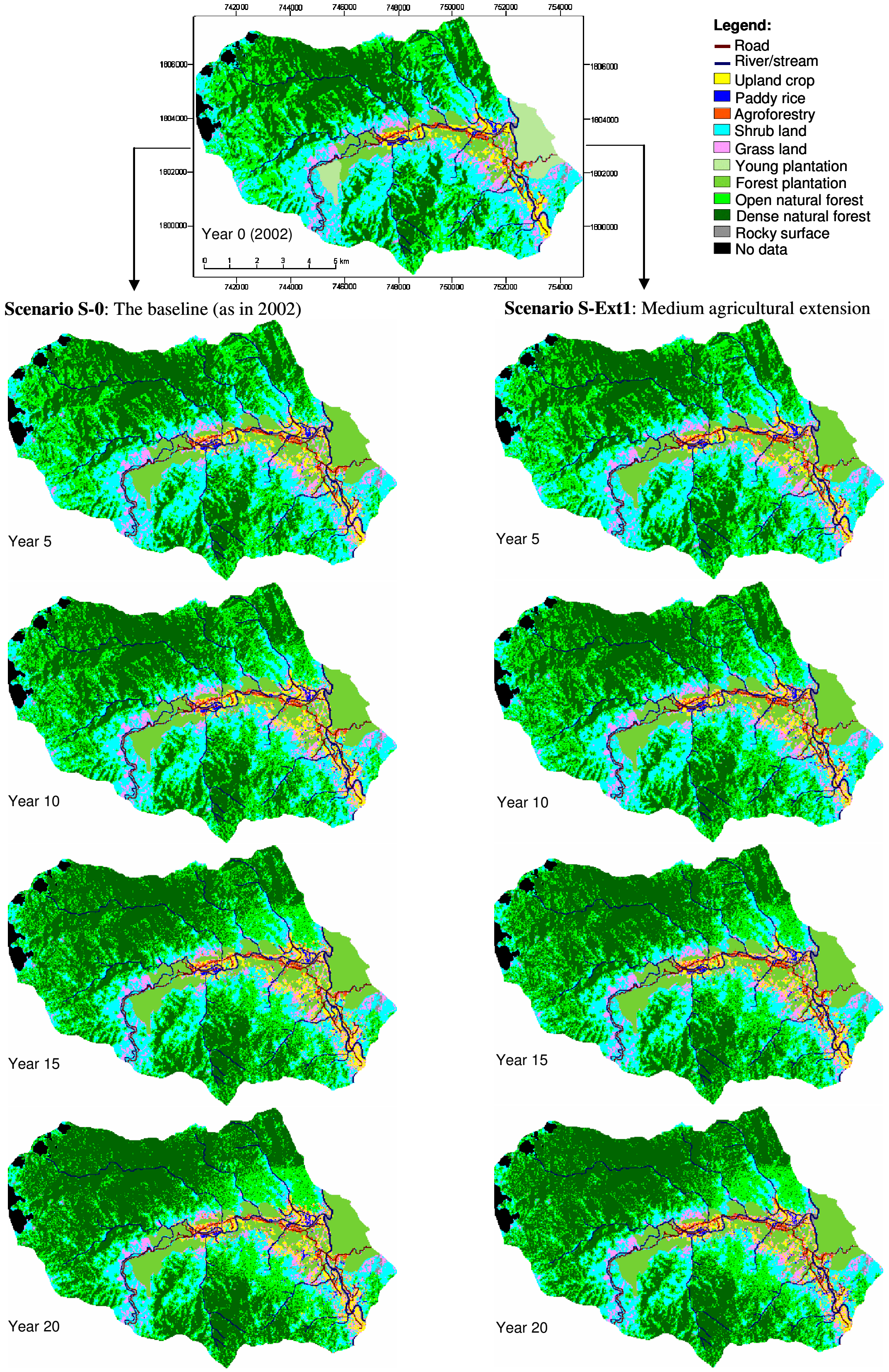


Figure 6.12 Simulated spatiotemporal land use/cover for policy scenario S-Ext1 (medium agricultural extension) in comparison to scenario S0 (the current trend). Source: data exported from VN-LUDAS's spatial outputs

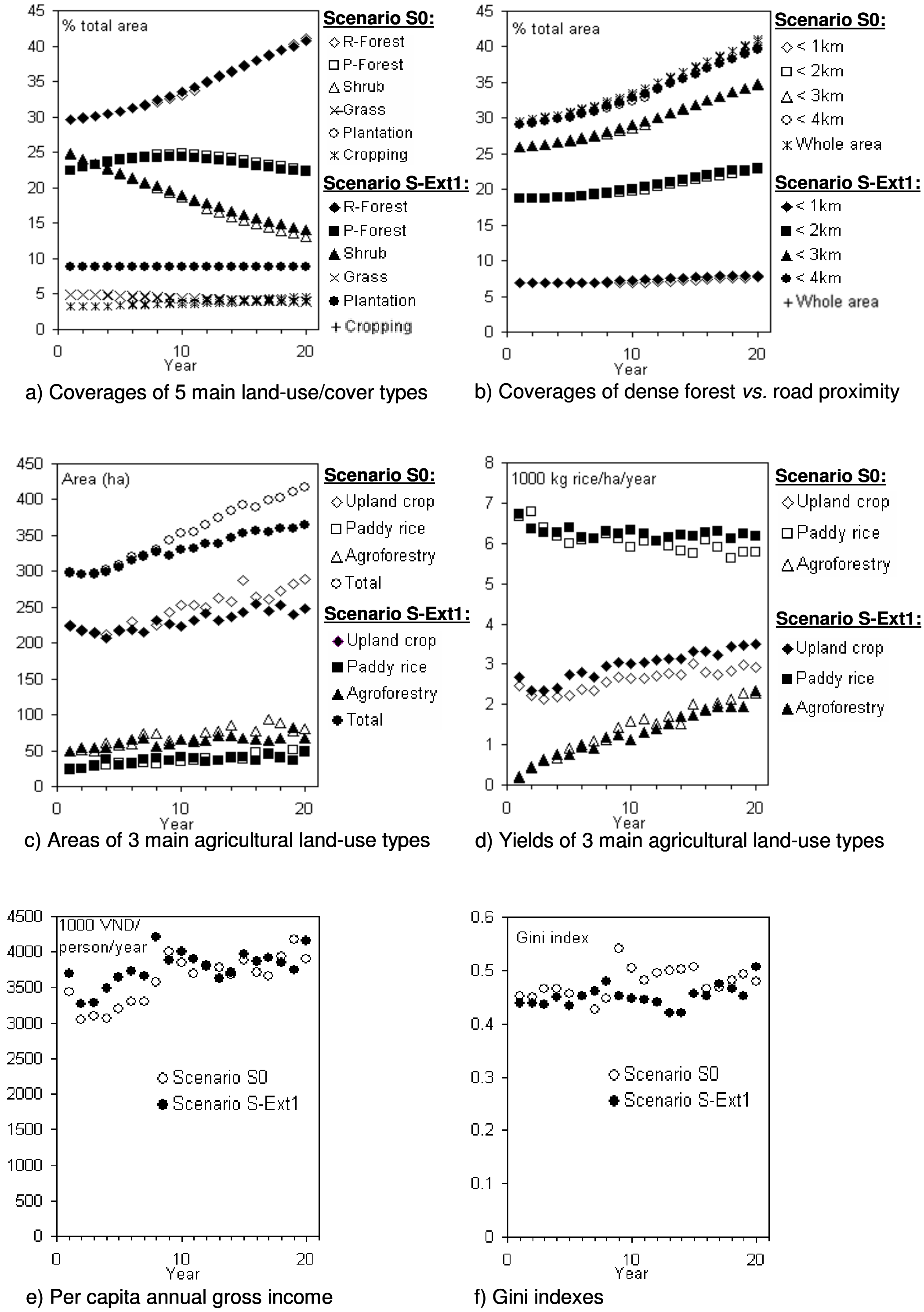


Figure 6.13 Simulated land-use/cover and socio-economic changes for policy scenario S-Ext1 (medium agricultural extension) in comparison to scenario S0 (current trend – high agricultural extension). Note: R-forest: rich/dense natural forest, P-Forest: poor/open natural forest. Source: data exported from VN-LUDAS's temporal outputs

The simulated average income (Figures 6.11e and 6.13e) and Gini index (Figure 6.11f and 6.13f) show that the scenario S-Ext1 (*medium extension*) may offer a slight increase in average income in the first 10 years and a more equal distribution of income (i.e., lower *Gini* index) in the later period.

6.4.4 Impacts of agrochemical subsidies on LUCC and community dynamics

The scenarios S-Sub0 (*no subsidy*), S0 (*baseline – low subsidy*), S-Sub1 (*medium subsidy*), and S-Sub2 (*high subsidy*) were run to explore the impacts of the change in the extent of agrochemical subsidies. The simulated spatiotemporal land-use/cover maps in Figures 6.14, 6.16, 6.18 as well as the time-series graphs in Figures 6.15a, 6.15b, 6.17a, 6.17b, 6.19a, and 6.19b show that the change in the agrochemical subsidies likely has no effect on land cover irrespective of distance to roads or villages.

Simulated results on the areas of agricultural types (Figures 6.15c, 6.17c, and 6.19c) and agricultural yields (Figures 6.15d, 6.17d, and 6.19d) show that the increase in the access to agrochemical subsidies, in general, leads to a decrease of agricultural land and an increase in agricultural productivity. However, the yield increment is too small in comparison to the subsidy increment. In the case of the scenarios S-Sub0, agrochemical subsidies still support higher yields in the case that 23% of the households were subsidized. The low fertilizer efficiency in crop production in Hong Ha (see Chapter 5) may be one of the reasons for this situation.

Simulated average household income (Figures 6.15e, 6.17e, and 6.19e) and Gini index (Figures 6.15f, 6.17f, and 6.19f) indicate that changes in the access to agrochemical subsidies would have no effects on household income and income equality. In this context, scenario S-Sub0 may be the most positive as there is no cost for subsidies in this case.

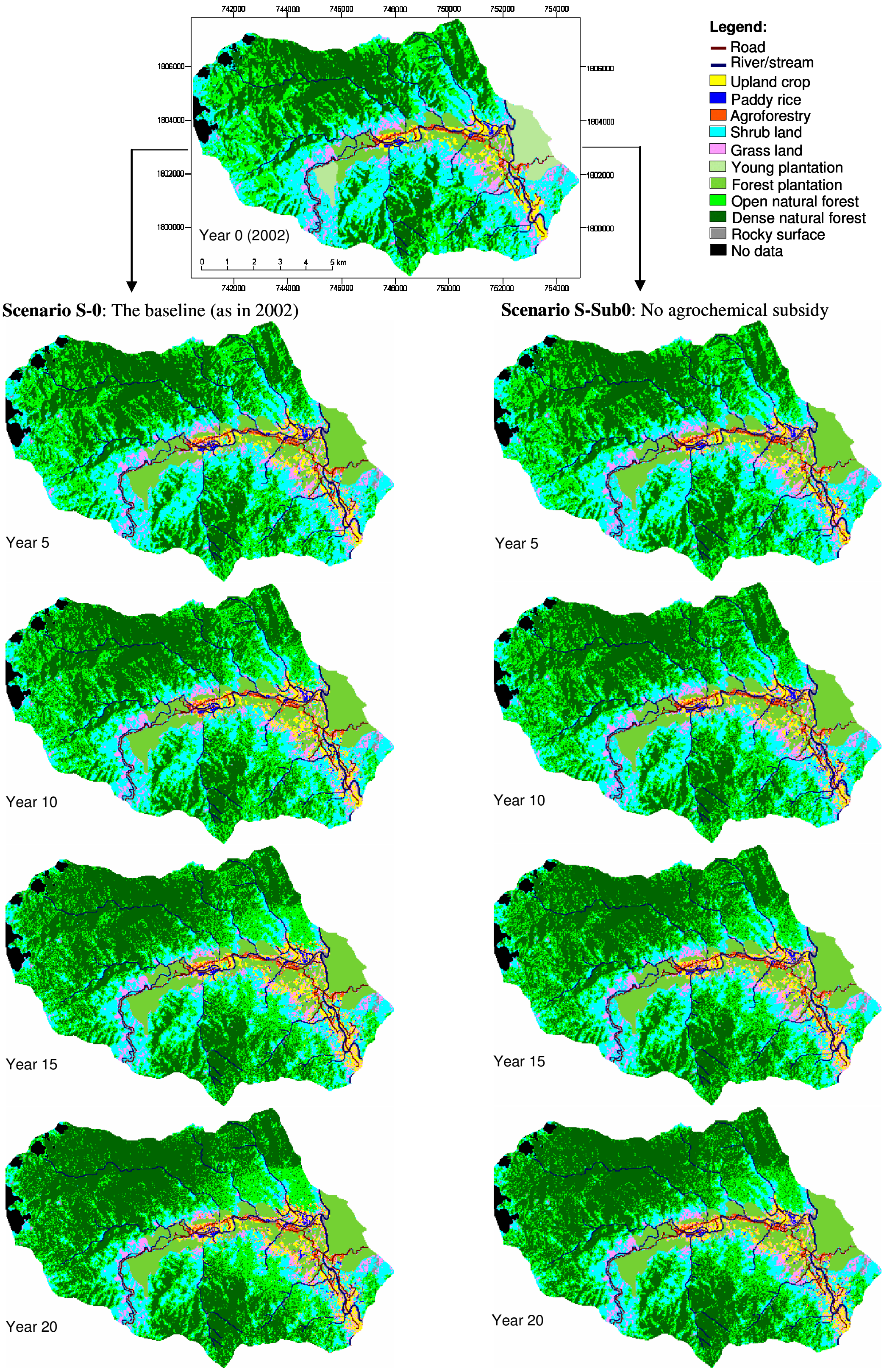
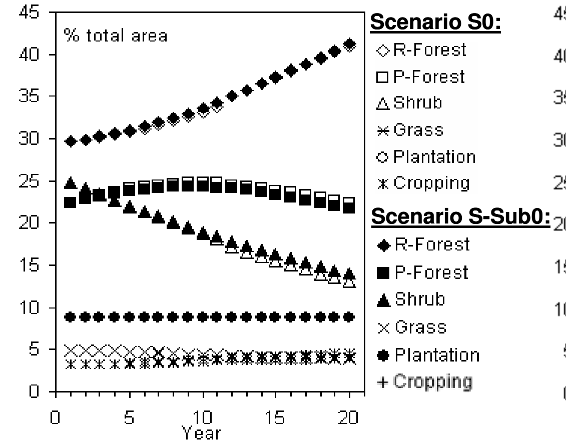
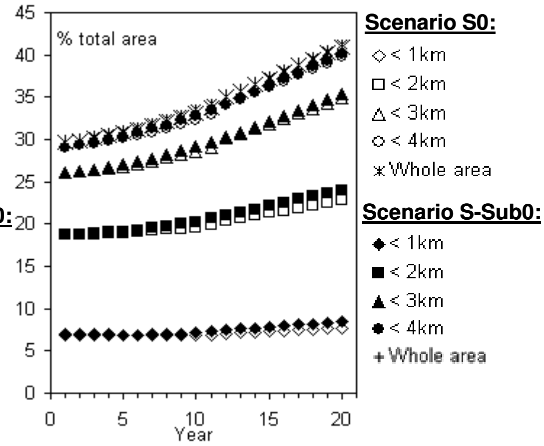


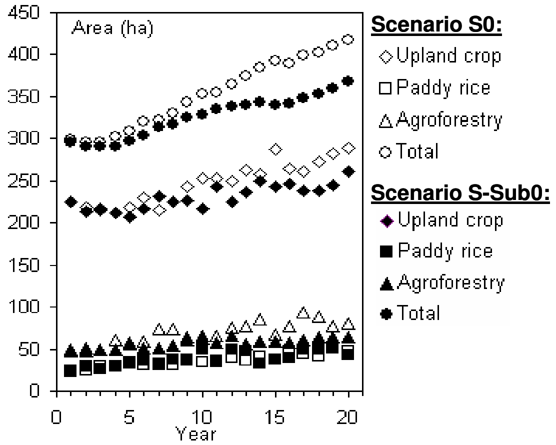
Figure 6.14 Simulated spatiotemporal land use/cover for policy scenario S-Sub0 (no agrochemical subsidy) in comparison to scenario S0 (the current trend). Source: data exported from VN-LUDAS's spatial outputs



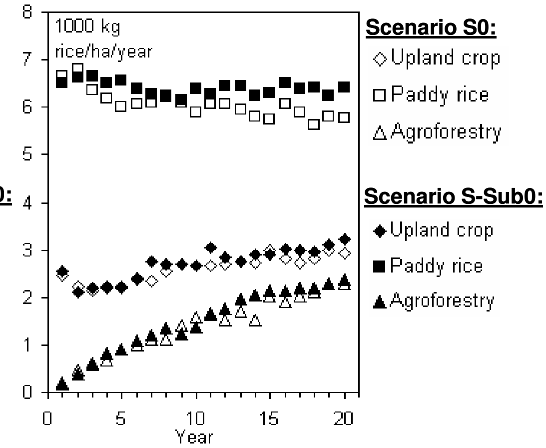
a) Coverages of 5 main land-use/cover types



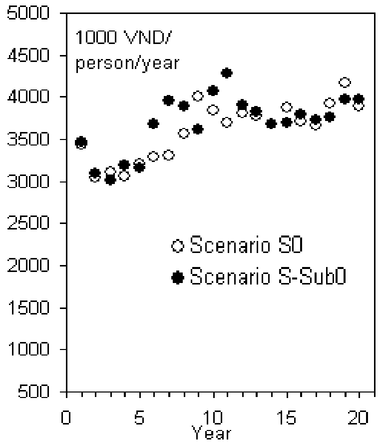
b) Coverages of dense forest *vs.* road proximity



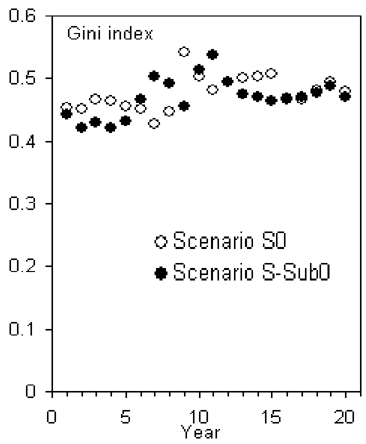
c) Areas of 3 main agricultural land-use types



d) Yields of 3 main agricultural land-use types



e) Per capita annual gross income



f) Gini indexes

Figure 6.15 Simulated land-use/cover and socio-economic changes for policy scenario S-Sub0 (no agrochemical subsidy) in comparison to scenario S0 (current trend – low agrochemical subsidy). Note: R-forest: rich/dense natural forest, P-Forest: poor/open natural forest. Source: data exported from VN-LUDAS's temporal outputs

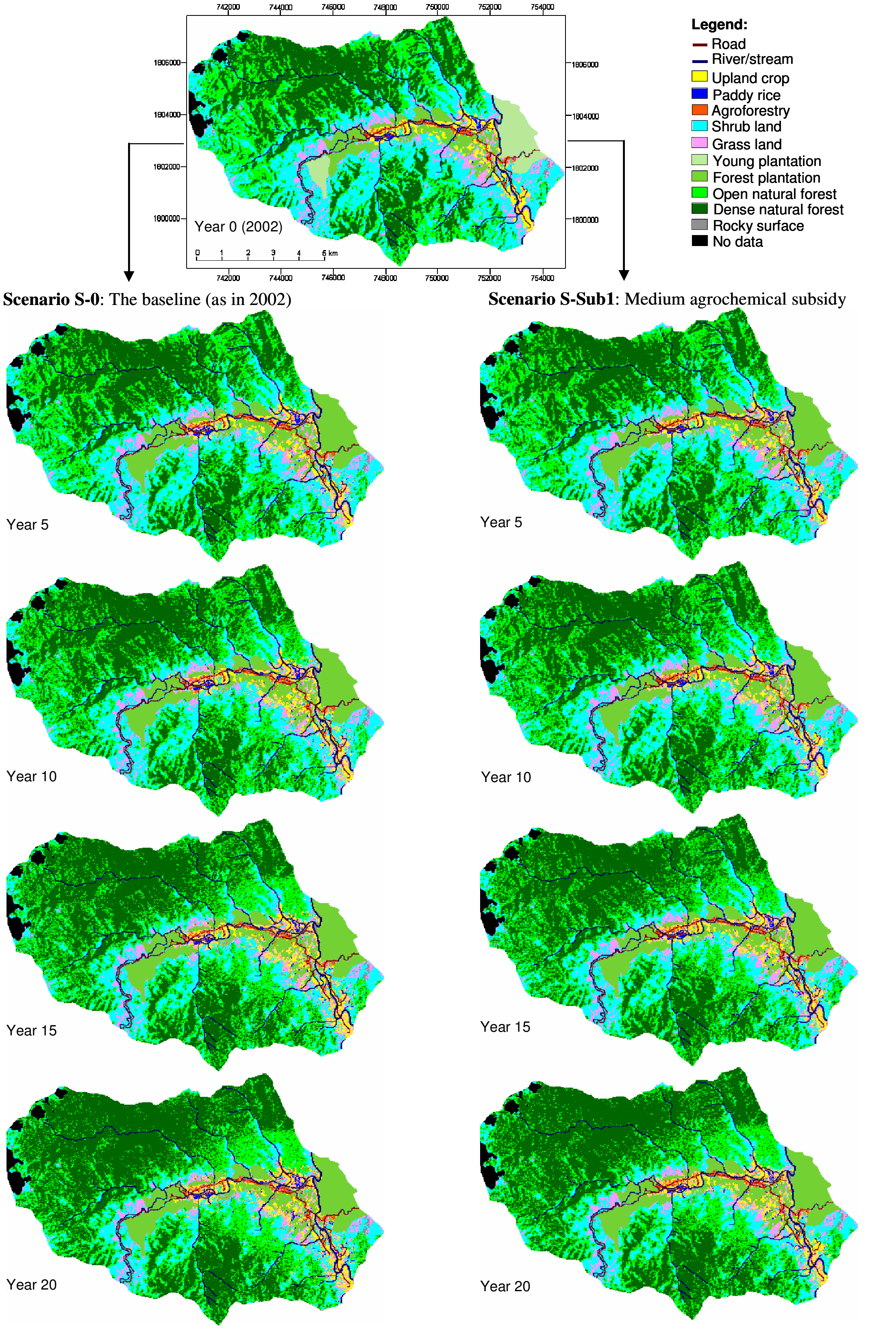


Figure 6.16 Simulated spatiotemporal land use/cover for policy scenario S-Sub1 (medium agrochemical subsidy) in comparison to scenario S0 (the current trend). Source: data exported from VN-LUDAS's spatial outputs

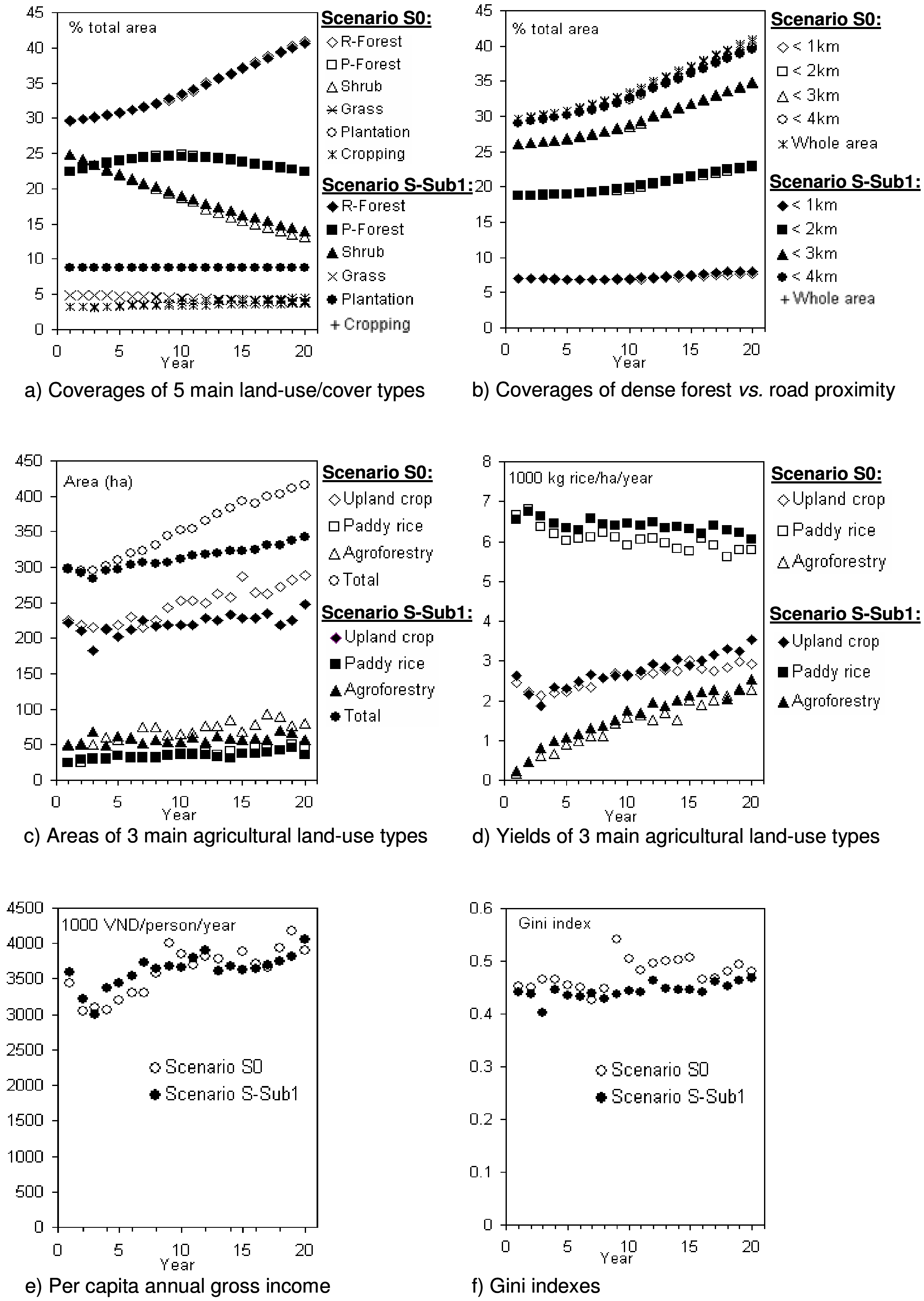


Figure 6.17 Simulated land-use/cover and socio-economic changes for policy scenario S-Sub1 (medium agrochemical subsidy) in comparison to scenario S0 (current trend – low agrochemical subsidy). Note: R-forest: rich/dense natural forest, P-Forest: poor/open natural forest. Source: data exported from VN-LUDAS's temporal outputs

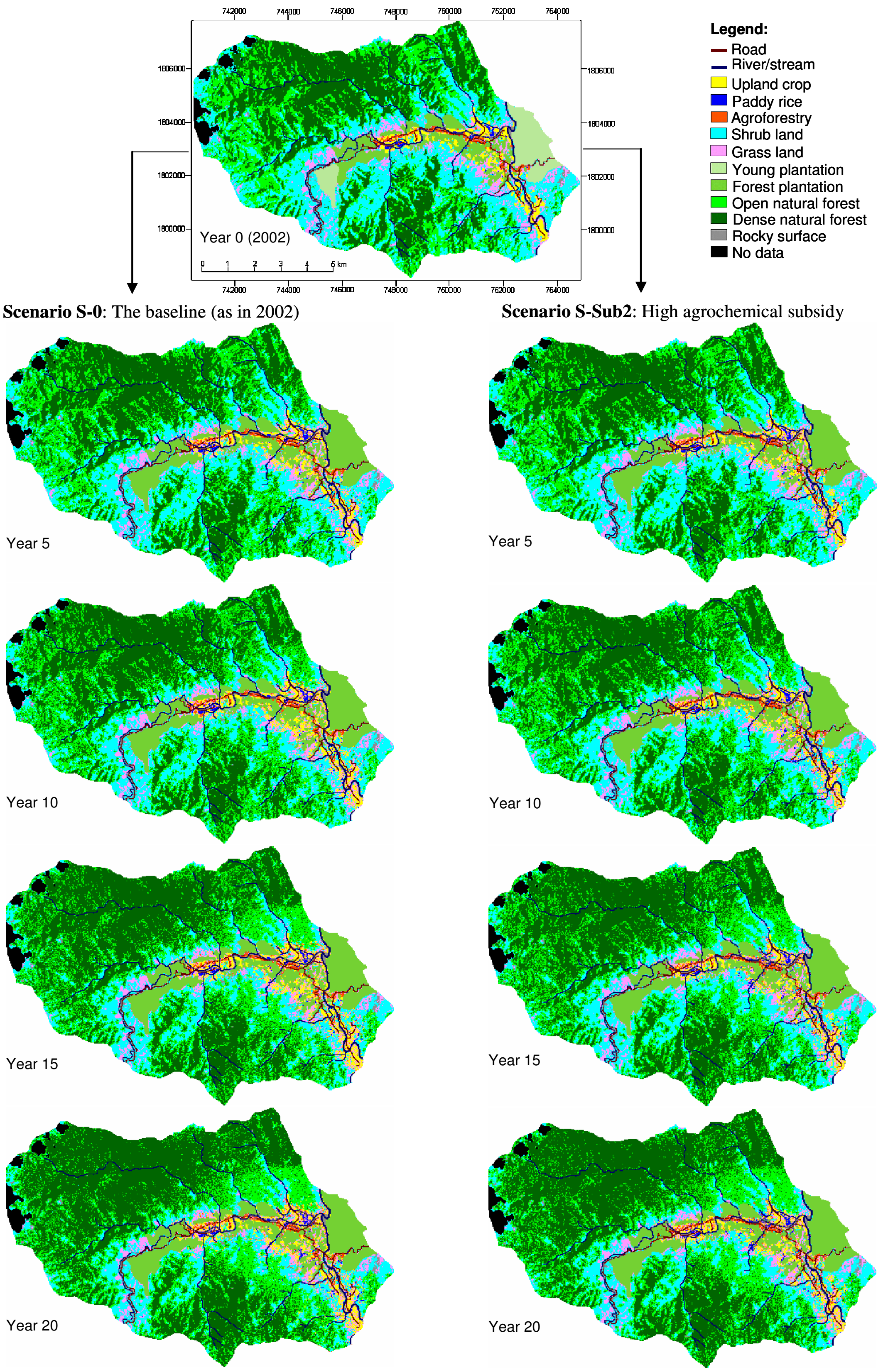
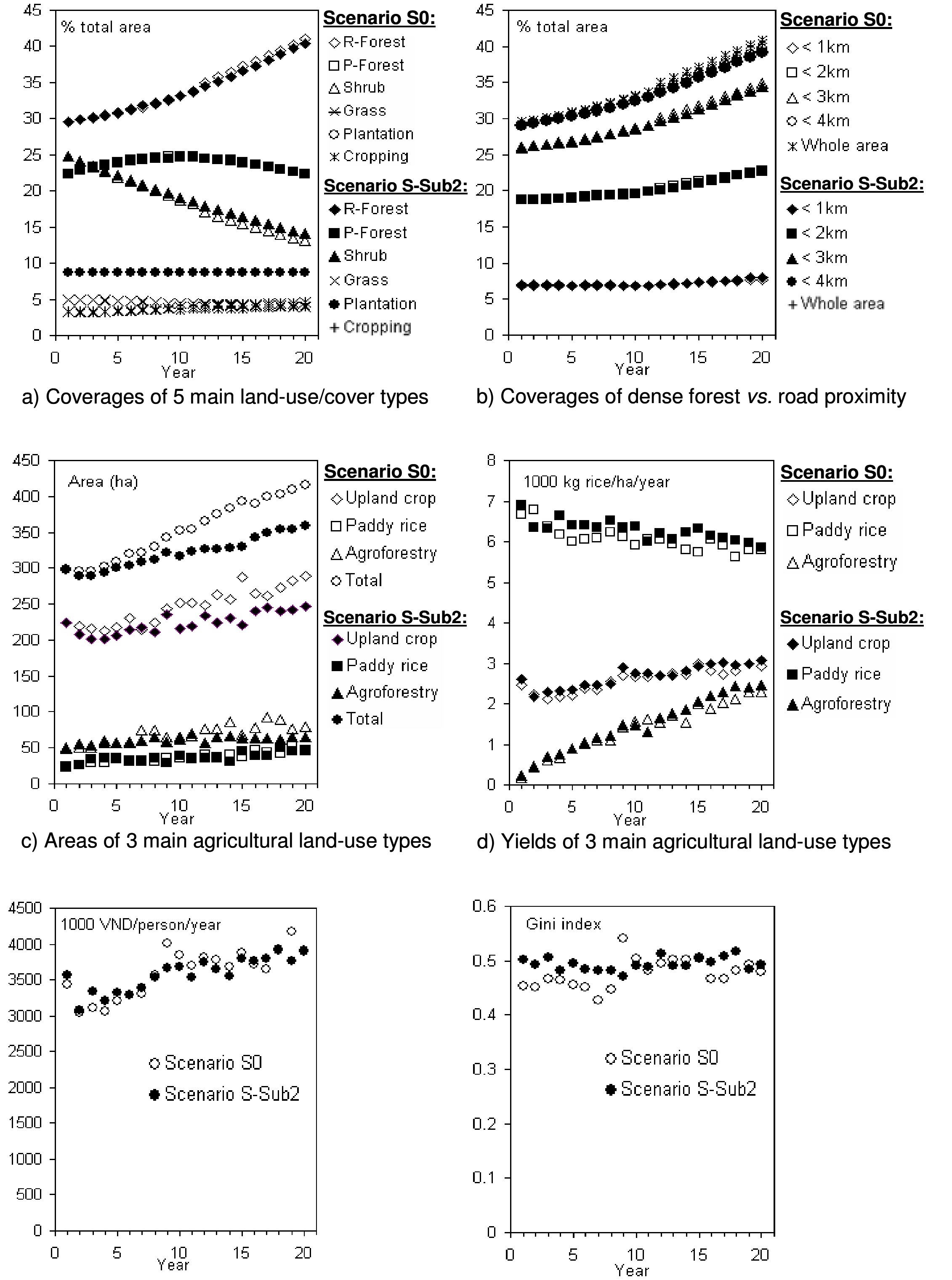


Figure 6.18 Simulated spatiotemporal land use/cover for policy scenario S-Sub2 (high agrochemical subsidy) in comparison to scenario S0 (the current trend). Source: data exported from VN-LUDAS's spatial outputs



a) Coverages of 5 main land-use/cover types

b) Coverages of dense forest *vs.* road proximity

c) Areas of 3 main agricultural land-use types

d) Yields of 3 main agricultural land-use types

e) Per capita annual gross income

f) Gini indexes

Figure 6.19 Simulated land-use/cover and socio-economic changes for policy scenario S-Sub2 (high agrochemical subsidy) in comparison to scenario S0 (current trend – low agrochemical subsidy). Note: R-forest: rich/dense natural forest, P-Forest: poor/open natural forest. Source: data exported from VN-LUDAS's temporal outputs

6.4.5 Combinational policy impacts on LUCC and socio-economic dynamics

The simulation results of the scenarios S-Pro0, S0, and S-Pro3 suggest that the degree of forest degradation, especially in the area near roads and villages, is negatively related to the degree of enforcement of the forest protection rules, and disregarding the protection zone to certain extent may not be a serious deforestation problem. Whether a large remote mountain area is declared a protected area or not, almost no villager goes there to cut trees due to natural inaccessibility and extremely high transaction cost for withdrawing the timber products. An alternative for the current protection zoning policy may be: pay more attention and resources to enforce the protection rule in limited and critical areas, rather than claim a too large area for protection but with poor enforcement of the rule. Based on this alternative thinking, a zoning threshold of 12.0 score (about 50% of the total area is protected) and an enforcement degree of 80 % seem a fair proposal to construct combined scenarios. A threshold of more than 12.0 may not be politically realistic, as the percentage of protected area would become two small (see Figure 6.2) and policy-makers would not accept it.

The comparative analysis for scenarios of different subsidy access showed that no agrochemical subsidy is not a problem. Therefore, access to agrochemical subsidies for only of 5 % of the total population provided with agrochemicals, may be all that is needed for local communities.

The comparative analysis for scenarios of different extension coverage showed that it is best when about 35 % of the population are reached by extension services. However, rural developers may argue that more extension is always better for agricultural development of the community. We can test both ideas using the VN-LUDAS.

Based on such a-priori assumptions, we formed the two combinational scenarios COM1 and COM2 as showed in Table 6.2.

The simulated spatiotemporal land-use/cover maps in Figures 6.20 and 6.22, as well as the time-series graphs in Figures 6.21a, 6.21b, 6.23a, and 6.23b show that forest degradation is reduced both in scenarios COM1 and COM2 compared to the baseline scenario. In both cases, dense/rich forests increase due to positive succession from open/poor forests. The recovery rate of dense forest coverage also increases significantly in areas near roads/villages compared to the baseline scenario.

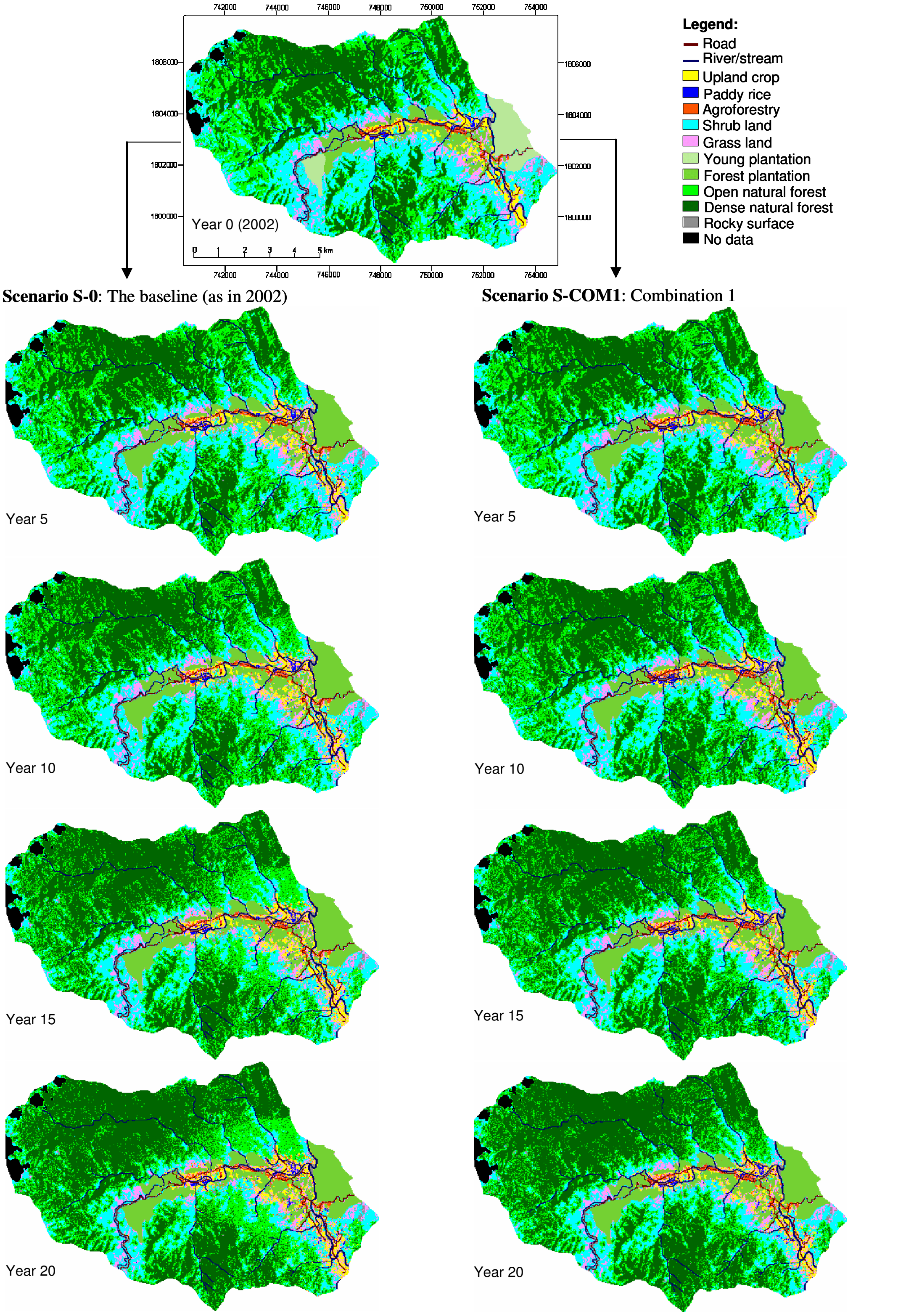
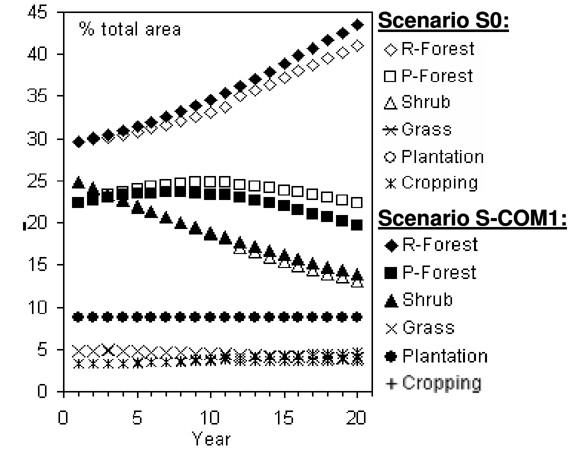
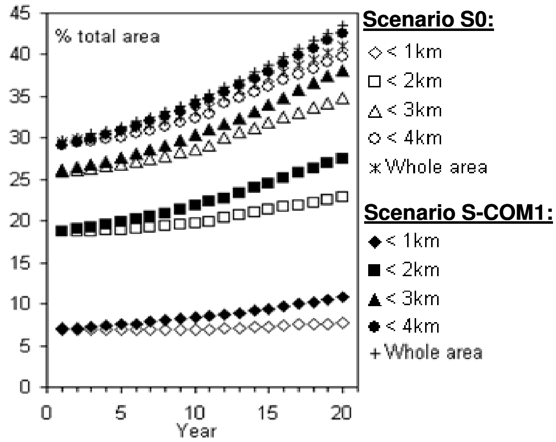


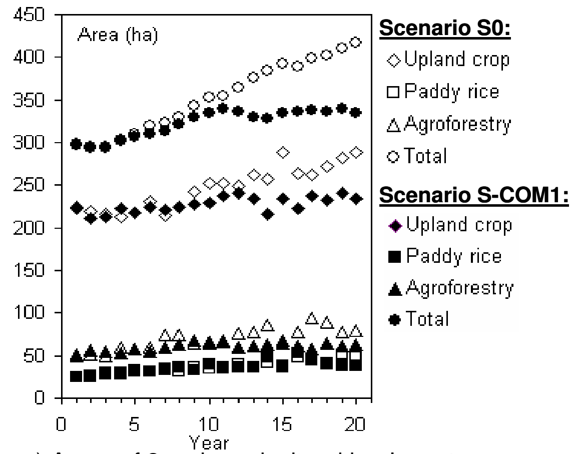
Figure 6.20 Simulated spatiotemporal land use/cover for policy scenario S-COM1 (combinational policy setting I: adjusted forest protection – high extension –minor agrochemical subsidy) in comparison to scenario S0 (the current trend). Source: data exported from VN-LUDAS's spatial outputs



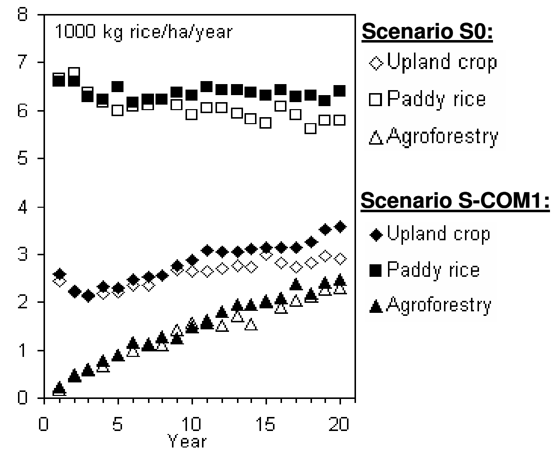
a) Coverages of 5 main land-use/cover types



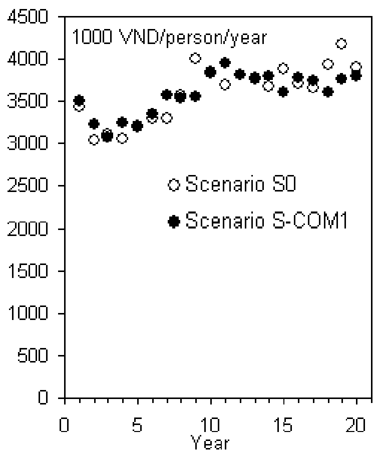
b) Coverages of dense forest *vs.* road proximity



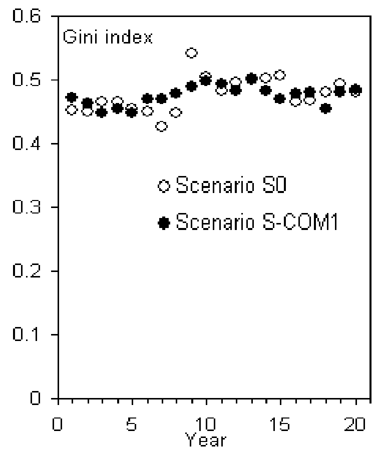
c) Areas of 3 main agricultural land-use types



d) Yields of 3 main agricultural land-use types



e) Per capita annual gross income



f) Gini indexes

Figure 6.21 Simulated land-use/cover and socio-economic changes for policy scenario S-COM1 (combinational policy setting I: adjusted protection – high extension – minor subsidy) in comparison to scenario S0 (current trend). Note: R-forest: rich/dense natural forest, P-Forest: poor/open natural forest. Source: data exported from VN-LUDAS's temporal outputs

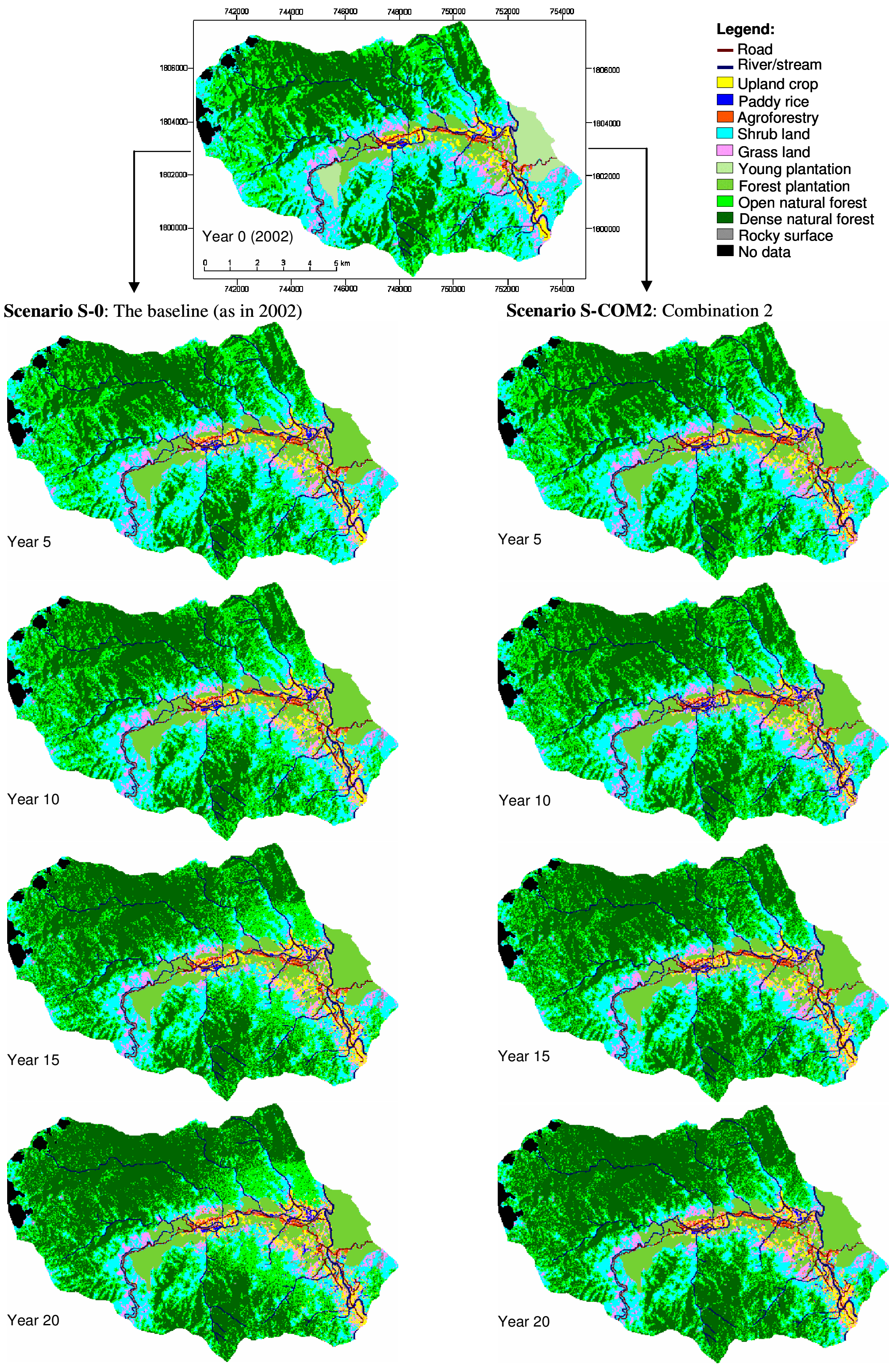


Figure 6.22 Simulated spatiotemporal land use/cover for policy scenario S-COM2 (combinational policy setting II: adjusted forest protection – low extension –minor agrochemical subsidy) in comparison to scenario S0 (the current trend). Source: data exported from VN-LUDAS's spatial outputs

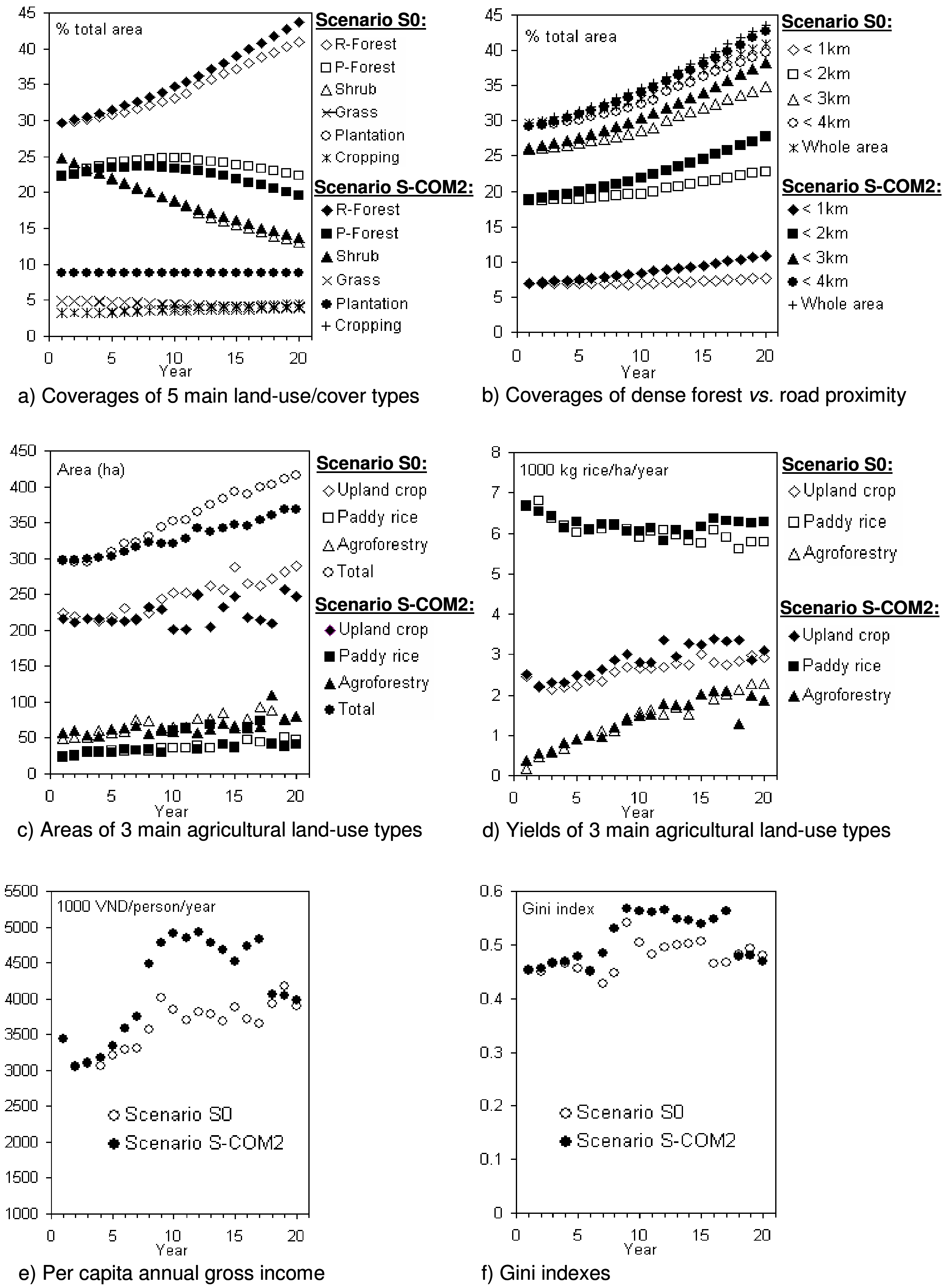


Figure 6.23 Simulated land-use/cover and socio-economic changes for policy scenario S-COM2 (combinational policy setting II: adjusted protection – low extension – minor subsidy) in comparison to scenario S0 (current trend). Note: R-forest: rich/dense natural forest, P-Forest: poor/open natural forest. Source: data exported from VN-LUDAS's temporal outputs

The simulated area of agricultural systems and agricultural productivity (Figures 6.21c, 6.21d, 6.23c, and 6.23d) shows a decrease in cropped land and slight increase in agricultural yields. The most dramatic difference between the two combinational scenarios is seen in the average household income (Figures 6.21e and 6.23e), as well as in the Gini index (Figures 6.21f and 6.23f). Whereas scenario COM1 does not deviate much from the baseline for these two indicators, scenario COM2 shows an increase of household income compared to the baseline, on average by about 15 %. However, the income inequality with scenario COM2 is also higher.

As in some of the previous cases, in-depth analysis of these integrated scenarios requires interpretation of the simulated world data to find underlying patterns and mechanisms for logically explaining the generated scenarios. This work is time demanding and requires micro-economic knowledge. In addition, a narrative analysis would be in order, requiring the modeler/users to visit the community and discus with stakeholders to seek explanations for the scenario outcomes. Therefore, we caution that although VN-LUDAS potentially is a good tool to generate quantitative scenarios of the co-evolution of the human and landscape systems, other quantitative interpretation and narrative studies are necessary to create meaningful and relevant scenarios.

6.5 Conclusions

This chapter aimed to implement an *operational* VN-LUDAS with functionalities of a modest DSS for land-use policy decisions in the Hong Ha watershed or other similar areas. It begins with an overview of the policy setting in the region, and procedures to identify "*puzzle decision points*" (i.e., *use cases*) of particular land-use policies in Hong Ha and policy interventions that should be tested for their likely impacts. We selected three important land-use policy issues and built the relevant policy parameters into the VN-LUDAS model.

We programmed the whole VN-LUDAS theoretical framework on a multi-agent modeling platform to produce an operational VN-LUDAS with functionalities of a decision support system (DDS) sensitive to land-use policies. The engine of the VN-LUDAS system is the MAS simulation program/protocol, which provides the scientific basis of the scenario development. The simulation program/protocol is coded in a multi-agent modeling platform (i.e., NetLogo) using an object-oriented design. With the object-oriented structure, the computer program is decomposed into several sub-programs/procedures, which can be tested independently and reused for other applications.

The VN-LUDAS system has an user-friendly visualization interface presenting the integrated scenarios. Simulation outputs are spatiotemporally explicit, including multi-temporal land-use/cover maps of the landscape environment and basic socio-economic indices of the community at different aggregate levels of human or landscape agents. By offering end-users possibilities to set policy parameters as needed and providing "*real-time*" responsive graphics, the system encourages interactive communications with users, thus enhancing their learning about the environmental consequences of human choices and facilitating rational consensus on actions.

We systematically developed integrated scenarios for different policy options with the purpose of focusing attention on the puzzle policy decision points (i.e., case uses). With the VN-LUDAS, we ran the scenarios in a systematic and organized process. Various policy options of watershed forest protection zoning, agricultural extension and agrochemical subsidy were cast in 10 policy scenarios. The results suggest that reducing the current proportion of protected area from 90 % to 50 % and increasing the enforcement of protection, together with provision of extension services for a third of the total population, and subsidizing 5 % of the population with agrochemicals ($ US 16 houeshold^{-1} year^{-1}) would, on average, increase per capita gross income by 15 % and significantly reduce forest degradation compared to the current scenario (i.e., the policy setting in 2002).

Although some technically-sound policy scenarios may be assessed with the VN-LUDAS, we submit that this kind of scientific reasoning is just one part of the information needed for actual decision making. This technical information supports stakeholders to develop options, enhances understanding, and evaluates the consequences of policy actions. However, in the end, human values must be applied on a participatory basis to determine what is a "*good*" policy for a given community on a specific issue.

7 CONCLUSIONS AND RECOMENDATIONS

7.1 Conclusions

Land-use and land-cover change (LUCC) is an essential environmental process that should be monitored and anticipated to provide a basis for proactive land management. However, studies on LUCC processes are often challenged because of the complex nature and unexpected behavior of human drivers and natural constraints. The aim of this thesis is to develop an integrated model that enables stakeholders in land management to explore alternative policy scenarios that may improve rural livelihoods and the environment, thereby providing them with support for making more informed decisions about land resources management.

The thesis begins with an in-depth review of multi-agent system simulation (MASS) as a new modeling paradigm of LUCC with coupled human-environment systems. The analysis shows that the complex the coupled nature of human-landscape system poses great methodological challenges for LUCC modeling, including the problems of scale dependencies, non-linear and transformative dynamics, socio-ecological heterogeneities, and emergent properties. So far we lacked an integrated modeling approach to overcome such problems. When reviewing relevant modeling approaches to different degrees of system complexity, we concluded that the MASS approach is well suited to capture the human-landscape system, which falls into the domain of organized complexity and likely resides on "*the edge of chaos*". The philosophy of the MASS as a new LUCC modeling paradigm built on MASS as an alternative to induction and reduction of doing science. MASS calls for a comprehension and clarification of the methods for system conceptualization before putting an intensive effort on detailed MAS exercises. Simulation as a virtual experimental vehicle for understanding system behaviour renders MASS especially adequate for scenario studies on LUCC for decision support purposes.

We conceptualized a MAS for representing the coupled human-landscape system in rural forest margins, named VN-LUDAS (Vietnam – Land-use Dynamics Simulator). VN-LUDAS falls into the class of "*all agents*" systems, in which the human population and the landscape environment are all self-organized interactive agents. The biophysical system is considered at the level of landscape agent, i.e., heterogeneous land patches with their own attributes and ecological response mechanisms with respect to environmental changes and human interventions. The human system is considered in terms of household agents, i.e., heterogeneous farm households with their own characteristics and decision-making mechanisms regarding land use. Interactions between household and landscape agents occur mainly through tenure relations and a perception-response loop. Tenure relations are institutional rules that regulate the household's access to land resources. The perception-response loop involves information flows between households and patches. The information flowing from household to patch reflects the decisions made by the household on land use on the patch. The information flowing from patch to household corresponds to the perceived bio-physical state and benefits that the household can derive from the use in arriving at decisions. Policy and other macro-drivers influence the system behaviour though modifying the functional relationships between the human and environmental system.

At a theoretical specification level, we divided the VN-LUDAS framework into four modules that represent the main components of the coupled human-landscape system in forest margins. The human module defines specific behavioral patterns of farm households (i.e., human agents) in land-use decision-making according to typological livelihood groups. The landscape module characterizes individual land patches (i.e., landscape agents) with multiple attributes, representing the dynamics of crop and forest yields and land-use/cover transitions in response to both household behavior and natural constraints. The policy module represents public policy factors that are assumed to be important for land-use choices. The decision-making module integrates household, environmental and policy information into land-use decisions.

The model specification, module-by-module and object-by-object, clearly shows an explicit and fully parameterized architecture, which accounts for the evolution of the coupled human-environment systems. In this first version of the model we nested the bounded-rational approach based on utility maximization using spatial multi-nominal logistic functions with heuristic rule-based techniques to represent decision-making mechanisms of households regarding land use. The proposed agent-based architecture allows integration of diverse human, environmental and policy-related factors into farmers' decision-making on land use and projection of subsequent accumulated outcomes in terms of spatiotemporally explicit patterns of the natural landscape and population. As the model's architecture is illustrated using graphic language, and the parameterization is in algebraic language, the model has better applicability to different contexts. Although many features of the complex processes of human decision-making have not yet been included, the agent-based system has a built-in flexibility for adaptation, upgrading and modification.

We calibrated and verified the the land-use choice model for heterogeneous household agents using standard inferential statistics. First, we presented methods for collecting household – parcel data that meet the requirements of MAS-LUCC modeling (i.e., VN-LUDAS model). A database of household – holding parcel for one third of the population in the study area was obtained through relevant field survey techniques, ranging from participarory rural appraisals (PRAs) to structured household interviews (aided by GIS and remote sensing technologies). The database includes household characteristics that are representative for different asset categories of the livelihood framework, biophysical characteristics of holding parcels and information indicating whether the households were subjected to particular policies. Second, we performed a sequential process of applying multi-variate statistical methods to categorize households and estimate the land-use choice model for each household category. Principal component analysis (PCA) followed by k-mean cluster analysis (k-CA) and descriptive statistics explicitly categorized and characterized the heterogeneous population. Furthermore, multi-nominal logistic regression analyses estimated the effects of these households and spatial variables on land-use decision-making of every household typological group.

The methods used for the household study could capture considerable heterogeneities in land-use choice behavior in the study community, and rigogously parameterized these heterogeneities. Key variables explaining most variations of household livelihoods which were extracted by PCA were used as criteria for regularly categorizing households in the VN-LUDAS. In general, households of all groups choose land use based on the mutual considerations of a range of personal characteristics, natural conditions of the environment, and particular policy factors. Therefore, the developed model of land-use choice provides a basis for coupling the human-environment systems under particular policy circumstances when simulating land-use changes. When applying these land-use choice analysis results to the VN-LUDAS in the study area, both the estimated coefficients and the standard errors of the estimates were used for the computation of land-use choice probabilities. Each household agent adopts random values of preference coefficients around the group's coefficient, bounded by the standard errors. Therefore, the land-use choice behaviors of households fluctuate within a behavioral template if they are in the same group, but are structurally different if they are in different groups.

We calibrated the heterogeneous landscape environment of the study site using remote sensing and GIS-based analyses. The important point addressed was how to capture the landscape reality utilizing available objective spatial datasets (e.g., remote sensing images and topographic maps) and a rigorous approximation of spatial variables that are hard to directly measure in the field (e.g., soil-water distribution). Because the path-dependent nature of land-use changes requires careful and accurate calibration of initial land-use/cover, current land-use/cover data were objectively extracted from fine-resolution satelite images (e.g., Landsat ETM and Aster). Soil-water spatial distribution, which is difficult to capture by field measurement, were proximated by topographic variables (e.g., slope gradient, upslope contributing area, and wetness index), which were extracted and calculated from a topographic map (i.e., UTM map 1:50,000).

We developed ecological models that were built into the landscape agents to enable them to respond to environmental changes and human interventions. The development of the empirical/statistical sub-model for agricultural yield dynamics is a typical example, illustrating how to take advantage of empirical data to build a bio-economic model that performs non-linear responses of landscape agents along three dimensions of heterogeneity: i) spatial heterogeneity, ii) temporal dynamics, and iii) household agent diversity. In contrast, the development of the theoretical sub-model for forest productivity dynamics is another example of "*theoretical guessing*" in the case of absence of data. The development of the sub-model for natural transitions of vegetation is an illustration of the modeling of land conversion (i.e., categorical/discrete change) derived from modification/growth (i.e., micro/continuous change) and ecological edge effects (i.e., neighborhood effects). Also, through the development of the three sub-models above, we explicitly arranged the possibility to couple different traditional modeling techniques into the modeling of landscape agents, including statistic (e.g., the sub-model of agricultural dynamics), system dynamics (e.g., the sub-model of forest yield dynamics) and cellula automata (e.g., the sub-model of natural transition of vegetation) approaches.

In sum, by building and calibrating sub-models for household and landscape agents we represented the human-environment system in a dynamic, adaptive, and realistic manner. The dynamic agent is such that it was designed to host processes operating in them; thus agents are able to change in ways beyond the control of other agents. The adaptive agent is such that it has specific capabilities to interact with and respond to the changes of the surrounding environment, including other agents. A realistic agent is one that which state and behavioral parameters are empirically grounded on real data.

We programmed the whole VN-LUDAS theoretical framework on a multi-agent modeling platform (i.e., NetLogo 2.1 (Wilenski (1999)) to produce an operational VN-LUDAS with functionalities of a decision support system (DDS) for particular land-use policies. The engine of the VN-LUDAS system is the MAS simulation program/protocol that provides the scientific basis for scenario development. The simulation program/protocol is coded using object-oriented design. With the object-oriented structure, the whole computer program is composed of several sub-programs/procedures, which can be retested independently and reused for other applications. From a model coupling view, VN-LUDAS is thus a tight (close) coupling of several small micro-economic and ecological sub-models.

Simulation outputs are spatiotemporally explicit multi-temporal land-use/cover maps of the landscape and basic socio-economic indices of the community at different aggregate levels of human/landscape agents. The VN-LUDAS system has an user-friendly visualization interface representing the integrated scenarios. By offering end-users possibilities to set policy parameters as needed and with a "*real-time*" responsive graphic interface, the system encourages interactive communications among users, thus enhancing their learning about environmental consequences of human land-use choices. The simulated world and the underlying data of all maps and graphs at any point in time can be exported to electronic files for further analyses/interpretations using standard GIS (e.g., ArcView and ArcInfo) and statistical packages (e.g., SPSS or S-Plus). Therefore, the system also shows a good communication capacity to academic users (e.g., students and researchers).

Integrated scenarios were developed for different policy options with the purpose of focusing attention to puzzle policy decision points (i.e., use cases). With the support of the VN-LUDAS, we conducted the scenarios development in a systematic and organized process. First, we defined the current policy setting to construct a *baseline scenario*. Second, each policy factor was shifted from the baseline to form a *scenario spectrum* (including 3 or 4 scenarios) to assess the impacts of the change in such a policy. Based on the comparison of the scenario results of single tests, we formed 2 complex scenarios that combine the most promising scenarios for each policy factor. For the policy areas of watershed zoning for forest protection, agricultural extension and agrochemical subsidies, we developed scenarios for 10 different policy options. The results suggest that reducing the current proportion of protected area from 90 % to 50 % and increasing the enforcement of protection, together with the provision of extension services to a third of the total population (from currently 67 %), as well as subsidizing 5 % of the population with agrochemicals ($ US 16 household^{-1} year^{-1}) (instead of 23 % as current) will, on average, increase per capita gross income by 15 % and significantly reduce forest degradation compared to the current scenario (i.e., policy setting in 2002).

Although technically-sound policy scenarios may emerge from using the VN-LUDAS, we argue that this is rather scientific reasoning and is just one part of the information needed for actual decision making. It provides information to stakeholders on policy options, and their consequences. However, in the end, human judgement must be applied to determine what is a “good” policy for a given community on a specific issue.

Finally, by passing through a cycle of a model-development process with the VN-LUDAS, we have produced a first version of a decision support tool that is based on a multi-agent system framework. This system can help stakeholders in land management planning and explore alternative policy options to improve rural livelihoods and the environment. The model was validated through the transparency and the scientific rigidity of the model-development process. The whole model development process gives an insight into how suitable the multi-agent system is for the study of the complex processes of land-use and land-cover change.

7.2 Limitations

This first version of the VN-LUDAS certainly has limitations. The first limitation is that some substantial interactions in both, the human and biophysical systems have not been explicitly represented. When accessing and using land resources, household agents often face the need to reach agreement on a variety of issues (e.g., the acquisition and use of common lands) and to exchange particular contracts, goods, and services (e.g., the exchange of labor in agricultural productions, the transfer of land rights, and sharing of agricultural benefits). Such negotiations among households are indispensable for their decision making. The formulization of human negotiations, therefore, has received a great deal of attention from the MAS community (Lomuscio *et al*., 2003). However, the current version of the VN-LUDAS has as yet no mechanism for these high-level human interactions yet. Moreover, in the formulizations of the landscape environment, surface processes (e.g., soil erosion and deposition) and their interactions with socio-economic components (e.g., responses and impacts of soil erosion/deposition on the socio-economic status of households) have not explicitly been modeled.

The second limitation is that this first VN-LUDAS version does not give users the choices between different decision-making mechanisms. As many competing techniques and theories for modeling human decision making exist, it makes sense for the VN-LUDAS to offer users a greater choice of appropriate decision-making mechanisms for different situations. However, in this first version of the VN-LUDAS, only a single decision-making mechanism of household agents is specified.

The third limitation is that the statistical approach for obtaining the parameters of household's land-use choices is somewhat data intensive in practice. The proposed methods for the classification of typological household groups and the estimation of land-use choice parameters required plot-explicit household interviews that are time consuming, especially when the study area/community is large. Therefore, one needs to look for some alternative methods with robust practical applications.

The fourth limitation is that the VN-LUDAS is only a tool for developing quantitative scenarios of LUCC and the associated socio-economic changes. The narrative (story) line and in-depth quantitative interpretations for the scenario outcomes must be done through follow-up social and micro-economic studies based on the simulated datasets.

Last, but not least, validation of the VN-LUDAS is still an open issue that has not yet been rigorously analysed in this thesis. Actually, the validation of MAS models is currently still a debatable issue. While classical validation methods, e.g., sensitivity analysis and comparing simulated data versus observed data, have turned out not suitable for MAS models (see Section 2.7), a number of different validation strategies are proposed (see Bousquet and Le Page, 2004; Parker *et al*., 2003) and debated.

7.3 Recommendations

Because no method is universally appropriate, our VN-LUDAS development will be version-by-version in accordance with the advancing knowledge and technologies, as well as regional contexts. The VN-LUDAS can be further developed in three following ways. First, from a methodological view point, the VN-LUDAS should be improved version by version in terms of its representative structure and simulation protocol. This would involve the addition of components, the fine tuning of the agent architecture, algorithms and the simulation program. After passing each round of the model development process as in Figure 2.6 (Chapter 2), an *improved version* of VN-LUDAS will be produced. Second, from the geographical context, the current VN-LUDAS can be adapted to socio-ecological conditions of other regions through *model variants*. Finally, putting the VN-LUDAS into practice in different contexts will create a basis for – yet unforeseen - further improvements.

Possible methodological extensions to the first version of the VN-LUDAS

- *Building some common alternative decision-making sub-models into household agents for wider choices of model users*. More research should be done on the formulization of different household decision-making strategies to examine whether particular agent decision-making formulizations are appropriate for particular decision-making situations. The VN-LUDAS should at the same time support some common methodological choices for household decision making. One alternative could be the use of *anthropological rule systems* as in the decision-making component of FLORES (Forest Land Oriented Resource Envisioning System) developed by scientists in Center for International Forest Research (CIFOR) (Haggith *et al*., 2004). Another alternative could be the BDI (Belief – Desire –

Intention) architecture (Woodridge, 1999; Bousquet and Le Page, 2004). The BDI normally assumes that human agents have beliefs about other human agents and environments coming from their individual experience or as a result of local reputation, which are used to guide their level of commitment to collective resource management (Bousquet and Le Page, 2004, Rouchier *et al*., 2001).

- *Adding a mechanism of negotiation and exchange of land goods and services to the decision-making sub-model*. Automated negotiations are a key form of interactions in agent-based systems, and such negotiations involve the design of high-level interaction protocols. A negotiation mechanism consists of a negotiation protocol together with the negotiation strategies for agents involved. The main parameters on which any automated negotiation depends and representative samples of some of the most prominent negotiation models are well identified and characterized in Lomuscio *et al*. (2003) and Van Dyke Parunak (1999), providing a theoretical basis for algorithm specifications of negotiations on tenure rights, goods and services of land resources in forest margins. Some simple mechanisms of negotiations on land resources, such as the re-allocation of common lands and sharing of crop/forest products, can be found in Haggith *et al*. (2004).
- *Adding a learning mechanism to the decision-making sub- model*. Learning capacity is an inherent characteristics of human decision making, and many learning models have been developed. For example, the Q-learning model (a form of reinforcement learning) is described in Kaelbling and Littman (1996), Watkins and Dayan (1992); and the EWA (experience-weighted attraction) learning model is proposed by Camerer and Ho (1999). These learning models can be adopted and modified to apply to the decision process on land allocations.
- *Building a sub-model of soil erosion/deposition and its impacts on household economy into landscape agents.* There are a number of options for this biophysical extension of the VN-LUDAS. One option is to the incorporation of one among a large number of available empirical soil erosion models, e.g., models based on Universal Soil Loss Equations (USLE). However, because USLE-based models are not temporally explicit, this will exclude temporal dynamics of soil erosion. An alternative is the use of a cellular automata (CA) model of soil erosion/deposition. However, this class of soil erosion models is in an early development stage. A

recent example of a CA soil erosion model specified in D'Ambrosio *et al*. (2001) could be implemented in VN-LUDAS. However, even if a CA soil erosion/deposition model is incorporated in landscape agents, the next challenge will be how to quantify the impacts soil erosion on agricultural productivity, and the social response of farmers to the soil loss phenomena. Although many models for crop yield response to soil erosion/deposition exist, but they are very data demanding for model parameter calibrations. Therefore, approximation of such complicated parameters sets based on a solid theoretical basis may need to be considered.

Model validation strategies

How to assess the credibility of MAS-LUCC models (especially in the context of scenarios studies) is still an open and debatable issue and a subject for further research. We proposed to verify the credibility of the VN-LUDAS using the following strategies given in current literatures (see Parker *et al*., 2003; Bousquet *et al*., 2004):

- *Rigorous representations of the structure of the system being modeled*. As analysed in Chapters 1 and 2, the credibility of MAS simulation models depends on how the model represents the structure of the system modeled. Therefore, rigorous verification of the model design and algorithms is important for model validations. The formulization and calibration of decision-making models play a crucial role in LUDAS validation, as the behavior of the whole system is derived from interactions governed by the formulized decision-making mechanism. For that purpose, a clear graphic presentation of the model architecture for expert assessments and comparative studies from model to model is required (Bousquet and Le Page, 2004; Parker *et al*., 2003; Axelrod, 1997). A graphic language commonly used for designing MAS models is the Agent Unified Modeling Language (AUML) (see Bauer, 2002). The design of the VN-LUDAS model using AUML also would allow easier implementation of the model in other platforms (e.g., SWARM, REPAST, ASCAPE, etc.).
- *Comparing the results of the LUDAS with other types of models*. An equivalence of the simulated results with analytical or empirical results may enhance the credibility of the model. For example, the current VN-LUDAS could be run in

Hong Ha over the period 1980 – 2000, and then the simulated results should be compared to the historical data on LUCC and community changes.

- *Rigorous calibration of interaction rules and the initial state using real measured data*. Because the behaviour of MASS models is stemming local interactions are path-dependent, careful calibration of parameters of interaction rules and the input data for system initialization are also crucial to increase the realism of the model. As sub-models are empirically calibrated/tested, input data are generated using scientific inferential methods, the model will likely gain greater acceptance and use.

Adaptation of the VN-LUDAS to different geographic regions

The adaptation the VN-LUDAS to different geographic regions, for instance to other areas in the Vietnam uplands, should firstly be based on a regional classification of socio-ecological conditions. The regional classification will structure a large region into many socio-ecological compartments that belong to a number of typological sub-regional units. Then, variants of the VN-LUDAS can be developed for each typological sub-regional through adjustment in structures and re-calibration of parameters of sub-models.

Putting the VN-LUDAS into practice for land management and planning

It is necessary that stakeholders on study sites actually make use of the VN-LUDAS to explore policy options and provide feedbacks on the model's performance and relevance. This requires a commitment to a client-based approach, to ensure that the VN-LUDAS can do what the stakeholders actually require. This effort includes working with potential users to check if the policy levers and indicators of performance are important to them, and how the simulation results (i.e., scenarios) should be presented in the most comprehensive way.

The VN-LUDAS can be used either as a stand-alone decision support tool for land management, or as a partial model in a larger decision support system. The use of the VN-LUDAS as a stand-alone land-use decision support tool was illustrated in Chapter 6. The use of the VN-LUDAS in conjunction with larger LUCC models or decision support systems can be done in two ways. First, the VN-LUDAS can be loosely coupled with other models in the system, i.e., the VN-LUDAS is positioned in the control system that allows it to exchange information/data with other models. Alternatively, the VN-LUDAS can be run separately to generate behavioural rules for every typological sub-regional unit, and then such generated rules can be used for other large-scale models of LUCC, hydrology, or climate changes.

Participation of stakeholders in the use of the VN-LUDAS is fostered by its client-oriented and user-friendly graphic interface. A possibility for farmer participation in the calibration of the VN-LUDAS is that the preference coefficients (β) in the land-use choice functions can be obtained through semi-quantitative participatory tools (e.g., matrix ranking) for each typological household group, instead of extracting such parameters through regression analysis. Then, an interesting comparative study would be to compare the simulation results of participatory-driven and the regression-based land-use choice models. Strength of the participatory-driven land-use choice model is that it is very much less data demanding, thus more acceptable for use by local stakeholders.

8 REFERENCES

Achard FD, Eva HD, Stibig H, Mayaux P, Gallego J, Richards T, and Malingreau J (2002) Determination of deforestation rates of the world's humid tropical forests. Science 297: 999 – 1002.

Agarwal C, Green G L, Grove M, Evans T, and Schweik C (2000) A review and assessment of land-use change models: dynamics of space, time, and human choice. Center for the Study of Institutions, Population, and Environmental Change (CIPEC), Indiana University, USA.

ALC (Advanced Learning Center) (2003) Agent and rules: understanding rules and interactions in complex adaptive system. Web-based Lecture Notes [http://netlogo.modelingcomplexity.org/about_cas.htm]. Grand Junction, Colorado: Advanced Learning Center (ALC), Mesa State College.

Alder D (1996a) Models for basal area dynamics of mixed tropical forest: neo-tropical experience and prospects for application in Ghana. In: Foli EG, Vanclay JK, and Ofosu-Asiedu A (eds) IUFRO conference on "Growth Studies in Moist Tropical Forests in Africa", Kumasi, 11-15 November. Forest Research Institute of Ghana, Kumasi, Ghana, pp. 93 – 105.

Alder D (1996b) SIRENA: a simulation model for tropical forest management in Northern Costa Rica - user's guide and technical notes. CODEFORSA/ODA Integrated Natural Forest Management Project, Costa Rica.

Alder D (1997) The ITTO permanent sample plots in Papua New Guinea: progress in growth model development. ITTO/PNG Forest Research Institute, Lae, Papua New Guinea.

Alder D (1998) PINFORM: a growth model for lowland tropical forests in Papua New Guinea. Consultancy Report, ITTO/PNG Project PD 162/91. Forest Research Institute, Lae, Papua New Guinea.

Alder D and Silva JNM (2000) An empirical cohort model for *terra firme* forests in Brazilian Amazon. Forest Ecology and Management 130: 141 - 157

Allen TFH and Starr TB (1982) Hierarchical-perspectives for ecological complexity, University of Chicago Press, Chicago, USA.

Alther C, Castella JC, Novosad P, Rousseau E, and Tran Trong Hieu (2002) Impact of accessibility on the range of livelihood options available to farm households in mountainous areas of Northern Viet Nam In: Castella JC and Dang Dinh Quang (eds) Doi Moi in the mountains: land use changes and farmers' livelihood strategies in Bac Kan Province, The Agricultural Publishing House, Hanoi, Vietnam, pp. 121 – 148.

Angelsen A and Kaimowitz D (1999) Rethinking the cause of deforestation: lessons from economics models. The World Bank Research Observer 14 (1): 73 - 98

April J, Glover F, Kelly JP, and Laguna M (2003) Practical introduction to simulation optimization. In: Chick S, Sánchez PJ, Ferrin D, and Morrice DJ (eds) Proceedings of the 2003 Winter Simulation Conference, New Orleans, 7-10 December. Institute of Electrical and Electronics Engineers, New Jersey, USA, pp. 71-78.

Aronson D (1998) Overview of systemic thinking. [http://www.thinking.net/Systems_Thinking/]

Ashley C and Carney D (1999). Sustainable livelihoods: lessons from early experience. DFID, London, UK.

Axelrod RM (1997) The complexity of cooperation: agent-based models of competition and collaboration. Princeton University Press, Princeton, USA.

Axelrod RM (2003) Advancing the art of simulation in the social sciences. Japanese Journal for Management Information Systems [in press].

Balmann A (1997) Farm based modelling of regional structure change: a cellular automata approach. Eur Rev Agric Econ 24: 85 – 108.

Barreteau O, Bousquet F, and Attonaty J (2001) Role-playing games for opening the black box of multi-agent systems: method and lessons of its application to Senegal River Valley irrigated systems. Journal of Artificial Societies and Social Simulation 4 (2). [http://www.soc.surrey.ac.uk/JASSS/4/2/5.html]

Bartelink HH (2000) Effects of stand composition and thinning in mixed-species forests: a modeling approach applied to Douglas-fir and beech. Tree Physiology 20: 399 – 406.

Batty M (2001) Editorial: agent-based pedestrian modeling. Environment and Planning B 28 (3): 321 – 326.

Batty M and Torrens P (2001) Modeling complexity: the limits to prediction. CASA Working Paper Series 36, London, UK.

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

Beckman M, Le Van An, and Le Quang Bao (2002) Living with the floods: coping and adaptation strategies of households and local institutions in central Vietnam. Stockholm Environment Institute, Stockholm, Sweden.

Becu N, Perez P, Walker B, Barreteau O and Le Page C (2003) Agent-based simulation of a small catchment water management in northern Thailand: description of the landscape model. Ecological Modelling 170: 319 – 331.

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

Berger T (2001) Agent-based spatial models applied to agriculture: a simulation tool for technological diffusion, resource use change and policy analysis. Agricultural Economics 25 (2-3): 245 – 260.

Bernabo JC (1998) Developing policy-relevant global climate change research. In: Mickler RA and Fox S (eds) The productivity and sustainability of Southern Forest Ecosystems in a changing environment. Ecological Studies 128: 55 – 70.

Bonabeau E (2002) Agent-Based Modeling: Methods and Techniques for Simulating Human Systems. PNAS 99(2): 7280 – 7287.

Bonabeau E, Dorigo M, and Theraulaz G (1999) SWARM intelligence: from natural to artificial systems. Oxford University Press, New York, USA.

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

Bousquet F, Le Page C, Antona M, and Guizol P (2000) Ecological scales and use rights: the use of multi-agent systems. In: XXIth IUFRO World Congress on “Forest and society: the role of research”, Kuala Lumpur, 2000. Kuala Lumpur, Malaysia, pp. 730 - 742.

Box PW (2002) Spatial units as agents: making the landscape an equal player in agent-based simulations. In: Gimblett HR (ed) Integrating geographic information systems and agent-based modeling techniques for understanding social and ecological processes. Oxford University Press, Oxford, UK, pp. 59 – 82.

Briassoulis H (2000) Analysis of land use change: theoretical and modeling approaches. Regional Research Institute, West Virginia University. [http://www.rri.wvu.edu/WebBook/Briassoulis/contents.htm].

Bui Dung The (2001) The economics of soil erosion and the choice of land use systems by upland farmers in central Vietnam. Economy and Environment Program for Southeast Asia (EEPSEA).

Bui Dung The (2003) Land-use systems and erosion in the uplands of the Central Coast, Vietnam. Environment, Development and Sustainability 5: 461 – 476.

Camazine S, Deneubourg JL, Franks NR, Sneyd J, Theraulaz G, and Bonabeau E (2001) Self-organization in biological systems. Princeton University Press, Princeton, NJ, USA.

Camerer C and Ho TH (1999) Experience-weighted attraction learning in normal form games. Econometrica, 67 (4): 827 – 874.

Campbell B, Sayer JA, Frost P, Vermeulen S, Ruiz Pérez M, Cunningham A, and Prabhu R (2001) Assessing the performance of natural resource systems. Conservation Ecology 5 (2): 22 [http://www.consecol.org/vol5/iss2/art22]

Carson MA and Kirkby MJ (1972) Hillslope form and process. Cambridge University Press, Cambridge, UK.

Castella JC and Erout A (2002) Montane paddy rice: The cornerstone of agricultural production systems in Bac Kan Province, Vietnam. In: Castella JC and Dang Dinh Quang (eds) Doi Moi in the Mountains: Land use changes and farmers' livelihood strategies in Bac Kan Province, Vietnam. The Agricultural Publishing House, Hanoi, Vietnam, pp. 175 – 196.

Castella JC, Tran Trong Hieu, and Eguienta Y (2002a) Combinations of participatory landscape analysis and spatial graphic models as a common language between researchers and local stakeholders. SAM Paper Series 12.

Castella JC and Tran Quoc Hoa, Husson O, Vu Hai Nam, and Dang Dinh Quang (2002b) The declining role of ethnicity in farm household differentiation: A case study from Ngoc Phai Commune, Cho Don District, Bac Kan Province, Vietnam. In: Castella JC and Dang Dinh Quang (eds) Doi Moi in the mountains: land use changes and farmers' livelihood strategies in Bac Kan Province, Vietnam. The Agricultural Publishing House, Hanoi, Vietnam, pp. 47 – 71.

Castella JC, Tronche NR, and Vu Nguyen (2002c) Landscape changes in Cho Don District during the Doi Moi era (1990 – 2000) and their implications for sustainable natural resource management in Vietnam's mountainous provinces. In: Castella JC and Dang Dinh Quang (eds) Doi Moi in the mountains: land use changes and farmers' livelihood strategies in Bac Kan Province, Vietnam. The Agricultural Publishing House, Hanoi, Vietnam, pp. 149-174.

Castella JC, Tronche NR, and Vu Nguyen (2002d) Scaling up local diagnostic studies to understand development issues in a heterogeneous mountain environment; an introduction to the SAM Program. In: Castella JC and Dang Dinh Quang (eds) Doi Moi in the mountains: land use changes and farmers' livelihood strategies in Bac Kan Province, Vietnam. The Agricultural Publishing House, Hanoi, Vietnam, pp. 1 - 18.

Chapman DG (1961) Statistical problems in dynamics of exploited fisheries populations. In: Neyman J (ed) Proceedings of the fourth Berkeley Symposium on mathematical statistics and probability (Vol. 4). University of California Press, California, USA, pp. 153 – 168.

Chase TN, Pielke RA, Kittel TGF, Nemani RR, and Running SW (1999) Simulated impacts of historical land cover changes on global climate in northern winter. Climate Dynamics 16: 93 -105.

Chomitz KM and Gray DA (1996) Roads, lands, markets, and deforestation: a spatial model of land use in Belize. The World Bank Economic Review 10(3): 487 – 512.

Cioffi-Revilla C and Gotts NM (2003) Comparative analysis of agent-based social simulations: GeoSim and FEARLUS models. Paper prepared for the Model-to-Model Workshop on comparing multi-agent based simulation models. Claudio Cioffi-Revilla and Macaulay Institute, Marseilles, France.

Conacher AJ and Dalrymple JB (1977) The nine unit landsurface model: an approach to pedogeomorphic research. Geoderma 18: 11 – 54.

Costanza R and Gottlieb S (1998) Modeling ecological and economic systems with STELLA (part II). Ecological Modelling 112: 81 – 84.

Costanza R and Ruth M (1998) Using dynamic modeling to scope environmental problems and build concensus. Environmental Management 22: 183 – 195.

Creed IF, Trick G, Band E, and Morrison K (2002) Characterizing the spatial pattern of soil carbon and Nitrogen pools in the turkey lakes watershed: a comparison of regression techniques. Water, Air, and Soil Pollution 2: 81 – 102.

Cropper M, Griffiths C, and Mani M (1997) Roads, population pressures, and deforestation in Thailand, 1976–89. Policy Research Working Paper 1726. Policy Research Department, World Bank, Washington D.C., USA.

Dale VH, Southworth F, O'Niell RV, Rosen A, and Frohn R (1993) Simulating spatial patterns of land use change in Rondinia, Brazil. Lectures on Mathematics in the Life Sciences 23: 30 – 55.

D'Ambrosio D, Di Gregorio S, Gabriele S, and Gaudio R (2001) A cellular automata model for soil erosion by water. Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere 26 (1): 33 – 39.

de Koninck R (1999) Deforestation in Vietnam. International Development Research Centre (IDRC), Ottawa, Canada.

De Roo APJ (1998) Modelling runoff and sedimentation transport in catchments using GIS. Hydrological Processes 12: 905 – 922.

Deadman PJ (1999) Modelling individual behaviour and group performance in an intelligent agent-based simulation of the tragedy of the commons. Journal of Environmental Management 56: 159 – 172.

Dietz T and Stern PC (1998) Science, values, and biodiversity. Bioscience 48(6): 441 – 445.

Do Dinh Sam (2001) Scientific bases adding silviculture techniques to increase productivity of natural forest after harvesting and plantations. In: Research results of forest science and technology 1996 – 2000 by Forest Science Institute of Vietnam. The Agricultural Publishing House, Hanoi, Vietnam.

Do Dinh Sam and Nguyen Ngoc Binh (2000) Assessment of potential productivity of forest lands in Vietnam. Statistical Publishing House, Hanoi, Vietnam.

Dolman AJ and Verhagen A (2003) Land use and global environmental change. In: Dolman AJ and Verhagen A (eds). Global environmental change and land use. Kluver Academic Publishers, Dordrect, The Netherlands, pp. 3 – 13.

Dorigo M, Di Caro G, and Gambardella LM (1999) Ant algorithms for discrete optimization. Artificial Life 5(3): 137 – 172.

Dumanski J and Craswell ET (1998) Resource management domains for evaluation and management of agro-ecological systems. In: International workshop on resource management domains. International Board for Soil Research and Management, Bangkok, Thailand.

Easterling WE and Kok K (2003) Emergent properties of scale in global environmental modeling - are there any? In: Rotmans J, Rothman D S, Rothman D S (eds) Scaling in integrated assessment. Swets & Zeitlinger Publishers, The Netherlands, pp. 263 – 292.

Eastmann JR (2001) Guide to GIS and image processing (volumm 2). Manual of IDRISI32 Release 2. Clark Labs, Clark University, Boston, USA.

EMWG (Ethnic Minorities Working Group in Vietnam) (2004) Map of Ethnic Minorities in Vietnam. [http://www.emwg.org.vn/images/ethnicity.jpg]

Eoyang G and Berkas T (1998) Evaluating Performance in a Complex Adaptive System. In: Lissack M and Gunz H (eds.) Managing complexity in organizations. Quorum Books, Westport, Connecticut, USA.

Epstein JM and Axtell R (1996) Growing artificial societies: Social science from the bottom up. MIT Press, Cambridge, MA, USA.

Faculty of Forestry (at Hue University of Agriculture and Forestry) (2002) Raw data sheets of plantation plots in Hong Ha commune in 2002. Faculty of Forestry, Hue University of Agriculture and Forestry, Hue, Vietnam.

Farrington J, Carney D, Ashley C, and Turton C (1999) Sustainable livelihoods in practice: early applications of concepts in rural areas. Natural Resource No. 42. ODI, London, UK.

Farrow A and Nelson A (2001) Accessibility modelling in ArcView 3.0: an extension for computing travel time and market catchment information. CIAT, Cali, Colombia.

Ferber J (1999) Multi-agent systems: an introduction to distributed artificial intelligence. Addison-Wesley, Reading, MA, USA.

Fernandez LE, Brown DG, Marans RW, and Nassauer JI (2003) Characterizing location preferences in an exurban population: implications for agent-based modeling. Environment and Planning B, Special Issue on Complex Systems. [in review].

Forrester JW (1980) Principles of systems (2nd edition). MIT Press, Cambridge, MA, USA.

Forrester JW (1987) Lessons from system dynmic modeling. System Dynamics Review 3 (2): 136 – 149.

Forrester JW (1995) The beginning of system dynamics. The McKinsey Quarterly 4: 4 – 16.

Fox F, Krummel J, Yarnasarn S, Ekasingh M, and Podger N (1995) Land use and landscape dynamics in Northern Thailand: assessing change in three upland watersheds since 1954. Ambio 24(6): 328 – 334.

Fox J, Kanter R, Yarnasarn S, Ekasingh M, and Jones R (1994) Farmer decision making and spatial variables in Northern Thailand. Environmental Management 18(3): 391 – 399.

Fox J, Truong DM, Rambo AT, Tuyen MP, Cuc LT, and Leisz S (2000) Shifting cultivation: a new old paradigm for managing tropical forests. BioScience 50 (6): 521 – 528.

Franklin D (1999) Teenage birth rates in California: what really matters? [http://www.csus.edu/indiv/w/wassmerr/Frankout.htm]. MS Thesis, Graduate Program in Public Policy and Administration, California State University, Sacramento.

Franklin S and Graesser A (1996) Is it agent, or just a program? A taxonomy for autonomous agents. In: Muller JP, Wooldridge MJ, and Jennings NR (eds) Proceedings of ECAI'96 Workshop (ATAL). Intelligent Agents III - Agent theories, architectures, and languages. Lecture Notes in Artificial Intelligence 1193, Springer-Verlag, pp. 21 - 35.

Freudenberger KS (1995) Tree and land tenure: using rapid rural appraisal to study natural resource management. Community Forestry Case Study Series. FAO, Rome, Italy.

Gardner RH (1998) Pattern, process, and the analysis of spatial scales. In: Peterson DL and Parker PT (eds) Ecological scale: theory and applications. Columbia University Press, New York, USA, pp. 17 - 34.

Geoghegan J, Prichard JRR, Ogneva-Himmelberger Y, Chowdhury RR, Sanderson S, and Turner II BL (1998) "Socializing the pixel" and "pixelizing the social". In: Liverman D, Moran EF, Rindfuss RR, and Stern P (eds) People and pixels: linking remote sensing and social science. National Academy Press, Washington D.C., USA, pp. 51 - 69.

Gessler PE, Chadwick OA, Chamran F, Althouse L, and Holmes K (2000) Modeling soil–landscape and ecosystem properties using terrain attributes. SSAJ 64: 2046 – 2056.

Gibson CC, Ostrom E, and Ahn TK (2000) The concept of scale and the human dimensions of global change: a survey. Ecological Economics 32: 217 – 239.

Gilbert N and Bankes S (2002) Platforms and methods for agent-based modelling. PNAS 99 (3): 7197–7198.

Gilbert N and Troitzsch KG (1990) Simulation for the Social Scientist. Open University Press, Buckingham, UK.

Ginot V, Le Page C, and Souissi S (2002) A multi-agents architecture to enhance end-user individual-based modelling. Ecological Modelling 157: 23 – 41.

Godet M and Roubelat F (1996) Creating the future: the use and misuse of scenarios. Long Range Planning 29 (2): 164 – 171.

Gomiero T and Giampietro M (2001) Multi-scale integrated analysis of farming systems: the Thuong Lo commune (Vietnamese uplands) case study. Population and Environment 22 (3): 315 – 352.

Gonzalez R (2000) Platforms and terraces: GIS as a tool for interactive land use negotiation. Ph.D. Thesis. Wageningen University and ITC, Wageningen and Enschede, The Netherlands.

Goodall B (1987) Dictionary of human geography. The Penguin, London, UK.

Goodchild M (2001) Issues in spatially explicit modeling. In: Paker DC, Berger T, and Manson SF (eds) Agent-based models of land-use and land-cover change. Proceedings of an International Workshop, October 4 - 7, 2001. Center for Study of Institutions, Population, and Environmental Change (CIPEC), Indiana University, USA, pp. 13 - 16.

Green DG (1993) An introduction to cellular automata [http://parallel.hpc.unsw.edu.au/complex/pubs.html]. Complexity On-line Tutorials, University of New South Wales.

Greene WH (1995) LIMDEP version 7.0: user's manual. Macmillan Publishing Company, New York, USA.

Gupta V (1999) SPSS for beginners. 1stBooks Library.

Hafeez MM (2003) Water accounting and productivities at different spatial scales in a rice irrigation system: a remote sensing approach. Ecology and Development Series No. 8, Cuvillier Verlag, Göttingen, Germany.

Haggith M, Muetzelfeldt RI, and Taylor J (2003) Modelling decision-making in rural communities at the forest margin. Small-scale Forest Economics, Management and Policy 2(2): 241 – 258.

Havel JJ (1980) Application of fundamental synecological knowledge to practical problems in forest management I: theory and methods. Forest Ecology and Management 3: 1 – 29.

Hiratsuka M, Toma T, Yamada M, Heriansyah I, and Morikawa Y (2003) A general algometric equation for estimating biomass in *Acacia mangium* plantations. In: International Conference on Tropical Forests and Climate Change. Conference proceedings. Manila, Philippines (in press).

Houghton RA, Hacker JL, and Lawrence KT (1999) The U.S. Carbon budget: contribution from land-use change. Science 285: 574 – 578.

Huigen M (2004) First principles of the Mamluke multi-actor modeling framework for land-use change, illustrated with a Philippines case study. Journal of Environmental Management 72: 5 – 21.

Huigen M (2003) Agent-based modelling in land use and land cover change studies. International Institute for Applied Systems Analysis (IIASA), Lanxenburg, Austria.

James YW and Cutforth HW (1996) Crop growth models for decision support systems. Can. J. Plant Sci. 76: 9 – 19.

Jamieson NL, Le Trong Cuc, and Rambo AT (1998) The development crisis in Vietnam's mountains. East-West Center (EWC), Honolulu, Hawai, USA.

Jansen H and Stoorvogel JJ (1998) Quantification of aggregation bias in regional agricultural land use models: application to Guacimo County, Costa Rica. Agricultural System 58: 417 – 439.

Johnsen K, Samuelson L, Teskey R, McNulty S, and Fox T (2001) Process models as tools in forestry research and management. Forest Science 47 (1): 2 – 8.

Jorgensen S, Patten B, and Straskraba M (2000) Ecosystems emerging: growth. Ecological Modelling 126: 249–284.

Kaelbling LP and Littman ML (1996) Reinforcement learning: a survey. Journal of Artificial Intelligence Research 4: 237 – 285.

Kaimowitz D and Angelsen D (1998) Economic models of tropical deforestation: a review. Centre for International Forestry Research (CIFOR), Bogor, Indonesia.

Kanaroglou P and Scott D (2002) Integrated urban transportation and land-use models for policy analysis. In: Dijst M, Schenkel W and Thomas I (eds) Governing cities on the move: functional and management perspectives on transformations of European urban infrastructure. Ashgate, Aldershot, UK, pp. 42 – 72.

Kemp-Benedict E (2004a) From narrative to number: A role for quantitative models in scenario analysis. In: Pahl-Wostl C, Schmidt S, Rizzoli AE and Jakeman AJ (eds) Complexity and integrated resources management. Transactions of the 2nd Biennial Meeting of the International Environmental Modelling and Software Society. iEMSs, Manno, Switzerland.

Kemp-Benedict E (2004b) Number in scenario studies. IPAT-S Scenario Notes 2. [http://ipat-s.kb-creative.net/Resources/]

Kennedy J, Eberhart R, Shi Y (2001) SWARM intelligence. The Morgan Kaufmann Series in Artificial Intelligence.

Kitamura T, Kagatsume M, Hoshino S, and Morita H (1997) A theoretical consideration on the land-use change model for the Japan case study area. International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria.

Koczberski G, Curry G, and Gibson K (2001) Improving productivity of the smallholder oil palm sector in Papua New Guinea: A socio-economic study of the Hoskins and Popondetta schemes. Australian National University, Canberra, Australia.

Kohler TA (2000) Putting social sciences together again: an introduction to the volume. In: Kohler TA and Gumerman GJ (eds) Dynamics in human and primate societies: agent-based modeling of social and spatial processes. Oxford University Press, New York, USA, pp. 1 - 18.

Korfmacher KS (2001) The politics of participation in watershed modeling. Environmental Management 27 (2): 161 – 176.

Lambin EF, Baulies X, Bockstael N, Fischer G, Krug T, Leemans R, Moran EF, Rindfuss R R, Sato Y, Skole D, Turner II BL, Vogel C, Nunes C, and Augé JI (1999) Land-use and land-cover change implementation strategy. IGBP Report No. 48, Stockholm, Sweden.

Lambin EF, Geist HJ, and Lepers E (2003) Dynamics of land- use and land-cover changes in tropical regions. Annu. Rev. Environ. Resour. 28: 205 – 241.

Lambin EF, Turner II BL, Geist HJ, Agbola SB, Angelsend A, Bruce JW, Coomes OT, Dirzo R, Fischer G, Folke C, George PS, Homewood K, Imbernon J, Leemans R, Li X, Moran EF, Mortimore M, Ramakrishnan PS, Richards JF, Skanes H, Steffen W, Stone GD, Svedin U, Veldkamp TA, Vogel C, and Xu J (2001) The causes of land-use and land-cover change: moving beyond the myths. Global Environmental Change 11: 261 – 269.

Le Sau and Tran Huy Thiep (1997) Technical guide of watershed protection zoning for projects under the 327 Program. Hanoi, Vietnam: Forest Inventory and Planning Institute of Vietnam (FIPI) [in Vietnamese].

Le Quang Bao (1998) Plants as potential indicators for soil contamination by heavy metals in Kup Kap Stream Valley, Mae Taeng District, Chiang Mai Province. M.Sc. Thesis. Chiang Mai, Thailand: Faculty of Science, Chiang Mai University, 122 pp.

Le Quang Bao (1999) Community-based natural resource management in the Thua Thien-Hue uplands. A briefing report for the period 1994 – 1998 prepared for IDRC's governors. Hue University of Agriculture and Forestry, Hue, Vietnam.

Le Quang Bao (2002) Forest management in Hong Ha commune and practical lessons learned. In: Le Van An (ed) Community-Based Natural Resource Management on the Uplands. The Agricultural Publish House, Hanoi, Vietnam, pp. 99 - 120 [in Vietnamese].

Le Trong Trai, Tran Hieu Minh, Tran Quang Ngoc, Tran Quoc Dung, and Hughes R (2001) An Investment Plan for the Establishment of Phong Dien Nature Reserve, Thua Thien Hue Province, Vietnam. Conservation Report No. 15. Birdlife International - European Union - Forest Institute for Inventory and Planning, Hanoi, Vietnam.

Le Van An (1999) General characteristics of Hong Ha commune, its potentials, problems and solutions in community development. In: Le Van An and Nguyen Thi My Van (eds) Community-based upland natural resource management project Hong Ha commune, Aluoi, Thua Thien-Hue, Vietnam. Annual report of Year 1, Phase II (1998 - 2001). Hue University of Agriculture and Forestry – CIAT – IDRC, Hue, Vietnam, pp. 15 – 21.

Leary RA (1985) Interaction theory in forest ecology and management. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Leigh EG (1999) Tropical forest ecology: a view from Barro Colorado Island. Oxford University Press, New York, USA.

Lenton TM and van Oijen M (2002). Gaia as a complex adaptive system. Phil. Trans. R. Soc. Lond. B 357: 683 – 695.

Levin SA (1992) The problem of pattern and scale in ecology. Ecology 73: 1943 - 1967

Ligtenberg A, Bregt AK, and Van Lammeren R (2001) Multi-actor-based land use modelling: spatial planning using agents. Landscape and Urban Planning 56: 21 - 33.

Ligtenberg A, Wachowicz M, Bregt AK, Beulen A, and Kittenis DL (2004) A design and application of a multi-agent system for simulation of multi-actor spatial planning. Journal of Environmental Management 72 (1, 2): 43 – 56.

Liu DS, Iverson LR, and Brown S (1993) Rates and patterns of deforestation in the Philippines: applications of geographic information system analysis. Forest Ecology and Management 57: 1 – 16.

Loibl W and Toetzer T (2003) Modeling growth and densification processes in suburban regions: simulation of landscape transition with spatial agents. Environmental Modelling & Software 18: 553 – 563.

Lomuscio AR, Wooldridge M, and Jennings NR (2003) A classification scheme for negotiation in electronic commerce. International Journal of Group Decision and Negotiation 12 (1): 31 – 56.

Maack, JN (2001) Scenario analysis: a tool for task managers. In: Krueger R, Cassey M, Donner J, Kirch S, and Maack JN (eds) Social analysis: selected tools and techniques. Social Development Paper 36. Social Development Department, World Bank, Washington D.C., USA, pp. 62 – 87.

Maddala GS (1992) Introduction to econometrics. Prentice-Hall, New Jersey, USA.

Mäkelä A, Landsberg JJ, Ek AE, Burk TE, Ter-Mikaelian M, Ågren G, Oliver CD, and Puttonen P (2000) Process-based models for forest ecosystem management: current state-of-art and challenges for practical implementation. Tree Physiology 20: 289 – 298.

Malczewski J (1999) Geographic Information System (GIS) and multi-criteria decision analysis. John Wiley & Sons, New York, USA.

Manson SM (2001) Simplifying complexity: a review of complexity theory. Geoforum 32 (3): 405 – 414.

Martens P and Rotmans J (2002) Transitions in a globalising world. Swets & Zeitlinger, The Netherlands.

Mather R, de Boer M, Gurung G, and Roche N (1998) Aerial photographs and "photo-maps" for community forestry (Paper 28e). Rural Development Forestry Network, London, UK, pp. 13 – 22.

Matthews R (2002) Crop management. In: Matthews R and Stephens W (eds) Crop-soil simulation models. CAB International, Wallingford, UK, pp. 69 - 79.

Meadows S H (2000) The power of feedback. Energy World 282: 21.

Minar N, Burkhart R, Langton C, and Askenazi M (1996) The SWARM simulation system: a toolkit for building multi-agent simulations. [http://www.santafe.ed/project/swarm]

Moore I D, Gessler PE, and Nielson G A (1993) Soil attribute prediction using terrain analysis. SSSAJ 57: 443 – 452.

Moore ID and Wilson JP (1992) Length-slope factors for the revised universal soil loss equation: simplified method of estimation. Journal of Soil and Water Conservation 475: 423 – 428.

Morrison E and Dubois O (1998) Sustainable livelihoods in Vietnam uplands: land allocation and beyond. Forestry and Land Use Series 14. International Institute for Environment and Development (IIED), London, UK.

Müller D (2003) Land-use change in the Central Highlands of Vietnam: a spatial econometric model combining satellite imagery and village survey data. Ph.D. Dissertation. Faculty of Agriculture, Göttingen University, Göttingen, Germany.

Nelson A (2000) Accessibility, transport and travel time information. CIAT, Cali, Colombia.

Nelson G, De Pinto A, Harris V, and Stone S (2004) Land use and road improvement: a spatial perspective. International Regional Science Review 27 (3): 297 – 325.

Nguyen BT (1998) Natural resources policies in the highlands of Vietnam: a synthesis of NGO experience. World Resource Institute (WRI), Washington D.C., USA.

Nguyen Ngoc Binh (1996) Forest soils in Vietnam. The Agricultural Publishing House, Hanoi, Vietnam.

Nguyen Xuan Hong (2002) Cultural characteristics of the KaTu, Taoi (Paco, Pahy), Bru-Van Kieu people and their impacts to the development of the economy, natural resources and environment in Hong Ha commune. In: Le Van An (ed) Community-based natural resource management on the Thua Thien-Hue uplands. The Agricultural Publish House, Hanoi, Vietnam, pp. 61 - 69 [in Vietnamese].

O'Brien P (2001) Scenario planning: a strategic tool. Bureau of Rural Sciences, Canberra, Australia.

O'Neil RV, Johnson AR, and King AW (1989) A hierarchical framework for the analysis of scale. Landscape Ecology 3: 193 – 205.

Orland B, Budthimedhee K, and Uusitalo J (2001) Considering virtual worlds as representations of landscape realities and as tools for landscape planning. Landscape and Urban Planning 54: 139 – 148.

Page EH (1994) Simulation modeling methodology: principles and etiology of decision support. Ph.D. Dissertation. Virginia Polytechnic Institute and State University

Pallaris K (2000) Terrain modelling for soil erosion risk assessment in the Cabuyal River catchment: comparison of results with farmer perceptions. Advances in Environmental Monitoring and Modelling 1 (1): 149 – 177.

Park SJ and Van de Giesen N (2004) Delineation of soil-landscape units along hillslopes to identify the spatial domains of hydrological processes. Jounral of Hydrology 295: 28-46.

Park SJ and Vlek PLG (2002) Environmental correlation of three-dimensional soil spatial variability: a comparison of three adaptive techniques. Geoderma 109: 117 – 140.

Park SJ and Vlek PLG (2003) Comparison of adaptive techniques to predict crop yield response under varying soil and land management conditions. Submitted to Agricultural Systems.

Park SJ, McSweeney K, and Lowery B (2001) Prediction of soils using a process based terrain characterization. Geoderma 103: 249 – 272.

Park SJ, Van de Giessen N, and Vlek PLG (2003) Optimum spatial scale for modeling land use change processes in a Savanna landscape in Northern Ghana. Agriculture, Ecosystems and Environment (accepted).

Parker DC, Manson SM, and Berger T (2001) Potential strengths and appropriate roles for ABM/LUCC. In: Parker DC, Berger T, and Manson SM (eds) Agent-based models of land-use and land-cover change. LUCC Report Series No. 6, Louvain-la-Neuve, pp. 6-12.

Parker DC, Manson SM, Janssen MA, Hoffmann MJ, and Deadman P (2002) Multi-agent system for the simulation of land-use and land-cover change: a review. Annals of the Association of American Geographers 93: 314 – 337.

Pegden CD, Shannon RE and Sadowski RP (1995) Introduction to simulation using SIMAN, (2nd edition). McGraw-Hill.

Penning de Vries FWT, Jansen DM, Berge HFM and Bakema A (1989) Simulation of ecophysiological processes of growth in several annual crops. Simulation Monographs. Pudoc/Internation Rice Research Institute, Wageningen, the Netherlands.

Peter RH (1991) A critique for ecology. Cambridge University Press, Cambridge, UK.

Pichon FJ (1997) Settler households and land-use patterns in the Amazon frontier: farm-level evidence from Ecuador. World Development 25: 67 – 91.

Polhill G, Gotts N, and Law A (2001) Imitative versus non-imitative strategies in a land use simulation. Cybernet. Syst. 32, 285 – 307.

Quang Bao Le (2003) A questionnaire for interviewing households in land-use and land cover change study. Memeo. 19 pp.

Quintero R, Baros, R, Dávila J, Monero, N, Tonella G, Ablan M (2004) A model of the bio-complexity of deforestation in tropical forest: Caparo case study. In: Pahl-Wostl C, Schmidt S, Rizzoli AE and Jakeman AJ (eds) Complexity and Integrated Resources Management. The 2nd Biennial Meeting of the International Environmental Modelling and Software Society. iEMSs, Manno, Switzerland.

Ramankutty N and Foley JA (1999) Estimating historical changes in global land cover: cropland from 1700 to 1992. Global Biogeochemical Cycles 13: 997 - 1027.

Rambaldi G and Callosa J (2000) Manual on participatory 3-D modeling for natural resource management. Essentials of protected area management 7. National Integrated Protected Areas Programme. European Commission and DENR, Quezon City, Philippines.

Rambaldi G, Fernan MS, and Siar SV (1998) Participatory resource mapping: a tool for community-based coastal resource management. Journal of Agriculture and Environment for International Development 92 (1): 69 – 79.

Rambo AT (1995) Perspectives on defining highland development challenges in Vietnam: new frontier or cul-de-sac? In: Rambo AT, Le Trong Cuc, and Digregorio MR (eds) The challenges of highland development in Vietnam. Eastwest Center (EWC), Honolulu, Hawai, USA.

Raskin P, Swart RJ, and Robinson J (2004). Navigating the sustainability transition: the future of scenarios. In: Biermann F, Campe S, and Jacob K (eds) Knowledge for the sustainability transition: the challenge for social science. Conference on the human dimensions of global environmental change, Berlin, 2002. Global Governance Project, Berlin, Germany, pp. 53 – 66.

Ratkowsky DA (1990) Handbook of nonlinear regression models. Marcel Dekker, New York, USA.

Resnick M (1994) Turtles, termites and traffic jams: explorations in massively parallel micro worlds. MIT Press, Massachusetts, USA.

Resnick M and Wilensky U (1993) Beyond the deterministic, centralized mindsets: new thinking for new sciences. American Educational Research Association.

Retherford RD and Choe MK (1993). Statistical models for causal analysis. John Wiley & Sons, New York, USA.

Reynolds JF, Stafford-Smith DM, and Lambin EF (2003) Do human cause desserts? An old problem through the lens of a new framework: The Dahlem dessertification paradigm. In: Allsopp N, Palmer AR, Milton SJ, Kirman KP, Kerley GIH, Hurt CR, and Brown CJ (eds) Proceedings of the VIIth International Rangelands Congress. Durban, South Africa, pp. 2042 – 2048.

Richards FJ (1959) A flexible growth function for empirical use. Journal of Experimental Botany 10 (29): 290 - 300.

Robinson R (2003) Simulation: the practice of model development and use. Willey Publishers.

Rodriguez JP (2003) Transportation geography on the web. [http://people.hofstra.edu/geotrans/index.html]. Dept. of Economics & Geography, Hofstra University.

Rouchier J, Bousquet F, Requier-Desjardins M, and Antona M (2001) A multi-agent model for describing transhumance in North Cameroon: Comparison of deferent rationality to develop a routine. Journal of Economic Dynamics and Control 25: 527 – 559.

Russell SJ and Norvig P (1995) Artificial intelligence: a modern approach. Prentice and Hall, Englewood Cliffs, New Jersey, USA.

Sadoulet D, Castella JC, Vu Hai Nam, and Dang Dinh Quang (2002) A short history of land-use changes and farming system differenciation in Xuat Hoa Commune, Bac Kan Province, Vietnam. In: Castella JC and Dang Dinh Quang (eds) Doi Moi in the mountains: land use changes and farmers' livelihood strategies in Bac Kan Province, Vietnam. The Agricultural Publishing House, Hanoi, Vietnam, pp. 21 - 46.

Sala OE, Chapin FS, Armesto JJ, Berlow E, Bloomfield J, Dirzo R, Huber-Sanwald E, Hueneke LF, Jackson RB, Kinzig A, Leemans R, Lodge DM, Mooney HA, Oesterheld M, Poff NL, Sykes MT, Walker BH, Walker M, and Wall DH (2000) Biodiversity: global biodiversity scenarios for the year 2100. Science 287: 1770 -1774.

Sanchez S and Lucas T (2002) Exploring the world of agent-based simulations: simple models, complex analyses. In: Yücesan E, Chen CH, Snowdon JL, and Charnes JM (eds) Proceedings of the 2002 Winter Simulation Conference, New Jersey. Institute of Electrical and Electronics Engineers, New Jersey, USA, pp. 116 - 126.

Sawhney A, Walsh K, and Mulky AR (2003) Agent-based modeling and simulation in construction. In: Chick S, Sánchez PJ, Ferrin D, and Morrice DJ (eds) Proceedings of the 2003 Winter Simulation Conference, New Jersey. Institute of Electrical and Electronics Engineers, New Jersey, USA , pp. 1541 - 1547.

Schoeneberger PJ, Wysocki DA, Benham EC, and Broderson WD (1998) Field book for describing and sampling soils. Nation Soil Survey Center. National Resources Conservation Service, USDA, Lincoln, NE, USA.

Shannon RE (1998). Introduction to the art and science of simulation. In: Medeiros DJ, Watson EF, Carson JS, and Manivannan MS (eds) Simulation in the 21st Century. Proceedings of the 1998 Winter Simulation Conference, New Jersey. Institute of Electrical and Electronics Engineers, Inc., New Jersey, USA, pp. 7 – 14.

Shugart HH and Urban DL (1988) Scale, systhesis, and ecosystem dynamics. In: Pomeroy LR and Alberts JJ (eds) Concepts in ecosystems ecology. Springer-Verlag, New York, USA, pp. 279 - 289.

Sinclair TR and Seligman NAG (1996) Crop modeling: from infancy to maturity. Agronomy Journal 88: 698 – 703.

Sit V and Poulin-Costello M (1994) Catalogue of curves for curve fitting. Biometrics Information Handbook No. 4. Province of British Columbia, Ministry of Forests, Victoria, Canada.

Sklar FH and Costanza R (1991) The development of dynamic spatial models for landscape ecology: A review and prognosis. In: Turner MG and Gardner RH (eds) Quantitative methods in landscape ecology. Springer-Verlag, New York, USA, pp. 239 – 288.

Smith K, Barrett CB, and Box PW (1999) Participatory risk mapping for targeting research and assistance: with an example from East African pastoralists. Social Sciencience Research Network Electronic Library. [http://ssrn.com/abstract=185369].

Stafford-Smith DM., and Reynolds JF (2002) Dessertification: a new paradigm for old problem. In: Reynolds JF and Stafford-Smith DM (eds) Global dessertification: Do human cause desserts? Dahlem Workshop Report 88. Dahlem University Press, Berlin, Germany, pp. 403 – 424.

Stave KA (2002) Using system dynamics to improve public participation in environmental decisions. System Dynamics Review 18 (2): 139 – 167.

Stephens W and T Middleton (2002) Why has the uptake of decision support systems been so poor? In: Matthews R and Stephens W (eds) Crop-soil simulation models. CAB International, Wallingford, UK, pp.129 – 147.

Sterman JD (1994) Learning in and about complex systems. System Dynamics Review 10 (2, 3): 291 – 330.

Sterman JD (2002) All models are wrong: reflections on becoming a system scientist. Jay Wright Forrester Prize Lecture 2002. System Dynamics Review 18 (4): 501 - 531.

Tan Quang Nguyen (2003) Differentiation in benefits from forest devolution: empirical study from Dak Lak province, Central Highland of Vietnam. Paper represented at Deutsche Tropentag 2003. Goettingen, Germany.

Tissue S and Wilensky U (2004) NetLogo: design and implementation of a multi-agent modeling environment. Paper presented at SwarmFest, Ann Arbor, May 9–11, 2004.

Tomobe K (2002) Agricultural production function, labor productivity, and wage embedded in Meiji peasant productive behaviors: a quantitative analysis of Nohi Ryogoku Choson Ryakushi, ca.1881. Paper prepared for the XIIth Congress of the International Economic History Association, Buenos Aires, 22 - 26 July 2002.

Tran Duc Vien, Rambo AT, Dzung PT, Lam NT, Yen NTB, Ha NTT, An NT, Huong PT, and Van DTH (2001) Household and community economics. In: Le Trong Cuc and Rambo AT (eds). Northern Vietnam Uplands: environmental and socio-economic issues. National Politics Publishers, Hanoi, Vietnam, pp. 117 - 167 [in Vietnamese].

Turner BL, Skole DL, Sanderson S, Fisher G, Fresco LO, and Leemans R (1995) Land-use and land cover change - Science/Research Plan. IGBP Report No. 35; HDP Repot No. 7, Stockholm and Geneva.

Turner MG (1999) Merging local and regional analyses of land-use change: the case of livestock in the Sahel. Annals of the Association of American Geographers 89: 191 – 219.

Van der Heijden K (1998) Scenarios – the art of strategic conversion. John Willey & Sons, New York, USA.

Van der Veen A and Otter HS (2001) Land use changes in regional economic theory. Environmental Modeling and Assessment 6: 145 – 150.

Van Dyke Parunak H (1999) Characterizing multi-agent negotiation. Center for Electronic Commerce ERIM, Michigan, USA.

Van Dyke Parunak H, Savit R, and Riolo RL (1998) Agent-based modeling vs. equation-based modeling: a case study and users' guide. Lecture Notes of Artificial Intelligence 1534: 10 – 25.

Van Noordwijk M, Verbist B, Vincent G, and Tomich TP (2001) Simulation models that help to understand local action and its consequences for global concerns in a forest margin landscape ASB. Lecture Note 11A. International Center for Research Agroforestry (ICRAF), Bogor, Indonesia.

Vanclay JK (1994) Modelling forest growth and yield: applications to mixed tropical forests. CAB International, Wallingford, UK.

Vanclay JK (1997) FLORES: a model to evaluate land use options at the forest frontier. In: Advances in Methodology and Software for Decision Support Systems. IIASA, Laxenburg, Austria.

Veldkamp A and Lambin EF (2001) Predicting land use change: editorial. Agriculture, Ecosystems and Environment 85: 1 – 6.

Veldkamp A and Verburg PH (2004) Modelling land use change and environmental impact: editorial. Journal of Environmental Management 72: 1 – 3.

Verburg P, Schot P, Dijst M, and Veldkamp A (2002) Land use change modelling: current practice and research priorities. GeoJournal: in press, 24 pp.

Verburg PH, de Groot WT, and Veldkamp AJ (2003) Methodology for multi-scale land use change modeling: concepts and challenges. In: Dolman AJ and Verhagen A (eds) Global environmental change and land use. Kluver Academic Publishers, Dordrecht, the Netherlands, pp. 17 - 51.

Victor DG and Ausubel JH (2000) Restoring the forests. Foreign Aff. 79 (6): 127 – 144.

Villa F (1992) New computer architecture as tools for ecological thought. Trends in Ecology and Evolution 7: 179 – 183.

Vitousek PM, Mooney, HA, Lubchenko J, and Mellio JM (1997) Human domination of Earth's ecosystems. Science 277: 494 - 499.

Vlek PLG, Berger T, Iskandarani M, Park SJ, and Van de Giessen N (2003) Integrated watershed project: GLOWA-Volta. In: Ehlers (ed) Earth system science in the antropence: emerging issues and problems. Springer Verlag (in press).

Vo Quy (1996) The environmental challenges of Vietnam's development. In: Draft report. Regional Seminar on Environmental Education on 19 – 22 March 1996. Center for Natural Resources and Environmental Studies (CRES), University of Hanoi, Hanoi, Vietnam, pp. 4 - 20.

Vo Quy and Le Thac Can (1994) Conservation of forest resources and the greater biodiversity of Vietnam. Asian Journal of Environmental Management 2(2): 55 – 59.

Von Bertalanffy L (1942) Theoretische biologie [Theoretical biology]. Berlin.

Von Bertalanffy L (1957) Quantitative laws in metabolism and growth. Quart. Rev. Biol. 32: 217 – 231.

Von Bertalanffy L (1968) General system theory: foundation, development, applications. George Braziller, New York, USA.

Vu Dinh Huong, Tung PV, Dung PT, Phuc HV, Binh NT, Duc HM, and Tron NT (2004) Site management and productivityof *Acacia auriculiformis* plantations in South Vietnam. In: Nambiar EKS, Ranger J, Tiarks A, and Toma T (eds) Site management and productivity in tropical plantation forests. Proceedings of Workshops in Congo, July 2001 and China, February 2003. CIFOR, Jakarta, Indonesia, pp. 121 – 137.

Vu Van Dzung (2002) Nature conservation and sustainable forest management in Quang Tri and Thua Thien – Hue provinces: a final report of desk study. TBI-Vietnam program, Hanoi, Vietnam.

Waldrop MM (1992) Complexity: the emerging science at the edge of order and chaos. Simon and Schuster, New York, USA.

Wallenius H (2004) Managerial economics: lecture notes spring 2004. [http://www.tuta.hut.fi/studies/Courses_and_schedules/Isib/TU- 91.113/ lecturenotes/]. Helsinki University of Technology.

Watkins C and Dayan P (1992) Q-learning. Machine Learning 8 (3/4): 279 – 292.

WCMC (World Conservation Monitoring Centre) (1996) The Socialist Republic of Vietnam [http://www.wcmc.org.uk/infoserv/countryp/vietnam/index.html]. WCMC, Cambridge, UK.

Weisbuch G, Kirman A, and Herreiner F (1997) Market organization. Lecture Notes Econ. Math. Syst. 456.

Western AW, Grayson RB, Blöschl G, Willgoose GR, and McMahon TA (1999) Observed spatial organization of soil moisture and its relation to terrain indices. Water Resources Research 35 (3): 797 – 810.

Westervelt JD and LD Hopkins (1999) Modeling mobile individual in dynamic landscapes. International Journal of Geographical Information Science 13 (3): 191 – 208.

Wikramanayake E, Dinerstein E, Hedao P, and Olson D (1997) A conservation assessment of terrestrial eco-regions of the Indo-Pacific Region. WWF-US Conservation Science Program, Washington D.C., USA.

Wilensky U (1998) NetLogo Wealth Distribution model [http://ccl.northwestern.edu /netlogo/] Center for Connected Learning and Computer-Based Modeling (CCL), Northwestern University, Evanston, IL, USA.

Wilensky U (1999) NetLogo [http://ccl.northwestern.edu/netlogo/]. Center for Connected Learning and Computer-Based Modeling (CCL), Northwestern University, Evanston, IL, USA.

Wilson JP and Gallant JC (2000) Digital terrain analysis. In: Wilson JP and Gallant JC (eds) Terrain analysis: principles and applications. John Wiley & Sons, New York, USA, pp. 1 - 28.

Wollenberg E, Edmunds D, and Buck L (2000) Anticipating change: scenarios as a tool for adaptive forest management: a guide. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Wooldridge M (2002) An introduction to multi-agent systems. John Wiley & Sons, New York, USA.

Wu F (1998) An experiment on the generic polycentricity of urban growth in a cellular automatic city. Environment and Planning B: Planning and Design 25: 731 - 752.

Wu F and Yeh AGO (1997) Changing spatial distribution and determinants of land development in Chinese cities in the transition from a centrally planned economy to a socialist market economy: a case study of Guangzhou. Urban Studies 34 (11): 1851 – 1879.

Zambonelli F, Jennings N R, and Wooldridge M (2003) Developing multi-agent systems: the Gaia methodology. ACM Transactions on Software Engineering and Methodology.

Zambonelli F, Jennings NR, and Wooldridge M (2001) Organisational Abstractions for the Analysis and Design of Multi-agent Systems. In: Ciancarini P and Wooldridge M (eds) AOSE 2000, LNCS 1957. Springer-Verlag, Berlin, Germany, pp. 235 – 251.

Zingerli C, Castella JC, Manh PH, and Cu PV (2002) Contesting policies: rural development versus biodiversity conservation in the Ba Be National Park area, Vietnam. In: Castella JC and Dang Dinh Quang (eds) Doi Moi in the mountains: land use changes and farmers' livelihood strategies in Bac Kan Province, Vietnam. The Agricultural Publishing House, Hanoi, Vietnam, pp. 249 - 276.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my great gratitude and appreciation to Professor Paul L.G. Vlek, my principal supervisor, for his excellent supervision and constant encouragement from the beginning to the end of this doctoral research. Especially, I am very much grateful for his clear and constructive guidance on how to approach the problem of development research, and the need to be persistent and realistic to accomplish the goal of this modeling research. Likewise, I am deeply indebted to Asst. Prof. Dr. Soojin Park, my scientific adviser, for not only his methodological guidance, but also his dedicating and friendly tutorials that brought about my scientific enthusiasm to pursue this research. I am grateful to Prof. Dr. Eckart Ehlers, PD. Dr. Christopher Martius, Prof. Dr. Wulf Amelung, and Prof. Dr. Armin B. Cremers for their willingness to be members of the Examination Committee, as well as their valuable comments and suggestions on the dissertation.

My sincere thanks to Church Development Service (EED), viz. the Scholarship Desk for giving the financial support for my stay in Germany and in Vietnam during my fieldwork. I would like to express my deep regards to Dr. Rudolf Ficker and Mrs. Susanne Werner at the Scholarship Desk of EED for their kind encouragement during my doctoral study.

I am very grateful to the Center for Development Research (ZEF) for offering me an excellent study environment that efficiently supports my interdisciplinary learning and research. I am grateful to Dr. Eric Craswell for fruitful discussions on my research topic and his helpful comments on the first chapter of this thesis. I am indebted to Dr. Stefanie Engel for her suggestions on the socio-economic part of my research proposal and my household surveys. My appreciation is given to Dr. Mathias Braun for his tutorial on remote sensing analyses. I extend my sincere thanks to Dr. Gunther Manske, Mrs. Hanna Peters, Mrs. Rosemarie Zabel, Mrs. Sabine Aengenendt-Baer, Mrs. Andrea Berg, and Ms Miryam Schneider for the excellent cooperation and support during my study at ZEF. My appreciation also goes to Mrs. Margaret Jend for editing the English of this dissertation.

My sincere thanks are extended to Dr. Peter Kerridge, Dr. Peter Taylor, Professor Vo Hung and Professor Tran Van Minh for giving me their kind recommendations of my doctoral study at ZEF Bonn.

My great appreciation is given to academic members of the Faculty of Forestry at Hue University of Agriculture and Forestry, especially Mr. Nguyen Thanh Thu and Dr. Duong Viet Tinh, for their valuable support and encouragement during my doctoral study.

My sincere thanks are extended to farmers in Hong Ha commune, especially Mr. Le Hoai Nam and Mr. Le Van Hua, for their hospitality, providing living facilities and nice cooperation during my fieldwork. I am very grateful to Mr. Thai Minh Bao, Mr. Ngo Huu Toan, Mr. Ho Ngoc Gioi, Mr. Lai Ngoc Vien (and his forest inventory team) for their invaluable helps in the fieldwork. Mr. Nguyen Phuoc Buu Nguyen and staffs at Hong Ha Forest Protection Station are truly acknowledged for their kind collaborations.

I express my sincere thanks to Dr. Le Van An, Dr. Le Duc Ngoan, Dr. Peter Kerridge, Dr. John Graham, Dr. Sam Fujisaka and other research team members of the Community-based Upland Natural Resource Management Research Project in Hue University of Agriculture and Forestry for sharing research experiences, as well as collaborations during the field work.

I am thankful to my colleagues at ZEF for the friendship and collaborations during my doctoral study. Mr. Lulseged Tamene, Mr. Mekuria Argaw (Ethiopia), Mrs. Nguyen Thi Phuong Hoa, Mr. To Xuan Phuc (Vietnam) and Mr. Ho Dac Thai Hoang (at Freiburg University) are specially thanked for sharing experiences and nice friendship.

Finally, I give my heartful thanks to my parents, Mr. Le Quang Liem and Mrs. Le Thi Thanh Minh, my sisters, Mrs. Le Thi Thanh Chi and Mrs. Le Thi Thanh Quy, and my brother, Mr. Le Quang Tho, for their unending love, constant encouragement and moral support.