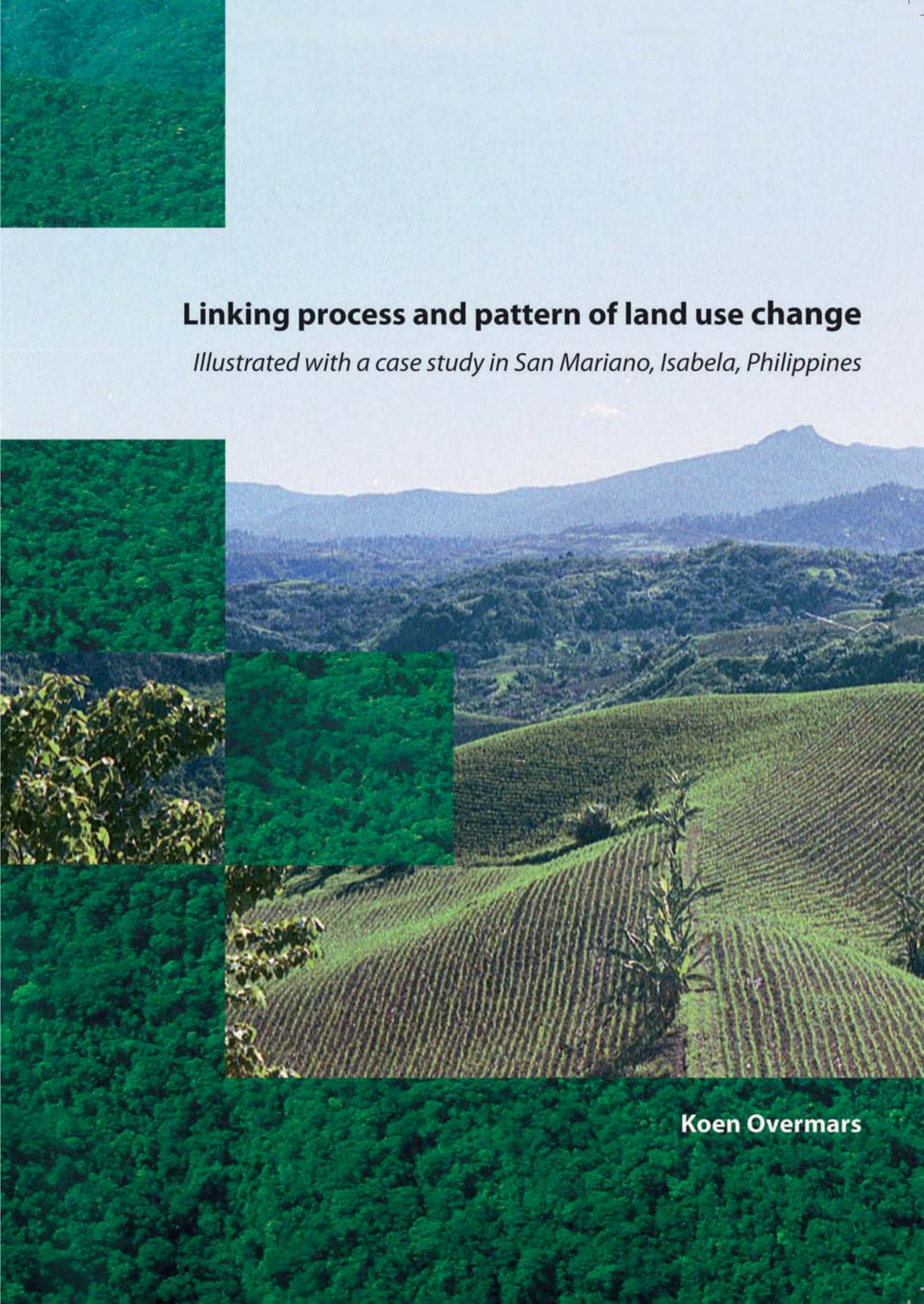
Linking processes and pattern of land use change

Overmars, K.P.

Linking process and pattern of land use change

Illustrated with a case study in San Mariano, Isabela, Philippines



Koen Overmars

Linking process and pattern of land use change

Illustrated with a case study in San Mariano, Isabela, Philippines

ISBN-10: 90-9020463-6
ISBN-13: 978-90-9020463-5

Cover and lay out: Stijn Koole & Koen Overmars
Photos: Koen Overmars & Tessa Minter

Linking process and pattern of land use change

Illustrated with a case study in San Mariano, Isabela, Philippines

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Dr. D.D. Breimer,
hoogleraar in de faculteit der Wiskunde en
Natuurwetenschappen en die der Geneeskunde,
volgens besluit van het College voor Promoties
te verdedigen op maandag 19 juni 2006
klokke 15.15 uur

door

Koen Pieter Overmars

geboren te Noordoostpolder
in 1976

Promotiecommissie:

Promotoren:	Prof. Dr. ir. W.T. de Groot Prof. Dr. ir. A. Veldkamp (Wageningen Universiteit)
Co-promotor:	Dr. ir. P.H. Verburg (Wageningen Universiteit)
Referent:	Prof. Dr. R.J. Aspinall (Arizona State University)
Overige leden:	Prof. Dr. J.N. Kok Prof. Dr. E. van der Meijden Prof. Dr. H.A. Udo de Haes

The research reported in this thesis was funded by the Foundation for the Advancement of Tropical Research (WOTRO) of the Netherlands Organisation for Scientific Research (NWO).

Preface

The research reported in this thesis was conducted as one of three projects that together formed an integrated program called "Integrating macro-modelling and actor-oriented research in studying the dynamics of land use change in northeast Luzon, Philippines", which was funded by the Foundation for the Advancement of Tropical Research (WOTRO) of the Netherlands Organisation for Scientific Research (NWO). In this program three other researchers conducted their studies: Marco Huigen, Peter Verburg and Cecile Mangabat. Marco was based at the Institute of Environmental Sciences Leiden (CML) of Leiden University conducting a PhD study on agent-based modelling of land use change at the micro-level. Peter was working as a post-doc at the Department of Environmental Sciences at Wageningen University on spatially explicit modelling of land use change in the Philippines as a whole. The aim of my research was to combine and to link these two spatial scales at the intermediate level and to link micro-level processes to observed patterns of land use at the landscape level. This work was carried out in both Wageningen and Leiden. Furthermore, Cecile was appointed as a local counterpart to study the impact of all forest related policies in the study area. The program was supervised by Wouter de Groot and Tom Veldkamp. With this set up we had a multidisciplinary research team to study the interdisciplinary research questions of land use change.

At the start of the project in May 2001 I lived in Leiden and spent most of my time at the CML to become familiar with the various sociological and anthropological methods and ideas in environmental studies, in which the CML has its expertise. In this period I got to know a completely different aspect of land use science and although I thought I had a quite interdisciplinary mindset I got many new insights and ideas about the functioning of the world of science in general and land use science in particular. After a first fieldwork period in the Philippines in 2002, I moved to Utrecht. From that time on, I spent half my time at the CML and the other half at the Laboratory for Soil Science and Geology in Wageningen. This way I was exposed to the different disciplinary inputs of both institutions on a weekly basis. Although working at two places brings about some organisational difficulties I have always enjoyed working in both institutions. For the interdisciplinary aspect it has been a great benefit to participate in both research groups and I learned a lot from balancing between the disciplines that are represented by these institutes.

In the first part of 2004 a second fieldwork was conducted in which I tried to link the land use research with biodiversity research and nature conservation. After that trip I worked towards finishing the PhD study. In October 2005 I was ready to send a final draft to my supervisors.

Acknowledgements

This dissertation is the result of four years of research and would not have been possible without the help, cooperation and support of many people. Writing these acknowledgements my thoughts go back to all the good things that have happened these past four years and I like to mention some of the experiences and people that were involved in this.

First of all, I would like to acknowledge the support of all my colleagues at both the Institute of Environmental Sciences (CML) in Leiden and the Laboratory of Soil Science and Geology in Wageningen. Although I spent only half my time in both institutions, I feel at home in both.

I would like to name a few people with whom I have worked most closely. Firstly, I like to thank Marco Huigen with whom I shared a room at the CML during these years, with whom I spent time in the Philippines and at many nice trips to conferences and workshops. Marco, maybe our characters are a little different, but finally we managed to work ourselves through this project. We have had tough discussions, but in the end I think it is precisely these tough discussions that most deeply influenced my understanding of what interdisciplinarity means. In Wageningen I shared a room with Peter Verburg and Kasper Kok. Peter, your support has been crucial for my research. I have learned a lot from you regarding land use science and I hope our collaboration will continue on projects in the future. Kasper, you always handled my moods elegantly and were always a good discussion partner. Cecile, thank you for your contributions to the project by giving us insight in the way forest polices work in the Philippines. Merlijn, I would like to thank you for the inspiring collaboration for the research and fieldwork on biodiversity and land use that resulted in the paper on which Chapter 6 is based.

During the four years of research that was spent on creating this dissertation nearly a year was spent in the Philippines. I gratefully acknowledge the support and commitment of CVPED in Cabagan, and especially Andy Masipiqueña and Jan van der Ploeg, who were the coordinators. Special thanks go to Noel Perez, who has been my field assistant during the fieldwork in the Philippines. Noel, without your help I could not have carried out the fieldwork. I very much enjoyed working with you and my thoughts go back to the many nice hikes we have had in the field.

I would especially like to acknowledge all government officials of the Municipality of San Mariano. Your cooperation has been essential to the success of my work. In the field I was helped by many people that offered me a place to stay for the night, cooperated with the research or simply enjoyed a merienda with us. In that respect I especially like to thank Jose and Patricia Wanol. I will never forget your hospitality to me and the students on all the occasions that we visited your home.

I would like to thank the students that dedicated part of their MSc studies to do field work in the Philippines within the framework of this project. Fenny, Nol, Marijn, Sander and Wouter, I had a nice time working with you. Finally, I like to thank everybody that made the periods of fieldwork in the Philippines a very nice, pleasant and memorable time.

The ISIS (Incitation à l'utilisation Scientifique des Images SPOT) program is acknowledged for providing SPOT images that are used for the construction of the land use map for the area in Chapters 5 and 6.

Finally, I am grateful for the support of my parents, family and friends. You made me realise that writing a dissertation is something special. On the other hand you were there to put things in perspective when I needed it.

Table of contents

1

General introduction

1.1 Relevance

The conversion of the earth's land surface by human actions has been extensive in the past and is still on going at a substantial rate (Vitousek *et al*., 1997). Although land use change is not the only component of global environmental change it has major impacts on climate change, ecosystem services, and sustainability (e.g. Moran *et al*., 2004; Rindfuss *et al*., 2004). Land use and land cover changes can induce climate changes directly through changes in albedo and transpiration rates. Land use influences climate indirectly through emissions of greenhouse gases from, for example, vegetation and soils (carbon dioxide) and rice paddies (methane) (e.g. Dale, 1997). Habitat destruction due to land use changes, for example tropical deforestation, forms an important threat to biodiversity (Tilman *et al*., 1994; Turner, 1996; Myers *et al*., 2000). Land use change can trigger soil degradation and soil erosion, which changes watershed properties and may cause flooding at local scales (Chomitz and Kamari, 1998; Bruijnzeel, 2004). Furthermore, unsustainable land use practices can affect soil properties causing loss of agricultural productivity with associated effects for local livelihoods and food security.

Land use change does not affect all regions in the world in a similar way. Some areas experience large changes with a high impact where other areas are hardly affected. One of the countries that is highly affected by land use change are the Philippines. In the past century, this country experienced large-scale deforestation (Kummer, 1992; ESSC, 1999), which was caused by intensive commercial logging and agricultural expansion. A large part of this agricultural expansion occurred in the upland areas (Garrity *et al*., 1993). When cultivating these uplands with arable crops like corn but without soil conservation measures the soils can easily erode (Coxhead and Buenavista, 2001).

The land use changes in the Philippines have major consequences for the landscape and the functions it can provide. The combination of severe loss of natural habitat and high numbers of endemic species makes the Philippines one of the most important conservation hotspots for biodiversity in the world (Myers *et al.*, 2000). The Philippines has the highest number (126) of endangered endemic species in the world (Brooks *et al.*, 2002). Fifty-three percent (92 species) of Philippine endemic forest bird species is threatened or near-threatened, mainly as a result of deforestation (IUCN, 2005). The catastrophic effects of land slides and flash floods after heavy rainfall, for example in eastern Luzon in December 2004, can mainly be attributed to on-going logging activities in the uplands, which destabilises slopes. Furthermore, many Philippine farmers have adopted unsustainable land use practices, especially cultivation of annual crops in upland areas, which leads to land degradation and restricts future opportunities for sustainable livelihoods (Coxhead and Buenavista, 2001).

These land use changes and their effects also apply to the study area of this dissertation, which is part of the municipality of San Mariano in the northeastern part of the Philippines (Figure 1.1). The area is situated in the transition zone between the lowlands of the Cagayan valley and the uplands of the Sierra Madre mountain range. The area experienced a high rate of deforestation, especially between the 1970s and the early 1990s. This is illustrated in Figure 1.1, which shows the forest cover in the study area in 1972 and 2001. Calculating the deforestation rate based on these maps shows a decrease of dense forest of 600 ha/yr in an area of 48,000 ha. One third of this dense forest changed into cleared area, which includes arable agriculture as well as grasslands, and two-thirds into 'low density forest', including logged-over forest, secondary growth and extensive banana plantations mixed with trees. Part of the study area is situated in the Northern Sierra Madre Natural Park, which is one of the largest contiguous areas of forest left in the Philippines and which is home to many (endangered) species of plants and animals. Although large-scale commercial logging stopped, the area is still a hotspot of change (Verburg and Veldkamp, 2004) due to agricultural expansion, (illegal) logging activities, and on-going immigration of people that search for land to cultivate.

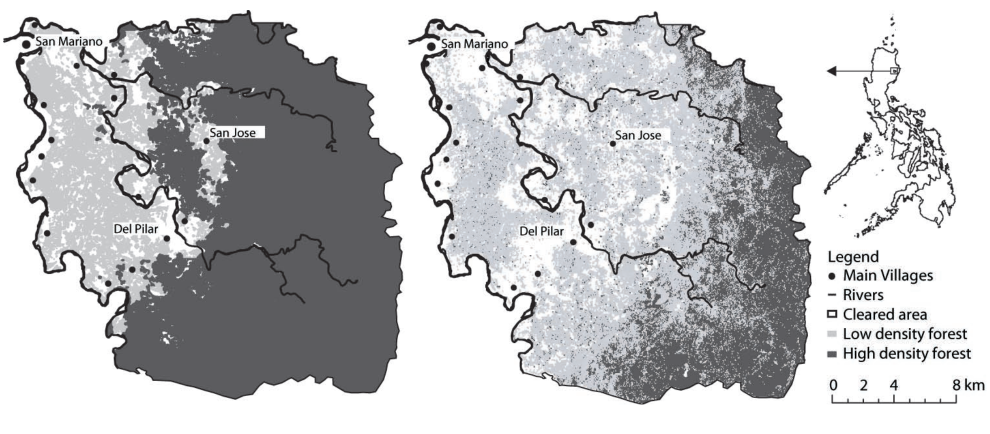


Figure 1.1: Location of the study area in the Philippines (right) and the forest cover in the study area in San Mariano, Isabela, Philippines in 1972 (left) and 2001(center). Interpreted by the authors from aerial photos from the DENR (Department of Environment and Natural Resources) (1972) and satellite imagery (2001).

Since land use change is considered to play an important role in global environmental change it has been given substantial attention in the past and has received even more attention during the last decade. Land use practices influence the global environment and, vice versa, the environment is an important factor in land use decisions. The recognition that land use forms the interface where the human and the natural system interact resulted in a combined project of the IGBP (International Geosphere-Biosphere Program) and IHDP (International Human Dimensions Program on global environmental change) (Turner *et al*., 2004). This so-called LUCC (Land Use/Cover Change) project (Turner *et al*., 1995; Lambin *et al*., 1999) and its successor the Global Land Project (GLP, 2005) aim at integrated social and biophysical research to study the causes and effects of land use change.[1] This dissertation aims to contribute to the methodological questions raised within these projects, as Section 1.3 will detail further.

1.2 Methodological approaches in land use studies

Land change science is by nature a field of science which involves many disciplines including natural, social and geographical information sciences (Rindfuss *et al*., 2004; Turner *et al*., 2004). To study land use, various disciplines have developed their own paradigms and methods. For example, land use studies have been carried out from the perspectives of geography (e.g. Tobler, 1979), economics (e.g. Alonso, 1964), sociology (e.g. Ostrom, 1990), and remote sensing (e.g. Lambin and Ehrlich, 1997). However, these disciplinary approaches can only cover part of the complex system responsible for land use change. It is especially the interaction between the human and the environmental system where land use and land cover change emerges from (e.g. Rindfuss *et al*., 2004; Turner *et al*., 2004). Therefore, in many instances land use scientists have argued that a more integrated, multidisciplinary (or interdisciplinary) methodology is necessary to understand the dynamics of land use change. This view on land use research is the starting point for this dissertation.

Within land use change research three broad categories can be identified (Rindfuss *et al*., 2004): (1) Observation and monitoring of land use change, which involves remote sensing, land use classification systems and quantification of land use changes in the past; (2) Identification of the drivers of land use change and the factors that determine the land use pattern to describe causal processes and (3) Modelling of land use (change) with computer models, which enables combining categories 1 and 2 in a dynamic and integrative manner. Land use change models are important tools in land change science to link information from various sources (Briassoulis, 2000; Verburg *et al*., 2004d). These models can be used to study the processes and dynamics of the land use system and allow researchers to make projections of scenarios of the future. These projections can be visualised to inform policy-makers and to provoke discussion among stakeholders. The work presented mainly covers the fields of driver analysis, process description and dynamic land use modelling (categories 2 and 3).

In this dissertation the concepts of process and pattern of land use change play a central role. Pattern of land use refers to the spatial pattern of land use or land use change over the years, which is represented in maps with a certain resolution and extent. Processes of

---

[1] *A historical overview of the LUCC project is in Moran et al. (2004) and Lambin et al. (2005) and a summary of the major achievements is in IHDP (2005).*

land use change refer to the underlying drivers and proximate causes that explain land use change (Geist and Lambin, 2002). A description of these processes includes land use (change) and the explanatory factors and their causal interactions (mechanisms) that lead to land use change.

To position the research approaches of this dissertation in the wide range of approaches that are used in land use science two distinctive and contrasting methodologies are identified: 'from pattern to process' and 'from process to pattern'. This classification can serve to broadly describe two basic starting points for studying land use change but is not intended to provide a complete classification of methods in land change science. The pattern-based method can be described as a spatially oriented, GIS (Geographical Information System) based approach. The approach starts out with analysing land use patterns by identifying correlations between these observed patterns and explanatory factors and aims at linking these with the processes that are responsible for those patterns. The process-based approach originates from the social sciences and starts with analysing actors and processes, focussing on actors' decision-making. In this approach the interactions between agents play a central role. The actors' decisions are then translated into mapped patterns of land use and land use change. Broadly speaking, the distinction between pattern-based and process-based research coincides with the distinction between inductive and deductive methodologies. The pattern-based approach induces the driving mechanisms from observed land use data. The process-based approach predicts land use change from causal assumptions and then may test these predictions. Examples of modelling from a pattern-based, inductive perspective are cellular automata (White *et al*., 1997) and neural networks based on land use patterns (Pijanowski *et al*., 2002). Many of the agent-based modelling approaches (Parker *et al*., 2003) fall in the category of process-based, deductive approaches.

To integrate process-based and pattern-based methods Geoghegan *et al*. (1998) suggest to 'socialise the pixel' and 'pixelise the social'. 'Socialising the pixel' refers to making remote sensing images more relevant to the social sciences and aims to push the pattern-based approaches beyond their biophysical dimensions. 'Pixelising the social' aims at making bottom-up, field-based approaches spatially explicit, integrate results with remote sensing information and test the social theory in a spatial explicit way. Roughly then, these two approaches appear as relatively concrete methods congruent with the inductive-deductive dichotomy. However, 'socialising the pixel' and 'pixelising the social' aim at bringing the extremes of the inductive 'from pattern to process' and the deductive 'from process to pattern' closer together in order to come to an integrated approach.

1.3 Objectives

The study in this dissertation was carried out as a project within a larger research program called "Integrating macro-modelling and actor-oriented research in studying the dynamics of land use change in North-East Luzon, Philippines", which was funded by the Foundation for the Advancement of Tropical Research (WOTRO) of the Netherlands Organisation for Scientific Research (NWO). Within this program three projects were carried out. One project aimed at spatially-explicit multi-agent modelling of land use change (Huigen, 2004). This project can be regarded as 'process to pattern' research and starts out from the actors and their decisions, builds rules of actor behaviour in a spatial environment and then arrives at a land use pattern. This project was carried out at the most detailed level, explicitly identifying separate land use managers and their fields. The second project applies a pattern-based approach at macro-level for the whole of the Philippines. This geographical, GIS-based approach aims at modelling macro-level processes to identify 'hotspots' of change within the country, which can be used to set priorities for research and policy-making (Verburg and Veldkamp, 2004). The project which this dissertation reports on has an intermediate position and aims at linking the micro-level to the macro-level while at the same time combining elements from various disciplines.

The main objective of this dissertation is to develop methodologies to identify important factors of land use and to integrate these factors in order to describe and model the complex land use system, including the mechanisms of change, in a comprehensive manner. To enable the study of the land use system from various perspectives and to facilitate the integration of human and natural sciences both 'pattern to process' and 'process to pattern' research is carried out. Through 'socialising the pixel' and 'pixelising the social', different methods are brought closer together and integrative methods are developed. The interdisciplinary nature of the research questions results in a series of methodological challenges, which are addressed in this study. These include bridging differences in spatial scales (extent, resolution), organisational levels (social, ecological) and temporal scales; identification of appropriate units of analysis that do justice to the research question; comparing and combining different disciplinary paradigms and developing a new approach that unifies the disciplines. Finally, the project aims to integrate all this information in a spatially-explicit modelling approach.

1.4 Outline

At the beginning of this study little information about land use was available for the study area. Land use in the municipality of San Mariano was studied qualitatively to some extent (Van den Top, 1998), but quantitative data, especially spatial explicit data, and analyses about land use change, its causes and effects were not available. The chapters that form this dissertation can therefore be regarded as progressive insight into the land use system and its context in the area.

In Chapter 2 an exploratory analysis is performed to identify the explanatory factors of land use in the area. Two datasets are analysed and compared: a household dataset starting from the people's perspective and a spatial dataset with land as the starting point. In order to make a first effort to 'pixelise the social' and vice versa, the household analysis is carried out first and the results are used to inform the spatial analysis. To make the household approach more spatially explicit and biophysical, the household analysis uses the field level as the unit of analysis to be able to incorporate land related variables like soils and slope. Household factors that show important relations with land use in the household analysis are included in the spatial analysis, besides a set of more traditional biophysical and geographical variables.

Chapter 3 compares the inductive and deductive approaches to model land use decisions. The chapter starts with the identification of six different steps between purely inductive and purely deductive methods and positions various land use studies on this ladder. Subsequently, a deductive and an inductive approach of analysing land use decisions are presented for the household level. The deductive approach makes use of actor-based, process-oriented research framework originating from the social sciences. The inductive approach uses a statistical approach to derive relations between land use and its explanatory factors. The decision-making theory is applied in a predictive model, tested in a real world case and compared with the results of the inductive approach. This chapter attempts to contribute to the development of interdisciplinary methodology of land use change by combining biophysical and social aspects of land use in one framework.

Chapter 4 deals with integrating different organisational levels and spatial scales by applying a multilevel analysis. This statistical, inductive approach explicitly defines multiple levels within the data and shows what proportion of the variance is explained at which level. The multilevel approach is a statistically sound model for the analysis of data that are hierarchically structured, which is often the case in land use analyses. Explanatory variables can be introduced in the model at their appropriate level, without the necessity to aggregate or disaggregate them before inserting the variables into the model. The construction of the statistical model is informed by the results from Chapters 2 and 3, especially in selecting appropriate variables to be included in the analysis.

In Chapter 5 the information from the analyses of Chapters 2, 3 and 4 is integrated in a dynamic spatial model, which is used to make projections of land use under different scenario conditions. The relations of the deductive model of Chapter 3 are translated to the spatial dataset to create suitability maps that are used in a modelling exercise using the CLUE-S model (Conversion of Land Use and its Effects at Small regional extent, Verburg *et al*., 2002). This approach is compared with a CLUE-S model that incorporates suitability maps derived with a statistical, inductive analysis. This chapter discusses the differences in outcome and the differences in applicability of both modelling approaches in policy analysis.

In Chapter 6 the effects of land use change are assessed for biodiversity. For three land use scenarios land use maps are projected for the year 2015, using the deductive modelling approach of Chapter 5. These land use changes are examined for their effects on endemic bird species richness in the area in a spatially-explicit way by using the relation between landscape characteristics and the occurrence of birds. The scenarios differ in the level of agricultural expansion and forest management. The value of the approach to evaluate policy options for land use and conservation management is discussed.

In the final chapter the experiences and conclusions regarding the interdisciplinary approach of this study are discussed and the main methodological conclusions are summarised and used to formulate recommendations for further research.

2

Analysis of land use drivers at the watershed and household level: Linking two paradigms at the Philippine forest fringe

Abstract

Land use and land cover change (LUCC) is the result of the complex interactions between behavioural and structural factors (drivers) associated with the demand, technological capacity, social relations and the nature of the environment in question. Different disciplinary approaches can help us to analyse aspects of LUCC in specific situations, though paradigms and theories applied by the different disciplines are often difficult to integrate and their specific research results do not easily combine into an integrated understanding of LUCC. Geographical approaches often aim at the identification of the location of LUCC in a spatially explicit way, while socio-economic studies aim at understanding the processes of LUCC, but often lack spatial context and interactions. The objective of this study is to integrate process information from a socio-economic study into a geographical approach. First, a logistic regression analysis is performed on household survey data from interviews. In this approach the occurrence of the land use types corn, wet rice and banana is explained by a set of variables that are hypothesised to be explanatory for those land use types, with fields as the unit of analysis. The independent variables consist of household characteristics, like ethnicity and age, and plot and field information, like tenure, slope and travel time. The results of these analyses are used to identify key variables explaining land use choice, which subsequently are also collected at watershed level, using maps, census data and remote sensing imagery. Logistic regression analysis of this spatial dataset, where a ten percent sample of a 50 by 50m grid was analysed, shows that the key variables identified in the household analysis are also important at the watershed level. Important drivers in the study area are, among others, slope, ethnicity, accessibility and place of birth. The differences in the contribution of the variables to the models at household and watershed level can be attributed to differences in spatial extent and data representation. Comparing the model with a mainstream geographical approach indicates that the spatial model informed by the household analysis gives better insight into the actual processes determining land use than does the mainstream geographic approach.

Based on: Overmars, K.P., and Verburg, P.H. 2005. Analysis of land use drivers at the watershed and household level: Linking two paradigms at the Philippine forest fringe. International Journal of Geographical Information Science 19 (2), 125-152.

2.1 Introduction

Land use and land cover change (LUCC) research has received much attention during the past decade, because of the pivotal role of LUCC in many urgent issues like global climatic change, food security, soil degradation and biodiversity (Turner II *et al*., 1995; Lambin *et al*., 2001; Geist and Lambin, 2002). LUCC research involves many disciplines, since it operates at the interface of natural and human sciences. LUCC is the result of the complex interaction of behavioural and structural factors associated with the demand, technological capacity, social relations and the nature of the environment in question. A theory of land use change, therefore, needs to conceptualise the relation between the driving forces and land use change, relations among the driving forces, and human behaviour and organisation. Different disciplinary theories can help us to analyse aspects of land use change in specific situations. The synthesis of these theories is essential, but the paradigms and theories applied by the different disciplines are often difficult to integrate and their specific research results do not easily combine into an integrated understanding of LUCC. Up to now researchers have not yet succeeded in integrating all disciplines and complexity of the land use system into an all-compassing theory of land use change (Verburg *et al*., 2004d). Conclusions drawn from disciplinary LUCC studies can vary substantially between disciplines (Lambin *et al*., 2001), which implies that the complexity of the land use system as a whole is not completely understood.

From a geographical perspective LUCC studies have been carried out mainly at national and sub-national level, using available geographic information from maps, census data and remote sensing. These data are used to construct driving factors of land use change that are used to explain the location of land use change (Veldkamp and Fresco, 1997; Kok and Veldkamp, 2000; Serneels and Lambin, 2001; Nelson *et al*. 2001; Pontius *et al*., 2001). What is often lacking in these studies is explicitness about processes and human behaviour. The drivers used are proxies for the processes that determine land use change. The identified relations between land use change and the supposed driving factors are valid at the pixel level and do not straightforwardly translate into the determinants of LUCC at the household level, the level that is central in decision-making. The strength of this geographical approach is its spatial explicitness that enables to explain land use pattern, which can be directly used in geographical modelling approaches (e.g. Pontius *et al*., 2001; Verburg *et al*., 2002, Pijanowski *et al*., 2002). This approach contrasts with the approach of the social sciences that generally conduct micro-level studies aiming at the understanding of people environment relations (Turner, 2003).

Socio-economic studies often focus at the household level to gain insight in the factors that influence land use decisions. These studies provide information about decision-making processes and human behaviour. But, in general, they do not incorporate a spatial component. Therefore, the relation between the households and the biophysical environment and their interactions and spatial dependencies are not represented, consequently disregarding the spatial nature of the problem (Geoghegan *et al*., 1998).

In literature it is acknowledged that for a better understanding of the land use system it is important to combine the strengths of both approaches and to come to an integrated approach by linking the social and geographic disciplines (Liverman *et al*., 1998; Walsh and Crews-Meyer, 2002; Fox *et al*., 2003). The process that enhances the link between the social sciences and the geographical sciences are often referred to as 'socialising the pixel' and 'pixelising the social' (Geoghegan *et al*., 1998).

'Socialising the pixel' can be described as moving from patterns to processes. Information within spatial imagery that is relevant for the social sciences is identified and used to inform concepts and theories (Lambin *et al*., 1999; Geoghegan *et al*., 1998). Some recent LUCC studies have presented preliminary results that link the pattern from geographical approaches to the human behaviour by incorporating landscape data in social data. A number of studies aim to link household level data directly to pixels in remote sensing images (e.g. Vance and Geoghegan, 2002; Walsh *et al*., 2003) to better understand the human-environment interaction. Mertens *et al*. (2000) aggregate household level data to the village level and combine the aggregated data at that level with spatial data. Walker *et al*. (2000) and Staal *et al*. (2002) base their analyses on household level data, but add spatial data to the household data using the geographical position of the households.
The other way around, 'pixelising the social' involves moving from processes to patterns. For example, socio-economic theory is tested in a spatially explicit way (e.g. Chomitz and Gray, 1996). Other approaches, like multi-agent modelling start with social and decision-making theories and move from there to construct spatial explicit models (Parker *et al*., 2002).
The approach applied in this study explores the results of statistical models based on socio-economic theories at the household level and uses the outcomes in the construction of geographical models in order to incorporate the theories about human decision-making in these spatially explicit models. This approach aims to link the widely used geographical approaches based on statistical models (Veldkamp and Fresco, 1997; Kok and Veldkamp, 2000; Serneels and Lambin, 2001; Nelson *et al*., 2001; Schneider and Pontius, 2001) and the socio-economic approaches using household level data (Walker *et al*., 2000; Staal *et al*., 2002; Vance and Geoghegan, 2002).

The objective of this chapter is to provide an alternative approach for the mainstream geographical studies that are applied in LUCC research in order to give more attention to the processes and behaviour that determine the land managers' decisions. The core of the approach is to use the understanding of socio-economic processes and environmental constraints at the household level and exploit those to create process related spatial variables at the watershed level ('pixelising the social'). With this new set of process-relevant variables an empirical model is constructed in which the variables are examined for their explanatory power to predict the current land use pattern. Using this approach we aim to construct a spatial model at the watershed level that has a better statistical fit than the mainstream geographical approach and gives better insight in what processes (driving forces) are important in the decision-making process of the land managers.
The socio-economic approach and the geographical approach often work at different scales and at different organisational levels. This alternative approach aims to provide tools and methods to facilitate the exchange of information between the two approaches.

2.2 Study area and data collection

2.2.1 Study area

The study area is situated in Cagayan Valley in the northeastern part of the island Luzon, The Philippines (Figure 2.1). The study area comprises 16 villages of the municipality of

San Mariano, Isabela province, and its size is approximately 26,000 ha. San Mariano is accessible by concreted road in a 30 minutes drive from the highway leading from Manila to the north. The study area is situated between the town of San Mariano in the west and the forested mountains of the Sierra Madre mountain range in the east. The mountainous area in the east consists of metamorphic and intrusive rocks as well as limestone and the hilly area in the west consists of dissected marine deposits. The elevation ranges from 40 to 800 m.a.s.l. The climate is hot and humid, but with strong spatial and temporal variations. A short dry season occurs between November and May (Van den Top, 1998).

The area is inhabited by approximately 16,500 persons (about 3,150 households) of various ethnic groups, among others: Ilocano, Ibanag and Ifugao, who are migrants or descendents of migrants that came to the area from the 1900s onwards, and Kalinga and Agta, who are the indigenous inhabitants of the area. In the migration history of the area some general patterns can be identified. A century ago the whole study area was covered with tropical rain forest and only few people lived in the area. From the 1900s to 1940 migrants from the nearby Cagayan valley settled in the area and started small scale (selective) logging for construction purposes and some local trade. In the same period some waves of migrants came from Ilocos to look for land to cultivate. In the period after World War II up to 1960 people from Cagayan valley and the Cordillera (central Luzon) came to look for prime agricultural land. From 1960 –1990 people entered the area for employment in the logging industry, coming from Cagayan valley, the Cordillera, and from other logging areas. The latter are predominantly Tagalog speaking people by origin, though currently they speak Ilocano. Between 1960 and 1990 corporate logging companies deforested large parts of the area. In 1989 a logging moratorium was issued in San Mariano. This moratorium was lifted in 1990, however, in 1992 another moratorium was enacted. By that time the logging companies had already pulled out of the area (Jongman, 1997). The moratorium made the people switch from logging based activities to agriculture. At present, most people in the area are farmers. From 1990 to present there are still migrants coming to the area. Those people migrate mainly because of livelihood problems in their own area, for land speculation or because they are invited by relatives that migrated before (Van den Top, 1998).

During the time of corporate logging activities accessibility of the area was relatively good. The companies constructed logging roads to transport logs out of the area. People and goods were transported with the same trucks as the logs. Most of the current roads still follow the former logging roads. Though, since the logging moratorium the accessibility decreased, because of a lack of maintenance of the roads, which was formerly done by the logging companies (Jongman, 1997). Currently, the situation is improving because of the efforts of the municipal government. All transport out of the area passes through San Mariano proper, which is the main market for selling products and buying agricultural inputs.

At present, the land use in the study area shows a gradient from intensive agriculture (mainly rice and yellow corn), near San Mariano, via a scattered pattern of rice, yellow corn, banana, grasses and trees, to residual and primary forest in the eastern part of the study area.

2.2.2 Data collection

Three datasets were collected: a spatial dataset for a mainstream geographic analysis, a household dataset based on questionnaires and an enhanced spatial dataset consisting of

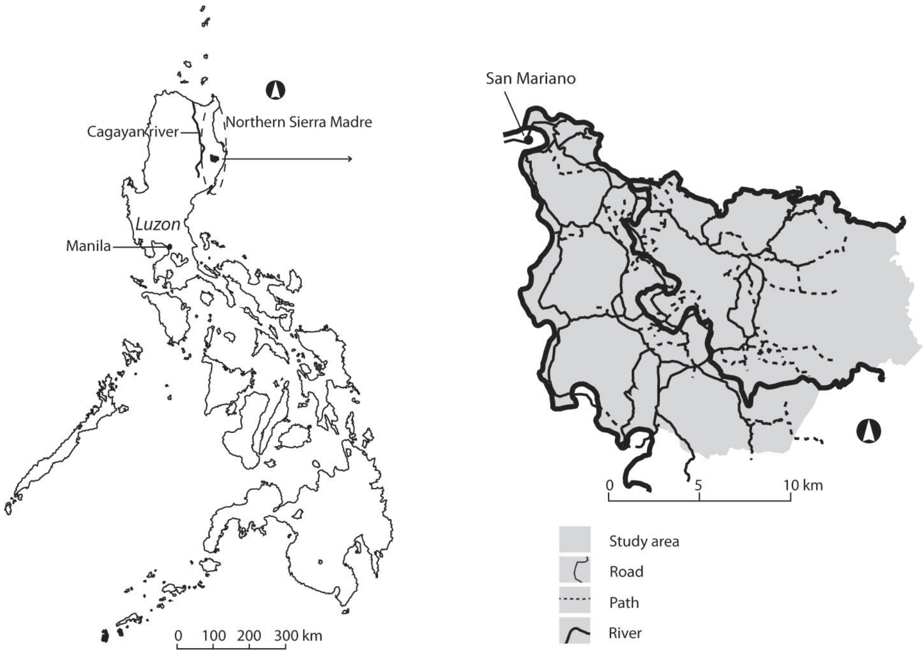


Figure 2.1: Location of the study area in the Philippines (left) and a close up of the study area (right)

maps with variables that were selected based on the household level analysis completed with other maps that are considered to be also explanatory for the land use in the study area. The spatial datasets are created independently of the household dataset; no information from the household level was aggregated to construct the spatial dataset, but instead other sources of information were used that more fully cover the whole area and give a better representation than aggregated household data.

Land use data for the two spatial approaches

Land use data were interpreted from Landsat ETM+ data (http://www.landsat.org) from June 2001 and ASTER data from March 2002. First, unsupervised classifications were made from subsets of both images. Second, the classes of the unsupervised classifications were recoded into a land use map according to a set of 96 observations of the present land use. Finally, the land use map was constructed by combining the classifications of the two images. In this procedure the ASTER image was first resampled from 15m resolution to the same grid as the Landsat image (30 by 30 m). Then, the land use classes of the 2 images were put in separate layers. In a GIS (Geographical Information System) these layers were combined, using overlay, in such a way that the best land use classification was established according to the field observations. For each land use type the image was used that best distinguished that land use type. For example, the ASTER image was best able to distinguish forested areas, so this classification was put on top the Landsat classification of a banana/secondary growth mixture that included parts of forested areas. Finally, the image was resampled to a 50 by 50 m grid that coincides with the other data. The classes in the final land use map are yellow corn (including some other arable crops), wet rice, grass, forest and a class that includes banana, secondary forest, reforestation and residual forest (Figure 2.2).

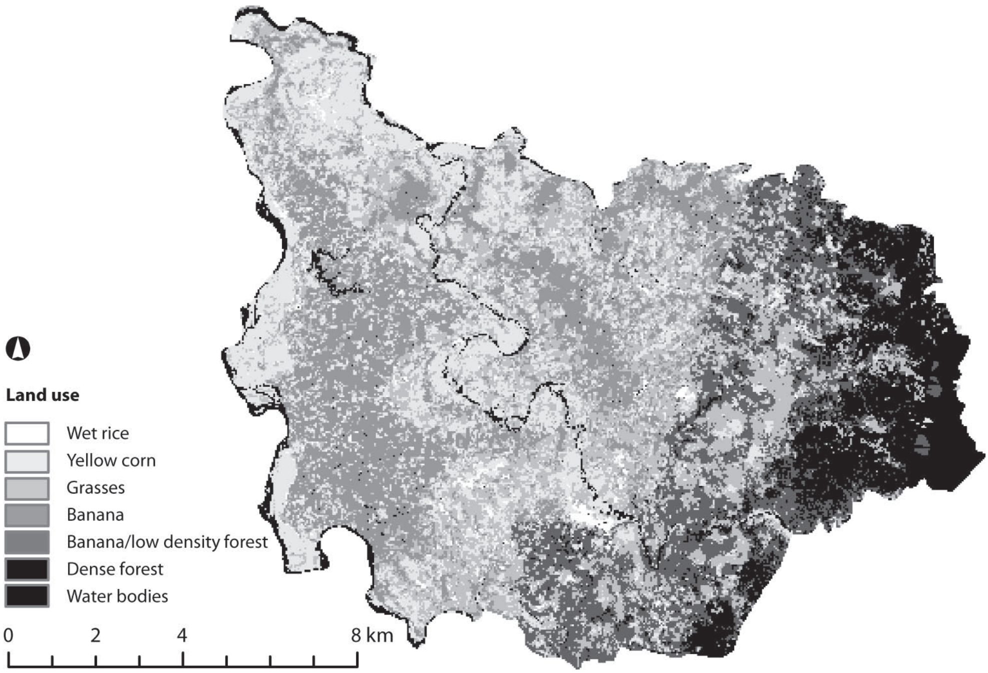


Figure 2.2: Land use map

Banana plantations and low-density forest types were difficult to distinguish, because the banana cultivation is quite extensive and often many trees grow in between the bananas. Especially close to the forest area this causes difficulties in classifying, because close to the forest many residual and secondary forest occurs. Therefore, a subset of the study area was created based on field observations. The western half of the area was identified as an area in which the class 'banana, secondary forest, reforestation, residual forest' can be considered to contain almost exclusively extensive banana plantations. In the analysis for banana only this area was used. The forested part of the study area can be regarded as land use (and not only as land cover) since all forest has been commercially logged in the past. Currently, the forest is mostly regrowing and in some parts small-scale logging takes place.
Classification accuracy of the land use map is 68 percent, which was calculated using an independent sample of 76 field observations (Verburg *et al*. 2004a).

Spatial data for mainstream geographic approach
Following the approach of the mainstream spatial geographical models (e.g. Verburg and Chen, 2000; Schneider and Pontius, 2001; Stolle *et al*., 2003) a dataset is constructed using data that are readily available. The spatial dataset is a set of maps in a GIS containing information derived from census data, maps, and field surveys (Table 2.1) collected at the watershed or meso-level. With these data spatial measures are constructed that are proxies for the processes that determine the location of different land use types. The data are converted into uniform grids with cell size 50 by 50 meter to facilitate the analysis.
Distance measures are calculated as the Euclidean distance of a cell to the nearest destination of interest, which is a method that is often applied in the mainstream spatial geographical

models of land use change. The destinations of interest are the market place in San Mariano, the nearest road (roads that are accessible by all vehicles during dry season), the *sitios* (villages and smaller settlements), and rivers and streams. A digital elevation model (DEM) was derived from the contour lines and elevation points of a 1:50,000 topographic map of the area (NAMRIA, unknown). From the 50 by 50 m DEM a slope map was derived. A population pressure map was constructed using a map with villages (as points) and the number of inhabitants per village. It is assumed that the population pressure is related to the number of inhabitants in the village and is higher close to the village than at distance. The assumption is that villagers want to have land nearby their house, because of accessibility and safety reasons, and land nearby is scarce. Therefore, the population pressure in a cell was calculated as the number inhabitants in a village divided by the distance to that village, summed up for all villages (after Haynes and Fotheringham, 1984). In this model the pressure is high near the village and diminishes quickly with increasing distance. In this approach the influence of a village stretches throughout the whole study area and does not stop at administrative village boundaries.

Household level data

To collect the household level data an interview campaign was carried out between June and November 2002 using a structured questionnaire. The selection of explanatory variables of land use to be incorporated in the questionnaire was based on literature (Doorman, 1991), theories from a range of disciplines and expert knowledge of the area. Some of the theories that were considered while constructing the questionnaire are the relation between land use and accessibility (e.g. Chomitz and Gray, 1996), land suitability, and household life cycles (Perz and Walker, 2002). The aim was to construct a questionnaire containing all variables that potentially have an influence on land use decisions of farmers in the area. The hypothesised relations are provided in the description of the data (Table 2.2). During a 2-month field survey in 3 different *barangays* (villages), the questionnaire was tested, a range of possible answers was determined and the questions were adapted to the understanding of the villagers. It is important to consider what questions will best fit to the purpose under study. The standardised questionnaire was written in English and was translated during the interview by the interpreter/field assistant in a local language (either Ilocano or Ibanag).

The selection of households to be interviewed was based on a combination of stratified sampling and systematic random sampling using population data available at the POPMAT (population manipulation action team) member in the village. Interviews were carried out in 13 of the 16 *barangays* under study. The sample was stratified according to these 13 *barangays*. This sampling strategy was selected to obtain an equal coverage of the households over the study area according to the relative population size of the village. In all 13 *barangays* every twentieth household was selected (systematic random sampling with sampling interval 20) from the POPMAT. Because the POPMAT data were structured by purok (neighbourhood) an extra spatial stratification was introduced. A total of approximately 151 households were interviewed. The number of interviews per barangay ranges from 6 in small *barangays* to 20 in the biggest.

The household survey is structured in a nested hierarchy (Figure 2.3), with at the top the household level and plot and field level underneath it. The household is defined as the group of persons sharing one housing unit. The plot is defined as a piece of land owned or used by the household. A field is defined as a specific part of the plot used for one land use type or crop. A household often owns or uses different plots at different locations and

Table 2.1: Description and descriptive statistics of the variables of the mainstream geographic and enhanced spatial dataset (n=9100)

						Hypothesised relations		
Variable name	**Description**	**Min.**	**Max.**	**Mean**	**St.dev.**	**Corn**	**Rice**	**Ban.**
Dependent variables (both reference model and spatial model)								
Corn	1 if cell is corn, 0 otherwise	0	1	0.21				
Banana	1 if cell is banana, 0 otherwise	0	1	0.35				
Wet rice	1 if cell is wet rice, 0 otherwise	0	1	0.02				
Independent variables mainstream geographical model								
Dist. to river	Distance to nearest river or stream (m)	0	1341	378	258	-	-	+
Dist. to village	Distance to nearest village (m)	0	4978	1419	987	-	-	no
Dist. to market	Distance to market (m)	427	24003	13481	5361	-	no	-
Dist. to road	Distance to nearest road (m)	0	7567	1129	1315	-	-	-
Elevation	Elevation (m.a.s.l.)	38	724	203	114	-	-	no
Slope	Slope (degrees)	0	42.72	8.21	6.06	-	-	+
Population pressure	Sum of (persons in village)/distance for all villages (pers./m)	1.93	31.81	4.94	2.22	+	+	no
Independent variables enhanced spatial model								
Slope	Slope (degrees)	0	42.72	8.21	6.06	-	-	+
Impr. dist. to market dry	Improved distance to market dry season, calculated as travel time (s)	525	33475	10836	4942	-	no	-
Impr. dist. to market wet	Improved distance to market wet season, calculated as travel time (s)	525	37885	13172	6716	-	no	-
Impr. dist. to village	Improved distance to nearest village, calculated as travel time (s)	2	26591	4718	4467	-	-	no
Impr. dist. to road	Improved distance to nearest road, calculated as travel time (s)	2	24239	3149	3601	-	no	-
Dist. to small river	Distance to nearest small river (m)	0	1504	440	281	no	-	no
Dist. to big river	Distance to nearest big river (m)	0	8638	1897	1630	+	no	no
Ethn. Ilocano	Sum of (persons of ethnicity Ilocano in village)/(distance) for all villages (pers./m)	0.71	12.25	1.85	0.81	+	+	no
Ethn. Ifugao	Sum of (persons of ethnicity Ifugao in village)/(distance) for all villages (pers./m)	0.05	4.86	0.15	0.12	-	+	no
Ethn. Kalinga	Sum of (persons of ethnicity Kalinga in village)/(distance) for all villages (pers./m)	0.03	2.36	0.09	0.07	no	no	no

Table 2.1: (Continued)

						Hypothesised relations		
Variable name	**Description**	**Min.**	**Max.**	**Mean**	**St.dev.**	**Corn**	**Rice**	**Ban.**
Ethn. Ibanag	Sum of (persons of ethnicity Ibanag in village)/(distance) for all villages (pers./m)	0.30	21.32	0.83	0.74	+	-	no
Tax declaration	Percentage of the village area that is registered to a tax payer	13	105	45.22	30.54	+	+	no
Org. municipality	Fraction of the village population that is born in the municipality of San Mariano	0.37	0.98	0.55	0.21	+	-	no
Project ISF	Area dedicated for DENR project Integrated Social Forestry	0	1	0.04		no	no	no
Project SIFMA	Area dedicated for DENR project Socialized Industrial Forest Management Agreement	0	1	0.19		no	no	no
Project FLMA	Area dedicated for DENR project Forest Land Management Agreement	0	1	0.02		-	-	-
Project IFMA	Area dedicated for DENR project Industrial Forest Management Agreement	0	1	0.04		-	-	-
Geo. limestone	Geomorphology: Mountainous with limestone parent material	0	1	0.08		no	-	no
Geo. terraces	Geomorphology: Terraces	0	1	0.12		+	+	-
Geo. marine sed.	Geomorphology: Hilly with marine sediments as parent material	0	1	0.50		-	no	+
Geo. active floodplain	Geomorphology: Active floodplain	0	1	0.06		+	+	-
Geo. Rock	Geomorphology: Mountainous with metamorphic and intrusive rocks	0	1	0.25		-	no	no

each plot might be cultivated with a different crop. Each of the variables was collected at the appropriate level, e.g. soil characteristics at the field level, accessibility at plot and household level and household structure at the household level (Table 2.2).
The location of the households was recorded. However, the location of the fields was not made spatially explicit due to time constraints, except for a few field checks. Studies that do map the fields often use this information to link data from other sources, like maps, to the fields. In this household analysis all information regarding the plots and fields, like size, land use, slope and soil, was obtained through questioning the respondents. Therefore, mapping of the fields was not strictly necessary. The consequence of obtaining all data through questioning is that the values represent the characteristic as perceived by the farmers instead of a more objective method.

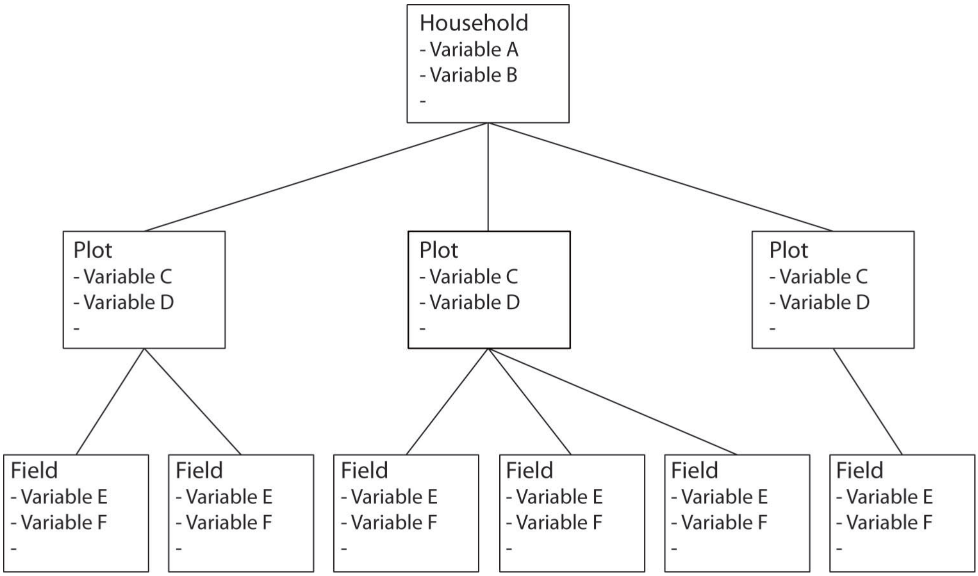


Figure 2.3: Hierarchical structure of the household level dataset presenting the relation between the levels Household, Plot and Field

Enhanced spatial dataset

The land use data as well as the slope data in this dataset are the same as in the mainstream geographic dataset. Besides this, additional variables were included according to the insights obtained in the household analysis (Table 2.1) about the explanatory factors for land use in the area. These variables are considered also to be possible drivers in the spatial analysis at the watershed level. To construct the enhanced spatial dataset we did not use the data of the household survey, but instead information was used from maps, census and field surveys that had the same theme.
Many variables that can be observed at the household level are difficult to represent in a spatially explicit way (e.g. age of a household member). However, it is possible to construct 'creative' spatial variables (Geoghegan *et al*., 1998) that best represent the processes affecting the land use decisions. For these spatial variables other data sources are used like census

data from the municipal office or data collected through other surveys. For example, the ethnicity of the owner of each grid cell was not determined by collecting this information through a survey, but separate population pressures were calculated for each ethnicity based on census data.

In contrast to the mainstream approach, the enhanced approach incorporates improved accessibility measures based on an in depth study on accessibility (Witte, 2003; Verburg *et al*., 2004a). Four accessibility measures were created for this study: travel time to market in the dry season, travel time to market in the wet season, travel time to the nearest village and travel time to the nearest road. Due to bad roads and higher water levels in the rivers a substantial difference exists between the travel time in wet and dry season. To study whether the limitations of the wet season or the opportunities in the dry season are most explanatory for land use both were taken into account in the analysis. Witte (2003) used the travel speed on different types of roads and travel speed off road depending on slope to calculate travel time to the destinations market, village and road.

The measure *distance to river* as used in the mainstream approach was separated in a measure for big rivers and a measure for the small rivers, because the small rivers can often be used for irrigation purposes, while the big rivers cannot unless pumps or large irrigation systems are available. Big rivers can be used as a way to transport goods in the wet season and illegally cut logs. For the distance measures to rivers the Euclidean distance was used.

It was not possible to obtain a map that depicts the ethnicity of the individual landowners, because in this study no database was available that links all land managers to their individual parcels. Instead, an indicator was created to represent ethnicity based on the population census data. For the four largest ethnic groups, Ilocano, Ibanag, Ifugao and Kalinga, an 'ethnic population pressure' was created. The procedure to calculate this measure is similar to the procedure used to calculate population pressure for the mainstream approach, though in the new measures the numbers of inhabitants were disaggregated into the number of people per ethnic group to create four ethnic population pressure maps.

Information about the place of birth and tenure were available at village level. So, a map of village territories was necessary. Therefore, GPS (global positioning system) recordings of all settlements were used to construct Thiessen polygons that delineate a map with the village boundaries. Place of birth is represented as the percentage of male inhabitants born in the municipality of San Mariano. The variable *tax declaration* is the percentage of land per village that is registered to a land manager by the municipal office.

Besides the variables above two sets of variables were included that are considered to have an important contribution to the spatial distribution of the land use, but which were not taken directly into account in the household survey. Forest-related land use policies (DENR-CENRO, 1998) were incorporated to see whether or not these programs have any influence on the land use in the area. Generally, these policies are related to forest conservation, protection and development. FLMA (Forest Land Management Agreement), ISF (Integrated Social Forestry), and SIFMA (Socialized Industrial Forest Management Agreement) are reforestation and agroforestry programs based on community participation. The participants are granted tenural security for 25 years and are committed to achieve the goals of the program regarding the planting of trees. The programs aim at providing sustainable livelihood for the occupants based on the sustainable use of forest products. A part of the area, which varies per program, is allowed to be devoted to agricultural crops. The IFMA (Industrial Forest Management Agreement) programs aims at developing industrial forest

Table 2.2: Description and descriptive statistics of the variables of the household analysis (n=187)

Variable name	Description	Min.	Max.	Mean	St.dev.	Hypothesised relation Corn	Rice	Ban.
Household level variables								
Transportation cost	Cost to transport a bag of corn to San Mariano (pesos)	7	45	25.19	12.85	-	no	-
Average age	Average age of household heads (years)	20.50	78	41.59	12.08	-	no	no
Education male	Education of the male household head (years)	0	14	5.80	3.36	+	no	no
Education female	Education of the female household head (years)	1	14	6.52	3.41	+	no	no
Ethn. Ilocano male	1 if male household head is Ilocano speaking, 0 otherwise	0	1	0.54		+	+	no
Ethn. Ibanag male	1 if male household head is Ibanag, 0 otherwise	0	1	0.27		+	-	no
Ethn. Kalinga male	1 if male household head is Kalinga, 0 otherwise	0	1	0.01		no	no	no
Ethn. Ifugao male	1 if male household head is Ifugao, 0 otherwise	0	1	0.14		-	+	no
Ethn. Ilocano female	1 if female household head is Ilocano speaking, 0 otherwise	0	1	0.60		+	+	no
Ethn. Ibanag female	1 if female household head is Ibanag, 0 otherwise	0	1	0.16		+	-	no
Ethn. Kalinga female	1 if female household head is Kalinga, 0 otherwise	0	1	0.05		no	no	no
Ethn. Ifugao female	1 if female household head is Ifugao, 0 otherwise	0	1	0.14		-	+	no
Place of birth male	1 if male household head is born in San Mariano, 0 otherwise	0	1	0.49		+	-	no
Place of birth female	1 if female household head is born in San Mariano, 0 otherwise	0	1	0.63		+	-	no
1st year of farming	Year that the respondent started his/her own farm	1945	2001	1983	11.74	+	-	no
Number of buffalos	Number of water buffalos currently owned by the household	0	7	1.54	1.38	+	+	no
Other income	No. of other activities from homegarden, fishpond, pigs, cows	0	3	1.48	0.91	-	no	no
Number of plots	Total number of plots owned and/or cultivated by the household	1	8	2.94	1.58	+	+	no
Total area	Total land area (ha)	0.25	87	8.70	19.37	-	-	+
Workshop	1, if workshop attended by one of heads, 0 otherwise	0	1	0.16		+	no	no
Farming 1st income	1, if farming is most important income generating activity, 0 otherwise	0	1	0.94		+	no	no
Farming 2nd income	1, if farming is second most important income generating activity, 0 otherwise	0	1	0.04		-	no	no
No. of non-dependents	No. of people currently living in the household older than 10 years	2	8	3.76	1.52	+	no	no
Dependents/non-dep.	No. of people younger than 11/ no that are 11 yrs or older	0	3	0.54	0.63	+	-	no
Plot level variables								
Plot size	Total size of the plot (ha)	0.13	45	2.84	6.69	-	-	+
Plot distance	Minutes walking to the plot (min)	0	600	23.67	50.6	-	-	no
Tenure position	1 if the plot is "in position", 0 otherwise	0	1	0.29		-	-	+

Table 2.2: (Continued)

						Hypothesised relation		
Variable name	**Description**	**Min.**	**Max.**	**Mean**	**St.dev.**	**Corn**	**Rice**	**Ban.**
Tenure tax	1 if there is a tax declaration for the plot, 0 otherwise	0	1	0.22		+	+	no
Tenure title	1 if the plot is titled, 0 otherwise	0	1	0.34		+	+	-
Tenure SIFMA	1 if the plot is a SIFMA, 0 otherwise	0	1	0.09		-	-	no
Acquire cleared	1 if the household acquired the plot by clearing the plot, 0 otherwise	0	1	0.14		+	+	no
Acquire inherited	1 if the household acquired the plot by inheritance, 0 otherwise	0	1	0.38		no	no	no
Acquire tenant	1 if the household is tenant of the plot, 0 otherwise	0	1	0.15		+	no	-
Acquire bought	1 if the household acquired the plot by buying, 0 otherwise	0	1	0.26		+	+	-
1st year on plot	Year that the respondent started farming on this plot	1950	2002	1989	10.8	+	no	-
Creek	1 if there is a creek or spring through or bordering the plot, 0 otherwise	0	1	0.58		no	+	no
Field level variables								
Yellow corn	1 if yellow corn, 0 otherwise	0	1	0.56		Dep. var.		
Wet rice	1 if wet rice, 0 otherwise	0	1	0.13		Dep. var.		
Banana	1 if banana, 0 otherwise	0	1	0.21		Dep. var.		
Flat slope	1 if slope category is "flat", 0 otherwise	0	1	0.42		+	+	-
Flat-moderate slope	1 if slope category is "flat to rolling/moderate", 0 otherwise	0	1	0.22		+	+	-
Moderate slope	1 if slope category is "rolling/moderate", 0 otherwise	0	1	0.29		-	-	+
Moderate-steep slope	1 if slope category is "rolling/'moderate to steep/hilly", 0 otherwise	0	1	0.06		-	-	+
Steep slope	1 if slope category is "steep/hilly", 0 otherwise	0	1	0.01		-	-	+
Flooding risk	1 if there is a risk of flooding of the field, 0 otherwise	0	1	0.20		no	no	-
Red soil	1 if soil colour is red, 0 otherwise	0	1	0.20		-	-	no
Black soil	1 if soil colour is black, 0 otherwise	0	1	0.18		+	+	no
Brown soil	1 if soil colour is brown, 0 otherwise	0	1	0.33		+	+	no
Brown/red soil	1 if soil colour is brown/red, 0 otherwise	0	1	0.16		-	-	no
Brown/black soil	1 if soil colour is brown/black, 0 otherwise	0	1	0.09		+	+	no
Fertile soil	1 if soil fertility category is high, 0 otherwise	0	1	0.11		+	+	no
Mod. fertile soil	1 if soil fertility category is moderate, 0 otherwise	0	1	0.57		+	+	no
Poor soil	1 if soil fertility category is low, 0 otherwise	0	1	0.28		-	-	no

plantation as an alternative and sustainable source of raw material for private corporations involved in forest based industries (Balagtas-Mangabat, 2002). Geomorphological variables (Van Egmond, 2003) were included to approximate landscape characteristics. The area was subdivided into five areas: active floodplain, terraces, marine sediments, limestone and metamorphic and intrusive rocks.

2.3 Methods

2.3.1 Analysis

In this chapter three logistic regression models are presented. The analyses focus on the current land use rather than land use change. First, a model is constructed using the spatial data of the mainstream geographic approach. This model is presented to illustrate the difference with the approach advocated in this study. Second, a model is presented using the data collected in the household survey. This model will be referred to as the household model. Third, a model referred to as the enhanced spatial model is constructed based on the explanatory drivers identified in the household level analysis supplemented with specific spatial drivers. This is the model aimed at in this study: a spatial model incorporating proxies for process information that does justice to the causal relations in land use change decision-making having a better predictive power than ordinary models. For all three analyses three land use types were analysed: yellow corn, wet rice and banana. Forest could also be studied in the spatial approaches, but this was not analysed in this study, because forest was not included in the household survey.

In the spatial models we are interested in the occurrence of a land use type relative to all other land use types including forest and other non-agricultural uses. Therefore, the logistic regression approach was chosen. For the household analysis a multinomial approach could have been appropriate, since only agricultural options are included in the model and in the dataset. In multinomial regression the categories are explained against a reference category. In this study we want to explain every land use type relative to all other options rather than relative to one specific land use type. Therefore, we decided to apply logistic regression analysis in this study.

Using logistic regression the assumption is made that all people in the area respond in a similar way to the variables. Though this does not have to be the case. A possible way to integrate the effects of communities (like villages or ethnic groups) and households within a single model is to use a multilevel model (Goldstein, 1995; Polsky and Easterling, 2001). In the multilevel approach the estimated parameters of the model are allowed to vary according to the hierarchical stratification of the data.

Beforehand, there was no complete insight in the processes determining land use in the area. Therefore, a stepwise procedure is used in this study to construct the logistic regression models in order to explore what variables may be explanatory for the observed land use.

To see whether the linkage and integration of the socio-economic and geographical approaches succeeded, the results of the household analysis and the watershed level analysis are compared and discussed. To assess the benefit of the alternative approach in comparison with the mainstream geographical approach the results of those models are also compared.

Mainstream geographic approach

In the mainstream geographic approach a logistic regression model is constructed in which the probability of the occurrence of a land use type at a location is estimated as a function of explanatory variables. For the selection of relevant factors explaining the pattern of land use a stepwise procedure was used (forward stepwise regression with probability levels of 0.01 for entry in the model and 0.02 for removal from the model). The independent variables are proxies of land use drivers and considered to explain the location of the different land use types. The following variables were included in the stepwise procedure: distance to market, distance to village, distance to road, distance to river, slope, elevation and population pressure. The hypothesised relations are listed in Table 2.1. A ten percent sample from the available grid cells was drawn to reduce spatial autocorrelation. This approach does not fully account for spatial autocorrelation and is in fact a loss of information (Overmars *et al*., 2003). However, it is commonly used and will minimise spatial autocorrelation to a level that it will not affect the results (Verburg and Chen, 2000; Serneels and Lambin, 2001; Stolle *et al*., 2003). Practical procedures that can fully account for spatial autocorrelation in logistic models are currently not available.

Household model

The household model is a logistic regression model in which the probability for a field to have a land use type or not is estimated as a function of explanatory variables. The model is based on the data collected in the household survey. All variables (Table 2.2) are hypothesised to be explanatory factors for land use. They are assumed to influence the preference of the land managers for a land use type at a certain location. From the variables a selection was made using a stepwise procedure (forward stepwise regression with probability levels of 0.05 for entry in the model and 0.10 for removal from the model) to select variables from the household survey to form a model to fit the land use data. Records with a missing value in one of the variables were removed from the dataset. Therefore, a subset of 187 observations (fields) from a total of 376 was used for this analysis. Models are constructed for the land use types yellow corn, wet rice and banana.

In most socio-economic studies of this kind the household level is the level of analysis since this is the level at which the land manager take his/her decisions. For example, whether or not a household adapts a certain agricultural technique or not is tested. But, using the household as the level of analysis, it is difficult to take field characteristics, such as soil quality and flooding risk, into account. These characteristics can vary between fields used and do influence the decision to use the land in one way or the other. Using the household as the unit of analysis it is also difficult to compare the results with a spatial analysis that uses grid cells or pixels as unit of analysis, which are also units of land. Therefore, in this analysis the field will serve as the unit of analysis. This enables us to use the physical characteristics of the site, together with the characteristics at the plot and household level (Figure 2.3), which are attached to the field level.

In the household analysis the assumption is made that the land use decision for a field is made independent from the land use on other fields of the same household. To test this assumption the standardised residuals of the models are regressed (linear regression) on variables containing the number of fields of the other land use types. For example, in case of the corn model these variables would contain the number of banana fields and the number of wet rice fields owned by the farmer. A significant relation of one of the variables with the residuals would suggest that the assumption might not hold. The residuals of the models for wet rice and banana did not have a significant relation with these variables, so the assumption holds for those models. The residuals of the corn model showed a significant (*p<0.05*) negative relation with the number of wet rice fields and banana fields. Using a somewhat richer specification of the model, by adding two variables, the relation with the number of banana fields turned out to be insignificant, though the relation with the number of wet rice field still appeared. So, evaluating the model in its current specification it seems that decision for corn is not taken completely independent from the decisions made for the other fields.

Enhanced spatial model

For the enhanced spatial dataset variables were derived that best represent the process identified by the factors that performed well in the household level model. This approach inherently assumes that the drivers at the household level correspond to the drivers at the watershed level. This assumption only holds when the same unit of analysis (resolution and extent) is used in both analyses, because otherwise scale dependencies (Walsh *et al*., 1999; Verburg and Chen, 2000) can play a role. The unit of analysis in the household analysis was chosen to be the field. In the spatial dataset used in this study the unit of analysis are grid cells of 50 by 50 m. That observation unit does not completely resemble the field of the household survey, since one field can be represented in the spatial data as several grid cells (in case of fields larger than 0.25 ha). This can cause spatial autocorrelation, because cells from the same field, which are neighbouring cells, have the same properties. Besides this, the probability for a field to be in the sample will be different for both datasets. Both effects might hamper a good comparison between the household and spatial models. The farmer's decision at the household level was made for the whole field, so this data representation suits the processes that caused the land use and will be applied to both datasets. So, ideally, fields should occur at most one time in the spatial dataset, like in the household dataset. Therefore, a sample of ten percent was drawn from the cells available in the grid, which approximates that every field occurs only once (at most) in the dataset and reduces spatial autocorrelation.

The enhanced spatial model is, like the mainstream geographic model, a logistic regression model in which the probability is estimated for a grid cell to have a land use type or not as a function of explanatory variables. A stepwise procedure was used (forward stepwise regression with probability levels of 0.01 for entry in the model and 0.02 for removal from the model) to select the relevant factors explaining land use.

2.3.2 Logistic regression

In all three models the dependent variable is binary, meaning that a certain land use type occurs at a certain location (value 1) or not (value 0). When the response variable is binary, a good way to describe the shape of the response function is a tilted S or its reverse. This response curve can be described mathematically by logistic response functions. Equation 2.1 is a linearised form of the logistic response function and is referred to as the logit response function (Neter *et al*., 1996).

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \qquad (2.1)$$

Where p is the probability for the occurrence, $ß_0$ is an intercept and $ß_n$ are regression coefficients to be estimated, and the X_n are a set of exogenous explanatory variables. The ratio *p/(1-p)* is called the odds, *log (p/(1-p))* is the log odds, also named 'logit'. The logit of Equation 2.1 can be converted to an expression for the odds and to an expression for the probability, but those are three different ways of expressing the same thing (Menard, 2001). The interpretation of the parameters $ß_n$ is facilitated by the odds ratio *(exp(ß_i))*. The odds ratio can be interpreted as the change in odds for the considered event upon an increase of one unit in the corresponding factor, while the other factors are considered to be unchanged. This means that the odds *(p/p-1)* are multiplied by $exp(ß_i)$ for every unit increase of the variable (Neter *et al*., 1996).
To estimate the relative contribution of different variables within a model a standardized logistic regression coefficient was used (Equation 2.2) (Menard, 2001).

$$b^*_{YX} = (b_{YX})(s_X)/\sqrt{s^2_{\log it(\hat{Y})}/R^2} = (b_{YX})(s_X)R/s_{\log it(\hat{Y})} \quad (2.2)$$

Where b^*_{YX} is the standardised logistic regression coefficient, b_{XY} is the unstandardised regression coefficient, s_X is the standard deviation of the independent variable X, $s^2_{logit(\hat{Y})}$ is the variance of logit($\hat{Y}$), $s_{logit(\hat{Y})}$ is the standard deviation of logit($\hat{Y}$), and R^2 is the coefficient of determination.
To indicate goodness-of-fit, the R^2 measure used in OLS regression cannot be applied in logistic regression. There are pseudo- R^2 measures available for logistic regression, but those can only be used to compare different specifications of the same model and can not be used to compare different models. Therefore, the ROC (Relative Operating Characteristic) (Swets, 1988) was used to indicate the goodness-of-fit of the models. This measure is capable of assessing the quality of the predictor and can be compared between different models. The ROC summarises the performance of a logistic regression model over a range of cut-off values by classifying the probabilities. The value of the ROC is defined as the area under the curve linking the relation between the proportion of true positives versus the proportion of false positives for an infinite number of cut-off values. The ROC statistic varies between 0.5 (completely random) and 1 (perfect discrimination).

2.4 Results

2.4.1 Models based on mainstream geographic approach

The logistic regression models of the mainstream approach are shown in Table 2.3. The stepwise procedure selected 4 (*corn*), 2 (*wet rice*) and 3 (*banana*) variables that have a significant contribution to the models. Distance measures turned out to be explanatory in all three models. *Distance to village* and *distance to road* are contributing significantly to all three models and *distance to market* to the corn model. Since these variables are highly correlated with *population pressure* this variable did not appear in any of the models. *Distance to river* also does not appear in any of the models. *Slope* appears in the mainstream geographic models for corn and banana. More detailed interpretations of these models are discussed in Section 5 when the models are compared with the results of the spatial models. The ROC values for these models are 0.77, 0.73 and 0.70 for respectively corn, wet rice and banana. No collinearity between the independent variables was found.

Table 2.3: Results of the mainstream geographic models

Variables	b	s.e.	sig.	b*	exp(b)
Corn					
Slope	-0.0876	0.0066	0.000	-1.342	0.9161
Dist. to market	-0.0001	0.0000	0.000	-1.111	0.9999
Dist. to road	-0.0003	0.0001	0.000	-1.041	0.9997
Dist. to village	-0.0004	0.0001	0.000	-1.020	0.9996
Constant	0.9631	0.0773	0.000		
ROC	0.7650	0.0058	0.000	(0.754-0.777)*	
Wet rice					
Dist. to road	-0.0010	0.0002	0.000	-9.926	0.9990
Dist. to village	-0.0007	0.0002	0.000	-5.291	0.9993
Constant	-2.8142	0.1518	0.000		
ROC	0.7340	0.0170	0.000	(0.700-0.767)*	
Banana					
Dist. to village	0.0008	0.0001	0.000	1.008	1.0008
Slope	0.0547	0.0066	0.000	0.529	1.0562
Dist. to road	0.0006	0.0001	0.000	0.508	1.0006
Constant	-2.2473	0.0758	0.000		
ROC	0.7030	0.0074	0.000	(0.689-0.718)*	

** 95% confidence interval*

2.4.2 Household models

The result of the household analysis is presented in Table 2.4. For all land use types a clear relation with slope can be observed, which is in line with the hypotheses. The variables *steep slope* and *moderate to steep slope* have a negative effect on the probability for a field to have corn. The model for wet rice shows a positive relation with *flat slope* and *flat to moderate slope*. On the contrary banana has a positive relation with steep and a negative relation with flat slopes. So, slope is a good determinant to make the distinction between corn and wet rice on the flatter fields and banana on the steeper fields. An explanation for this is that for corn and wet rice regular tillage is necessary and for wet rice irrigation has to be applied. Both activities can be best performed in the flatter areas. The remaining steep parts are cultivated with banana in which no regular tillage is applied.
The different preferences of ethnic groups turned out to be significant as well as the variable *place of birth*, which is partly related with the ethnic groups, because some groups have a distinct migration period. In general, the Ibanag have a tradition in growing corn and are less focussed on rice cultivation. The Tagalog speaking people, the Ifugaos and also the Ilocano people have a tradition in rice cultivation. In the corn model a positive relation was found with the Ibanag people and a negative relation with the Ifugao people. In the wet rice model a positive relation with the Ifugao and Ilocano people was found. But, besides ethnicity also the variable *place of birth* turned out to contribute significantly to the model. Being born in San Mariano has a positive effect on the probability of corn and a negative

Table 2.4: Results of the household models

Variables	b	s.e.	sig.	b*	exp(b)
Yellow corn					
Moderate-steep slope	-9.639	15.292	0.528	-4.709	6.515E-05
Steep slope	-9.608	41.932	0.819	-1.970	6.725E-05
Ethn. Ibanag female	1.782	0.774	0.021	1.286	5.943
Ethn. Ifugao male	-1.731	0.700	0.013	-1.194	0.177
Creek	-0.995	0.386	0.010	-0.979	0.370
Place of birth male	0.913	0.389	0.019	0.910	2.492
Moderate slope	-0.848	0.399	0.033	-0.767	0.428
Constant	0.909	0.378	0.016		
ROC	0.839	0.029	0.000	(0.782-0.896)*	
Wet rice					
Flat slope	5.590	1.387	0.000	267.686	8.430
Ethn. Ifugao male	4.601	1.414	0.001	99.614	4.869
Flat-moderate slope	2.957	1.431	0.039	19.241	3.742
Creek	2.199	0.698	0.002	9.019	3.323
Ethn. Ilocano male	2.114	1.118	0.059	8.284	3.223
Place of birth male	-1.353	0.703	0.054	0.258	-2.069
Constant	-9.474	1.942	0.000		
ROC	0.922	0.027	0.000	(0.869-0.975)*	
Banana					
Flat slope	-9.412	28.659	0.743	-11.379	8.175E-05
Steep slope	13.874	189.426	0.942	3.499	1.066E+06
Moderate-steep slope	5.506	1.426	0.000	3.308	246.119
Moderate slope	2.046	0.639	0.001	2.274	7.737
1st year on plot	-0.084	0.026	0.001	-2.200	0.920
Plot size	-0.080	0.037	0.031	-1.302	0.924
Constant	164.069	51.379	0.001		
ROC	0.924	0.021	0.000	(0.883-0.964)*	

* 95% confidence interval

effect on the probability of wet rice. This can be caused by the fact that newcomers are mainly people that grow rice traditionally. Another explanation for this could be that new migrants will focus primarily on subsistence and therefore cultivate rice, which is the main staple crop.

In the model for wet rice the variable *creek* has a significant positive contribution to the probability. In the corn model this relation is negative, where no relation was expected. The presence of a creek on a plot is important to grow wet rice, because this crop is irrigated. At these locations near corn is out-competed by rice. Corn is mainly grown on the large flat terraces near the big rivers and is primarily dependent on rain. These plots are not all connected with the big rivers, but even if they would be close to the big river natural irrigation would not be possible because of the height difference. So, the plots near the river cannot be used for irrigated rice.

Besides the strong relation with slope banana cultivation has a negative relation with the variable *'first year the farmer started on the plot'*, which means that recently acquired plots are often not cultivated with banana, but with other crops like corn and rice. Banana is negatively related with *plot size*. This is remarkable, because the hypothesis was that small field are cultivated with corn and rice and bananas are cultivated extensively on the larger plots. A possible explanation is that in the limited capacity for farmers to transport bananas from their fields leading to acreages planted with banana that do not exceed the transportation capacity.
The variables included in the models can be explained in how they affect land use decisions. On the other hand some variables that were hypothesised to be important to land use were not included in any of the models. Especially transportation costs and tenure were expected to be important. Transportation costs were hypothesised to influence the decisions for cash crops like yellow corn. Another strong relationship that was not included in these models is the positive relation between tenancy and corn.
The independent variables of the three models were investigated on any correlation to see if there is any collinearity within the models. This resulted in excluding *transportation cost* in the wet rice model, because this variable was highly correlated with *ethnicity Ifugao*. The other variables did not show high correlations among each other.
The ROC values that are used to give an indication for the goodness-of-fit are high for all three models, but especially for the land use types wet rice (0.92) and banana (0.92). The ROC for the corn model is 0.84. The high ROCs indicate that the variables used in the model can very well predict the occurrence of the land use types corn, wet rice and banana at the household level.

2.4.3 Enhanced spatial models

The enhanced spatial models are presented in Table 2.5 and the results are visualised in Figure 2.4. This figure shows the actual land use pattern (left) and the predicted probability (right). Especially for land use types corn and banana the predicted probabilities seem to fit the observed data well. For wet rice, the visual interpretation of the goodness-of-fit is less easy, because there are not many rice fields present in the area. The ROC values for the spatial models are 0.78, 0.76 and 0.74 for respectively the corn, wet rice and banana model.
The spatial models, incorporating the improved variables, have characteristics of both the household and mainstream geographic models (Table 2.5). The variable *improved distance to village* appears in the models for wet rice and banana. Wet rice is an intensive crop and is cultivated close to the villages and banana is situated further away. So, people grow the intensive crops close to their houses, or the other way around, people have settled where the conditions for growing their main crops is best. The variable *improved distance to market* wet is negatively related with corn. The (time) distance to market represents the costs to transport in- and outputs to the market and the negative relation indicates that the crop is grown at places with low transportation cost, generating the largest profits. The variable *distance to small river* turned out to be positively correlated with corn and negatively with banana. Only the small rivers are used for irrigation, since hardly any technology is used to use the water from the big rivers. Therefore, the variable *distance to small river* was created to approximate suitable locations for irrigation. *Slope* was selected in the corn and in the banana model and the relations are as hypothesised: negative for corn and wet rice

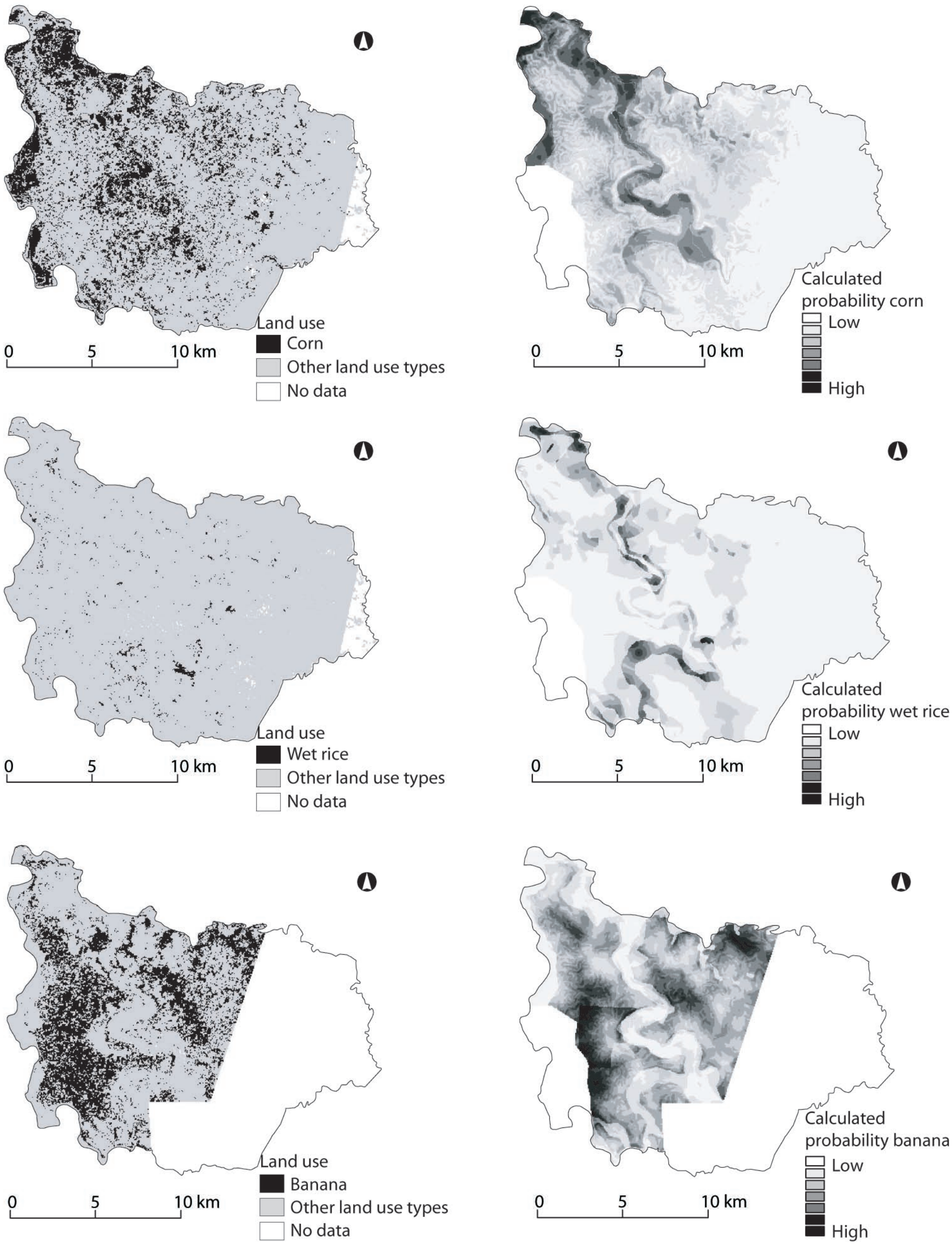


Figure 2.4: Actual land use (left) and predicted probabilities of the enhanced spatial models (right) for the three land use types corn (top), wet rice (centre) and banana (bottom)

and positive for banana. Geomorphological variables appear in all spatial models. These variables are not directly related to a process, but were introduced to approximate environmental characteristics like soil fertility and suitable landscape properties.

Table 2.5: Results of the enhanced spatial models

Variables	b	s.e.	sig.	b*	exp(b)
Corn					
Impr. dist. to market wet	-1.294E-04	6.236E-05	0.000	-2.142	1.000
Slope	-0.073	0.007	0.000	-1.099	0.930
Geo. active floodplain	1.047	0.100	0.000	0.608	2.848
Geo. terraces	0.732	0.077	0.000	0.576	2.080
Dist. to small river	4.644E-04	1.064E-04	0.000	0.320	1.001
Project ISF	-0.418	0.154	0.007	-0.194	0.658
Constant	0.213	0.094	0.024		
ROC	0.775	0.005	0.000	(0.763-0.786)*	
Wet rice					
Impr. dist. to village	-3.014E-04	5.365E-05	0.000	-11.191	1.000
Org. municipality	-2.462	0.618	0.000	-4.219	0.085
Tax declaration	0.013	0.004	0.001	3.255	1.013
Geo. terraces	0.996	0.184	0.000	2.657	2.709
Project ISF	0.904	0.340	0.008	1.420	2.470
Constant	-2.771	0.318	0.000		
ROC	0.759	0.018	0.000	(0.724-0.795)*	
Banana					
Impr. dist. to village	3.384E-04	2.145E-05	0.000	1.087	1.000
Geo. marine sed.	1.078	0.104	0.000	1.039	2.938
Dist. to small river	-0.001	0.000	0.000	-0.483	0.999
Project IFMA	0.864	0.124	0.000	0.444	2.373
Slope	0.038	0.007	0.000	0.368	1.038
Geo. active floodplain	-0.576	0.192	0.003	-0.358	0.562
Geo. limestone	1.717	0.360	0.000	0.298	5.567
Geo. rock	1.338	0.292	0.000	0.295	3.810
Constant	-2.408	0.118	0.000		
ROC	0.738	0.007	0.000	(0.725-0.752)*	

** 95% confidence interval*

In general the active floodplain and terraces are flat and have better conditions for crop production. On these sites the intensive crops, yellow corn and wet rice, are cultivated. The less favourable areas are left for banana, which is cultivated in an extensive manner in this region. The variable indicating place of birth (*original municipality*), which was introduced based upon the experiences with the household level analysis, appeared to be significant in the model for wet rice. The relation is the same as in the household model. The variable *tax declaration* also turned out to be relevant in the wet rice model. The positive relation suggests means that wet rice is cultivated more on plots with a relatively strong tenural security. The positive relations with forest policies in the wet rice (*project ISF*) and banana model (*project IFMA*) are the contrary to what was hypothesised. It was expected that all agricultural land use types would occur less or equal at locations designated for forest policies compared to other areas, because the forest policies are mainly focussed on reforestation and agroforestry. Though, the ISF program aims at developing only 20 percent of the area with forest trees and the remaining 80 percent with agroforestry and it aims at increasing upland production (Balagtas-Mangabat, 2002). So, the higher probability of wet rice in the ISF area could be well caused by the program. The relation between banana and IFMA could be caused by misclassification of the remote sensing images, since banana and secondary forest or reforestation are difficult to distinguish from each other.

The variables selected initially by the stepwise procedure were checked on collinearity. This resulted in excluding *improved distance to village* from the corn model (because of correlation with the variable *improved distance to market*) and *ethnicity Ibanag* from the banana model (correlated with *Geo. Rock*).

In contrast with the household models the spatial models do not incorporate any ethnicity variables, but do incorporate accessibility measures. These differences are discussed in the following section.

2.5 Discussion and conclusions

The mainstream geographic models explaining LUCC are based on readily available biophysical and geographical data, or on data that are easy to calculate from basic data, like distance measures. From the results of the mainstream geographic model in this study it could be concluded that the location of the land use types yellow corn, wet rice and banana are primarily determined by distance to village, road and market, and slope. These are location specific measures, which are basically physical characteristics. Though, the distance measures can be proxies for other process related variables like transportation cost and travel time to the field.

The household models show somewhat different results. From the household analysis it is clear that variables from all three levels present in the household survey (household, plot and field) play a significant role. Biophysical characteristics of the field, like slope and the presence of a creek, as well as social-economic characteristics of the household, like ethnicity and place of birth, are important. This advocates the incorporation of household characteristics in the spatial model, in addition to the geographical characteristics. The household model also shows the relevance of using the field as the unit of analysis in the household study. By taking the smallest unit as the level of analysis and linking the higher-level characteristics to this level the biophysical and socio-economic variables can be incorporated jointly in the statistical analysis. The result shows that both types of variables play a role in explaining the occurring land use.

The spatial model presented in this chapter combines the best of both. The dataset for this analysis was a combination of newly specified spatial variables that best represent variables that were significant in the household model and improved spatial variables complemented with spatial variables that are considered to have explanatory power at the watershed level. In these models a combination of accessibility measures, social and physical variables turned out to be the explanatory factors for land use. The main differences between the household models and the spatial models are that (1) the variables that are representing the same do not always have the same relative importance (indicated by the standardised b (b^*) in the models) and (2) that the goodness-of-fit (indicated by the ROC value) of the spatial models is substantially lower than the household models. The

following paragraphs will discuss the causes for these differences.
The differences in the relative importance of the parameters between the household and the spatial models could be caused by scale effects through differences in resolution or extent. The resolution of both analyses, however, is practically the same. The fields of the farmers in the area are generally between 0.25 and 3 ha. In the household study the field was taken as the unit of analysis and all fields are represented once. But, in the spatial dataset the larger fields consist of a number of cells, because in the spatial analysis cells of 50 by 50 m (0.25 ha) are used. By taking a ten percent sample from the data this was overcome. So, there is no difference in resolution between both datasets. They are both a sample of fields. Therefore, scale dependencies as a result of resolution, as mentioned by Walsh *et al*. (1999) and Verburg and Chen (2000), are not likely to occur. Though, there are differences in extent. In both analyses the study area consists of the same 13 *barangays* in San Mariano. However, the household data only represent the land that is occupied by farmers and the spatial data consist of the whole area, including forest and other land that is not occupied. This is the reason why certain variables, like accessibility and geomorphology, which can make the distinction between these land use types and the agricultural land use types, are relativity more important in the spatial analysis.
Differences can also occur because the two approaches use different sampling techniques. The household analysis uses a random draw from the households in the area, while the spatial approach draws randomly from the grid cells in the study area. This could be a possible explanation for the effect that different variables are found to be important in the two approaches.
Another important source of differences between the spatial model and the household model are the different ways of data representation. The two analyses are based on different datasets that were collected at different organisational levels. No household data is used in the spatial analysis, so aggregating errors can hardly occur. The information extracted from the household models is used as an indication what variables might be important at the watershed level. This study tried to represent household level variables at the watershed level in order to 'pixelise the social'. Social variables and/or processes and household characteristics are captured in maps. Because of these two different levels of organisation it is not always possible to represent the data in the same way.
A group of variables, like typical household characteristics such as ethnicity, can be mapped potentially if one would know what field on the map is used/owned by whom. This approach has recently been adopted by some authors (e.g. Vance and Geoghegan, 2002; Walsh *et al*., 2003). In this study the aim is to be spatially explicit in the whole study area, meaning that all land parcels of all people in the area should be mapped. This turned out to be impossible in this study due to its size and consequently time and financial constraints. Therefore, more aggregated variables based on census information were created to proxy the variable. For example, the ethnicity of the field's owner was not recorded for every field, but instead a value indicating population pressure was calculated per ethnic group for each pixel.
Another group of variables included in the household model are based on farmers' perceptions. As indicated before, the fields in the household analysis were not mapped. But instead, the field characteristics are collected through questioning the respondents. The approaches that do map fields can use this information to attach other mapped data to this field. In this study mapped data are only used in the two spatial approaches. The result of this approach is that in the household survey the variables are values as perceived by

the farmers in contrast to more objective sources of data. This has to be taken into account while interpreting the data. An advantage of this approach is that it saves time, because not all fields have to be visited. So, in the household survey the answers of the farmer are perceived values and therefore relative to the knowledge of that farmer. The scale used to rank certain variables varies between the farmers. A variable included in the spatial model, from more objectified sources, is based on only one scale. For example, let us assume that the most fertile soil of a farmer in the mountains is less fertile than the most fertile soil of a farmer cultivating in the lowlands and that the farmers' strategy is that they all grow corn on their most fertile soil. This would mean that soil fertility is only explanatory at the household level, but not at the spatial level, causing a difference in outcome of the two models.

Closely related to this are variables that are approximated by the respondents, like the question: 'how long do you travel from your house the this plot?' A calculated map will be more objective answer to this question than the estimation of many respondents. On the other hand, the calculated map is also based on a number of assumptions (e.g. the travel speed on all parts with the same slope is the same), which possibly does not represent the conditions that are perceived by the households.

Some variables are just difficult to obtain at the household level, because the respondents do not have any knowledge of the subject, like geomorphological units or governmental policies, whereas they might indirectly respond to these factors.

These different data representations and the difference in extent are causing the different behaviour of the variables in the household model and in the enhanced spatial model. They can even cause variables to occur in one model, but not in the other. Besides this it is widely acknowledged in literature that higher-level processes cannot be represented as simply the sum of lower level characteristics (Coleman, 1990). Despite our efforts to keep the extent and the resolution of the household and spatial models the same these effects can cause differences between the two approaches.

The differences in extent and the aggregated representation of some variables are also likely to cause the lower explanatory capacity (as indicated by the ROC) in the spatial models compared with the household models. The larger extent of the spatial model means that the spatial model has also to differentiate between the land use types corn, wet rice and banana and the areas that are not in use, whereas the household model has only to distinguish between corn, wet rice and banana. This can explain that the explanatory power of a variable is less in the spatial model compared to the household model. An alternative model specification, which would be closer to the specification of the household model, is to exclude the area that are not in use or owned by any farmer. Though, the interest in this study is in the whole area, so this alternative was not applied.

Furthermore, the spatial variables derived from aggregated data cannot completely represent the variability as perceived by the individual households. Information is lost in comparison with household level data, which will lead to a lower goodness-of-fit. Less accurate variables, like the ethnicity measures, tend to be less important in the spatial model and contribute less to the ROC.

The difference between the spatial model and the mainstream geographic model is clear. The mainstream geographic model is built out of readily available topographic and census data, like the approach followed in most regional level studies. The data was processed into variables that are mainly proxies for the processes at hand and give no insight in the processes that determine the land use. The enhanced spatial model does provide insight into those processes, because the variables used in the spatial model are constructed to represent a process with roots in the household model, which is the level where land use decisions are taken. These processes are described in Section 2.4 and will be described in greater detail in Chapter 3. The variables included in the geographical model are much more abstract and do not represent a specific process. However, the overall fit of the enhanced spatial model is only slightly higher than the mainstream model. ROCs for corn, rice and banana are 0.77, 0.73 and 0.70 for the mainstream model and 0.78, 0.76 and 0.74 for the enhanced spatial model. Based on the 95 % confidence interval the ROC values of the two corn and the two rice models are not significantly different. The ROCs of the banana models do show a significant difference. The added value of a spatial model compared to the household model is that it allows the inclusion of spatially relevant variables that provide the context for household level processes. This way we have included social processes in spatial models ('pixelising the social') and added spatial dimension to household level decision-making ('socialising the pixel').

This study can make a significant contribution to empirical land use change studies. The approach, which jointly analyses the household and watershed level, can serve as a connection between spatial models at a broader scale and more social research aimed at the explanation of the causal relations that drive land use change. In that respect the findings of this research can be a basis for spatial statistical models (e.g. Verburg *et al*., 2002) as well as more actor-based approaches studying the farmers' decision process (e.g. De Groot, 1992; Parker *et al*., 2003). Basically, the former will give insight in the dimensions and locations of the land use change and the latter can provide insights in how to influence the process, which can be relevant in policy-making.

3

Comparing inductive and deductive modelling of land use decisions: Principles, a model and an illustration from the Philippines

Abstract

Understanding the causes of land use change is of great importance for issues of tropical deforestation, agricultural development and biodiversity conservation. Many quantitative studies, therefore, aim to link land use change to its causal 'driving forces'. The epistemology of virtually all these studies is inductive, searching for correlations within relatively large, sometimes spatially explicit, datasets. This can be sound science but we here aim to exemplify that there is also scope for more deductive approaches that test a pre-defined explanatory theory. The chapter first introduces the principles and merits of inductive and more deductive types of land use modelling. It then presents one integrated causal model that is subsequently specified to predict land use in an area in northeastern Philippines in a deductive manner, and tested against the observed land use in that area. The same set of land use data is also used in an inductive (multinomial regression) approach.

With a goodness-of-prediction of 70 percent of the deductive model and a goodness-of-fit of 77 percent of the inductive model, both perform almost equally well, statistically. Because the deductive model explicitly contains not only the causal factors but also the causal mechanisms that explain land use, the deductive model then provides a more truly causal, as well as more theory-connected, understanding of land use. This provides land use scholarship with an invitation to add some more deductive (theory-driven and theory-building) daring to its methodological repertoire.

Based on: Overmars, K.P., De Groot, W.T., Huigen, M.G.A. 2006. Comparing inductive and deductive modelling of land use decisions: Principles, a model and an illustration from the Philippines. Human Ecology (Accepted).

3.1 Introduction

The face of the earth is rapidly changing, with great consequences for rural livelihoods, biodiversity conservation, urban quality of life and the global climate. Understanding land use change is therefore a matter of obvious import and urgency, reflected, inter alia, in LUCC, the joint land use program of the International Geosphere-Biosphere Programme (IGBP) and the International Human Dimensions Programme on Global Environmental Change (IHDP) (Turner *et al*., 1995).

Land use change is the result of the often complex interplay of underlying causal factors, usually referred to as 'driving forces', that may vary across scales and organizational levels, that may work directly or through longer causal routes and that may be associated with quite different societal and scientific realms, such as markets, policies, demography, culture and biophysical factors. How can such complexity be handled scientifically? One approach is to focus on only one or a few factors, and accept that explanations can only be partial. Generally, however, land use scientists desire to do a more 'integrated' (multi-factor) analysis. As shown, for instance, in the overviews of Walker *et al*. (2002) and Verburg *et al*. (2004d), the great majority of the present-day blooming of quantitative integrated (multi-factor) studies of land use change follows an inductive approach, sometimes guided by theory but without testing the theory as such. In the present chapter, we make a case that the present state of the art allows to perform integrated research and yet use a more deductive epistemology, and that this option, in interaction with inductive work, will enhance causal insight and cumulative scientific progress in land use science. We aim to strengthen our case by showing and discussing the performance of a deductive and an inductive approach, applied parallel to each other to explain the land use in a single example region.

The chapter is organized as follows. The following section discusses the principles and merits of inductive and deductive approaches to land use science. Since deductive work requires a theoretical model to be tested, the third section is devoted to the structure of the model for our case study. The material and methods section then introduces the study area and the data gathering methods. The fifth section formalizes the deductive model for our example region and the sixth quantifies the model. The seventh section describes the results of the deductive model as well as those of a multinomial regression model, which was used for the inductive modelling exercise. The ensuing discussion shows the value of the deductive modelling approach within a range of approaches from inductive to deductive.

3.2 Inductive versus deductive modelling

3.2.1 Deductive and inductive epistemology in integrated land use explanations

For most of us, the "empirical cycle" must have been the first concept taught in lectures about how science proceeds. First there is a theory; then a concrete predictive hypothesis is deducted from that theory. Then this hypothesis is tested in the real world and with that result, the theory is either falsified or strengthened. This, in short, is deductive epistemology. Contrasting with this approach, inductive methodology works the other way around. It begins with observations of reality and then tries to find regularities in these data. This regularity is then declared to be a general pattern (a model, a theory). This claim can be based, for instance, on the randomness of the sampling that was used. Statistical work can be part of both approaches; statistical testing is a characteristic part of deductive methodology, while multiple regression is often used for inductive approaches.

The present chapter does not find fault in this basic epistemological scheme. We do, however, think that for a proper understanding of how land use science proceeds in practice, it is necessary to define a number of methodological positions that lie in-between the deductive and inductive extremes. In order to arrive there, it helps to first specify what exactly is 'a model' or 'a theory', especially in terms of what may be called the degree of specification of that model or theory. A first case is that a researcher has no model or theory at all. Obviously then, the only methodology available is extreme induction, or data mining as it will be called below, in which the researcher attacks large datasets, basically 'correlating everything with everything else' in order to see if any patterns may be found. A second case is that the researcher has a notion of what factors may be relevant for the explanation of what he seeks to explain. This may be called a weak model. The researcher may then still fall back on data mining approaches, but he may also concentrate on the candidate factors to see if these indeed play a role as was hypothesized. A third case is when the researcher avails of what may be called a strong or structured model, that not only states what factors are important but also how they are important. According to Elster (1989), it is only then that true explanation comes within reach, because true explanation requires insight not only in the factors but also in the mechanisms. An example of a structured model is the well-known law of Liebig on plant growth that specifies not only that nutrients like nitrogen and phosphorus are important but also how they are, with the plant responding only to the nutrient that is 'in the minimum'. In formula: *plant growth = a + b*MIN [phosphorus, c*nitrogen]*. Note here how different this formula is from the usual structure of multiple regression, which would be: *plant growth = a + b*phosphorus + c*nitrogen*. In the latter formula, adding more nitrogen would always result in more plant growth even if phosphorus would be at zero. In the first formula (and in reality), the plant does not respond at all. In the case of the researcher availing of a strong model, he can try to induce the parameters for his particular case in the model's structure. He may also fall back on a traditional multiple regression with the nutrients arranged in the additive structure, or even on blind data mining. See for instance De Groot *et al*. (1987) for an example of induction using both Liebig's structure and traditional regression on plant growth. A final case is when the researcher avails of a fully quantified model, e.g. Liebig's structure with the parameters a, b and c specified. It is only then that true prediction, hence true deduction, is possible.

We can now come to see the deduction/induction pair of terms as defining a gliding scale between two extremes. On the one hand, there is extreme deduction of the Popperian kind (Popper, 1963), in which the empirical cycle is followed strictly and theory falsification rather than verification is seen as the key to progress. On the other hand, there is extreme induction, in which the researcher aims to find patterns in large datasets without any theoretical guide. Both extremes have their advantages in some cases, e.g. if very strong theories are available, or if no theory at all is as yet defined, respectively. Both have strong disadvantages too, however. In the social and economic sciences, extreme deductivism would lead to an endless rejection of theories because simply none of them is able to grasp the full complexity of the system described. Extreme inductivism, on the other hand, leads to an immense amount of correlations that cannot be interpreted as causes and never accumulate into a coherent theory.

One response to this dilemma is the validation of inductive models following the suggestion made (in passing) in many statistical textbooks to split one's data in half, use the first half for a free induction of any kind, and then use the other half to test the induced model. An example is in Nelson *et al*. (2004), who use 1/25th of their large and spatially explicit dataset to induce their explanatory land use model and then use the model to predict land use over the whole map. A more radical way out of the dilemma has been suggested by Brox (1990), discussing the epistemological status of 'grand theories', in his example the common property theory applied to fisheries. Brox' solution is that we forget about the empirical claim of such theories at all but rather regard them as analytical tools. Using the theory we may discover which part of reality behaves according to the theory (which is interesting), and which part does not (which is interesting too).

In most research practice, researchers find a less daunting solution by seeking or simply adopting a position, usually implicit and led by disciplinary traditions, somewhere on the continuum between extreme induction and extreme deduction. For the present chapter and including the two extremes, we may define six of these positions. We concentrate here mainly on quantitative work.

1. *'Extreme induction'*. This is the extreme of data mining, "knowledge discovery in databases" (Liao, 2003).[1]
2. *'Unstructured factors induction'*. Under this term we subsume all research approaches that apply a broad conceptual framework of some kind, usually derived from common sense or literature overview, in order to specify a usually long list of factors (roads density, slope, off-farm income, tenure security, distance to recreation sites, household composition and so on, often each with several variants of further specification and measurement) that are candidate to help explain land use or land use change. (Alternatively, some kind of theory may be invoked as well, e.g. as Nelson *et al*. (2004) do, saying that land users choose for the most profitable land use, but then these theories are in fact only serving as a broad conceptual framework.) Often, these factors are proxies of the actual factors that influence the land use process, since the processes themselves are not specified. The studies then leave it to the procedures of statistical inference to find the correlations between these variables.[2] Characteristically, these studies do not end with a discussion of theoretical perspectives but only with a discussion of the significance of correlation coefficients and suchlike in the specific case studied. Many land use change studies fall into this category (e.g. Geoghegan *et al*., 2001; Serneels and Lambin, 2001; Overmars and Verburg, 2005 (Chapter 2)).
3. *'Theory-guided factors induction'*. This term denotes all studies that take an explicit theory of land use change as point of departure to critically specify a theory-connected (and usually shorter) list of explanatory variables. Strictly speaking, this list is still unstructured; it is only a list, after all, without specification of how the

---

[1] *In quantitative research, this extreme is often, and understandably, seen as something to be done only very sparsely. In qualitative research, remarkably, extreme induction is often seen as the ideal basis for 'grounded' theory building (Glaser and Strauss, 1967), allowing respondents to speak in their own voice and analysing their visions without any preset notions of the researcher. Great progress has been achieved this way, e.g. the famous discovery of the 'ethics of care' (Gilligan, 1982).*

[2] *See Overmars and Verburg (2005) for a factors-led inductive study on the research area in the Philippines and Geist and Lambin (2002) for an inductive meta-analysis of 152 studies on tropical deforestation.*

variables are supposed to interact. On the other hand, the variables are not simply 'candidates' that are dropped if they do not contribute to the explanation. If they do not contribute, something is 'wrong' with the theory or its interpretation, which needs to be discussed. One quantitative example is in Perz and Walker (2002), focusing on secondary forest growth in Amazonia in connection with Chayanovian theory. Another example is by Rudel and Roper (1997) who arrived at their "frontier model" and "immiserization model" of tropical deforestation by a careful construction, examination and re-examination of a relatively small dataset rather than by blind force applied to a large one. Interesting results have also been reached in a more qualitative manner, exemplified by Ostrom (1990) who arrived at her well-known conditions for successful common property management by a stepwise induction of case studies. Characteristic for all studies of theory-guided induction is that the relevance of the results is wider then those of type 1 and type 2 studies. Guided by theory, induction can become theory building.

4. *'Imposed theory structure'.* The next rung on the induction/deduction ladder is formed by studies that impose not only theory-guided factors but also a theory-guided structure (the 'behavioural statements', as Walker (2004) says) on reality before multiple regression is applied in order to induce the parameters within that structure. If our theory would be, for example, that people only choose for a land use type to the extent that this land use type is both culturally appropriate and profitable, our model structure would look like Liebig's law, e.g. that the land use depends on $ß_0 + ß_1 *MIN [ß_2 *CULT, PROF]$. In the same vein Tadepally (1999) stated that in order to rehabilitate their village-level irrigation systems, villages should avail of both the capacity (specified by Tadepally as collective social capital) and the motivation to do so (specified by Tadepally as low rehabilitation cost), and found a strong relationship between these two variables and the success of NGO intervention for rehabilitation, with an imposed structure of $SUCCESS = ß_0 + ß_1 *MIN [ß_2 *CAP, MOT]$. (It is interesting to note that the 'imposed structure' approach can also be used in a more qualitative style. We then use a theory to 'tell the story' of a specific case of land use change as do, for instance, Walker and Solecki (2004) and De Groot (1999) who apply dynamic versions of Thünian theory to tell the land use history of the Everglades and of the Cagayan Valley in the Philippines, respectively. If the story is good, or at least significantly more insightful than others, this is a test that reality indeed works as the theory prescribes. This test will always remain soft, however, since qualitative theories and stories will always be quite malleable in the hands of good storytellers.)
5. *'Imposed theory'.* A purely deductive approach is reached when a land use theory is specified for a real world case in terms of both structure and parameters, and the land use thus predicted is tested against real land use. As an example, in the case study presented in this chapter we will develop a theory-based model structure, quantify it and then test it on a dataset from the Philippines.
6. *'Extreme deduction'.* We keep the 'Popperian' extreme separate here because in step 5, the model and the data gathering are not geared towards falsification and neither need theories to be dropped if they do not work adequately yet.

A few technical remarks are in order here. First, induction, deduction and the continuum between them, even though central tenets of epistemology, do not cover the full spectrum of scientific methodology. Creative inference ('abduction') and the heuristic concepts of 'event ecology' (Vayda and Walters, 1999) are cases in point. Second, we may note that the six rungs of the induction/deduction ladder are naturally not the only possible ones. Researchers may also find intermediate and mixed positions, or work sequentially, with more or less extreme induction generating patterns that may be later used for a more deductive approach or, the other way around, starting from a theory. We do not go into these issues here, however, and regard the listing as good enough to indicate what we mean when saying that land use studies could or should become 'more deductive'.

3.2.2 Could land use explanation studies become more deductive?

Overlooking the field of explanatory land use studies, we find a quite skewed distribution over the induction-to-deduction axis. Examples abound of unstructured factors induction. Theory-guided factors induction is present in much smaller numbers. Imposing of theory structure is virtually non-existent. This may have historical and cultural backgrounds. To begin with, strong theories that may be tested are simply not massively present in any young science field. Furthermore, theories and deduction are not really *en vogue* in post-modern times (they are top-down, they turn a blind eye to the multiple complexities and voices of social realities, etc.). And finally, the attraction that land use studies appear to have had to econometrists and GIS-based geographic data technology may have had a flipside too, namely to block growth towards more deductive, theory-guided work.

In our opinion, explanatory land use studies could become more deductive. We do have land use theories to use and test, if only simple. Examples are Neo-Malthusian theory speaking about poverty traps, neo-Boserupian theory speaking about the positive effects of population on land use sustainability and neo-Thünian theory speaking about moving frontiers and urban markets (coupled as in De Groot (1999), or decoupled as in Walker (2004)). And we have more general explanatory theories waiting to be applied and tested on land use situations, such as rational choice theory, cultural theory, theories of collective action and common property management, and so on.[3] Furthermore, much knowledge has accumulated and great datasets have been collected – knowledge and datasets that do not need to be used one-way but may also be re-used in more deductive manners. In the present chapter, our example shows that nothing difficult is at stake here, conceptually.

3.2.3 Should land use explanation studies become more deductive?

Should explanatory land use studies try to move up a bit on the deduction ladder? There are two main advantages of using deductive methods. First, deduction yields the intrinsically better proof of causality, *i.e.* true explanation. Let us take Nelson *et al*. (2004) as an example. Their causal model is that on each site, the most profitable crop is grown. However, this is not tested as such because, as Nelson *et al*. say, profitability is not measurable. Instead, factors such as land slope are used as independent variables. Slopes, obviously, do have an influence on profitability but they causally underlie many other values too. If, say, maize is found to be associated with medium slopes, would that be because of its relative profitability there? It could also be that traditions do not allow maize elsewhere, or because of risk

[3] *We disagree here with Couclelis (in Parker et al., 2002: 6), who rightfully states that predictive models should be structurally appropriate, but then adds that the process theories (i.e. theory specifying causal mechanisms of these model structures) are simply not available in the land use field.*

aversion, or because maize from these soils tastes best. The slopes/maize correlation does not establish any causality. If, however, Nelson *et al*. would have *calculated* the spatially explicit profitability of crops (based on prices, distance to road, slope etc.), then would have set the model to predict crop distribution on the basis of highest profitability and then would have found the model predicting maize on medium slopes, a strong proof of causality would have been delivered. In the words of Elster (1989), this is because not only the causal factor but also the causal mechanism (in this case, profitability) has been tested. The second benefit of a more deductive approach is that it better facilitates the accumulation of insight on the level of the discipline as a whole. Referring back again to the example of Nelson *et al*. (2004), the outcomes of type 2 studies are basically the strengths-of-correlation between land use and the usually long list of independent variables such as slopes, prices, household composition variables and so on. Conclusions then necessarily tend to remain largely stuck on that level, e.g. that maize tends to converge on certain slopes in this case, or that the number of children does not appear to have an influence in that case. In order to reach some degree of generalization, such studies then have to wait until enough of them have accumulated to themselves become data in a meta-analysis such as that of Geist and Lambin (2002) who, characteristically for an inductive approach in the meta-analysis of inductive studies, come up with a generalized and regionally patterned listing of proximate factors and underlying driving forces of tropical deforestation. Obviously useful as this may be, more progress would be made if not only the incidental meta-analyst but also the researchers themselves, in their own studies, would be able to participate in a permanent intertrade of generalization. This can be achieved if these studies would be more deductive, *i.e.* more theory-led. That way, all land use scientists could contribute to progress around a relatively limited number of theoretical themes in stead of only adding more detail about a very large number of empirical variables. Theory-led work, feeding back into theory, leads to theory building.

In all this, we assume that empirically based theories are good to have. In other words, we assume that land use scientists do not become addicted to theories, especially their own, to a degree that theories begin to block entry for the surprises of reality (Vayda, 1983) or become objects of counterproductive controversy (Brox, 1990).

3.2.4 Model choice and levels-of-deductivism of this chapter

As said, the objective of the present chapter is to expound and illustrate the deduction/induction dimension for integrated land use explanation. For the deductive part, we have therefore chosen to test a broad model that is able to take up all factors that should be comprised in a truly integrated approach, hence including cultural, economic and biophysical data. It does have to be a model, however, and not some underspecified agglomerate. For this deductive 'core structure', we have chosen for the Action-in-Context framework of De Groot (1992), which may be characterized as broad rational choice. For the inductive approach a multinomial logistic regression model is applied. As for the positions on the deduction ladder, we have chosen to compare an 'unstructured factors induction' (rung 2) with a fully deductive approach, 'imposed theory' (rung 5), hence a true test without any subsequent fitting on the dataset. (In the remainder of this chapter these two approaches are referred to as the inductive and deductive approach or model, respectively). Logically too, we put all emphasis on the comparison and not on the cultural or land use intricacies of the study area.

3.3 The Action-in-context framework and decision model

Action-in-Context (AiC) (De Groot, 1992; Verburg *et al*., 2003) is a framework designed for the explanation of human actions, especially in the environmental field. Based on the concept of progressive contextualization (Vayda, 1983), the idea of AiC is to start out from the action to be explained, then identify the (individual or collective) actors directly causing this action, then identify the range of options available to these 'primary' actors and the motivations attached to these options, and then identify other ('secondary') actors and factors influencing these options and motivations, thereby putting the action in its relevant causal context without *a priori* bias towards any scientific discipline (Vayda and Walters, 1999). With that, AiC is a fully actor-based framework, which is a logical choice for explanatory work because actors, not systems, are the social entities that cause change directly.[4] AiC may be used as a framework to guide the research process, but can also be used as a template for models. These models can be, for example, detailed multi-agent models that model individual agents (an example is in Huigen, 2004), or models that explain the choices of a smaller number of large actor categories. The latter is of course much simpler to implement and the way we will proceed in this study.

Action-in-Context has four interconnected components. (1) The first is an often repeated "core element", comprising of the action, the actor, his options and his motivations. In Elster (1989), the latter two are called "opportunities" and "desires" but the structure is of the same simplicity: in order to act, people must have both the capacity and the will to do so. The other components of AiC are elaborations of the core element. (2) The "actors field" is an aspect of AiC that is, to our knowledge, unique in the social sciences. It describes the chains of social influence (causality, power) that run from the primary actors outward to other actors. Such chains may run, for instance, from farmers to NGOs, big landowners, traders, government agencies and the World Bank; an example is in Verburg *et al*. (2003). The method of constructing actors fields is by posing the question what actions (hence what actors) have an influence on the options and/or motivations of the primary actors. The secondary actors thus identified have their own options and motivations for these actions, which then may lead to the identification of tertiary actors, and so on. Moving from primary to secondary and further actors in AiC is the actor-based way of moving from proximate factors to underlying drivers *sensu* Geist and Lambin (2002). (3) The next component in AiC may be mixed freely with the preceding one and consist of a "deeper analysis" of the options and motivations of selected actors, distinguishing, *inter alia*, between elements of knowledge, resources, economic merit and culture. Figure 3.1 is AiC's broadly rational decision model designed to support this step, which will be discussed in some detail below. The deeper analysis is a second way to connect proximate factors to underlying culture and structure. (4) The final component of AiC is called the "actor model", which defines how the actor evaluates the options and motivations to come to his decision. In qualitative research, the actor model can often remain implicit. In such cases, the researcher 'puts himself in the place of the actor' (Vayda, 1983) and trusts that his audience can do the same, thus understanding the logic of the actor's choice without

[4] *Also Blaikie (1985) has this basic notion of explanation by putting actions of actors in context, but his contexts are conceptualised as systems rather than other actors. In AiC, explanations may reach up to the global level but this level is then still present as actors, e.g. the IMF in its own global 'life-world' and with its own options and motivations to act.*

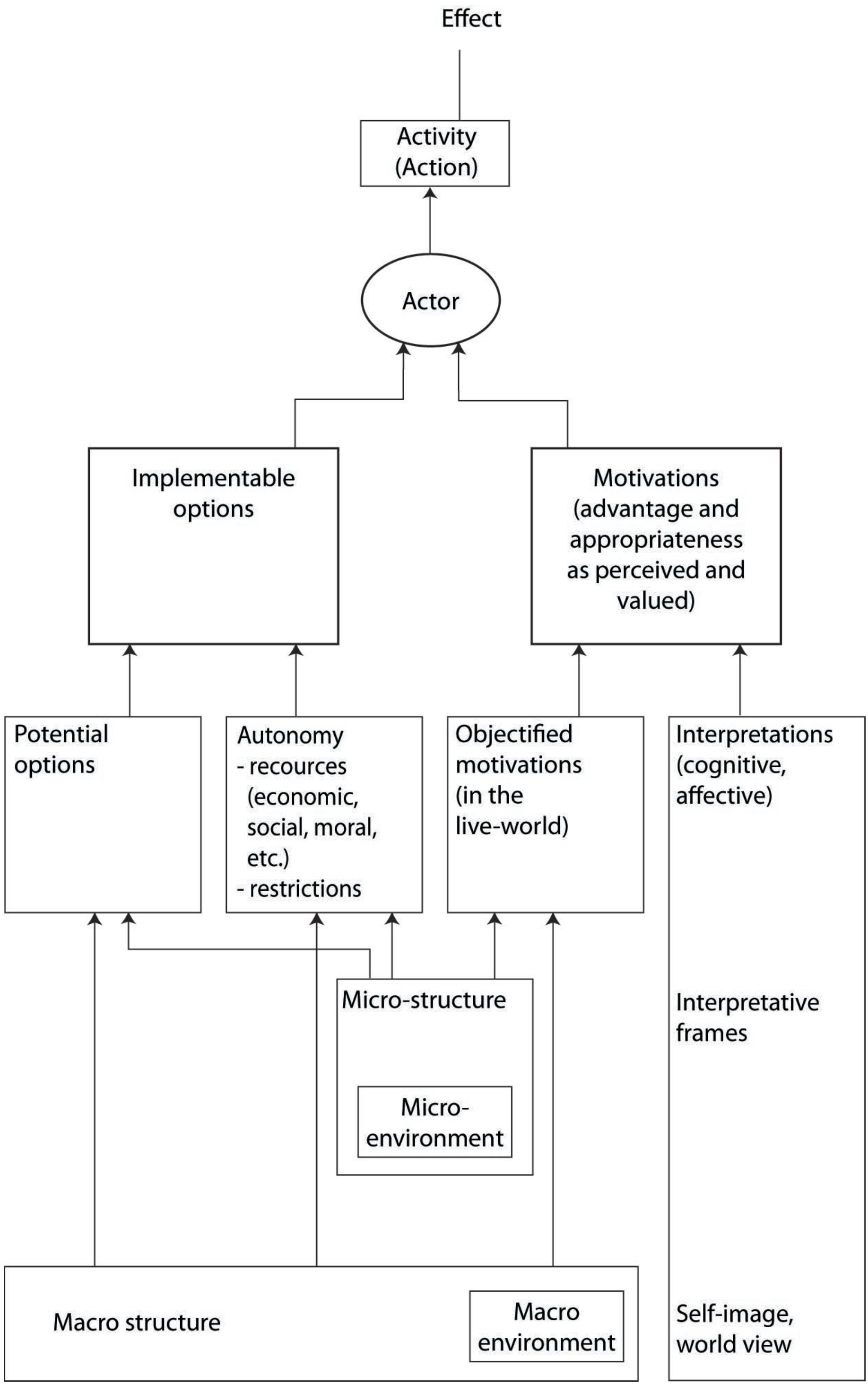


Figure 3.1: The decision model structure of AiC

a formalized model being needed. Another actor model is rational choice, which is to say that the actor chooses for the option of maximum merit – or utility, or profit, in which the definition of these terms marks the difference between narrow and broad rational choice. Broad rational choice may also be formalized in the shape of a multi-criteria table, as done by Hobbes *et al*. (n.d.). Specific for AiC is that it also offers an actor model for cases where a deeper reflection on the logics of human decision-making is warranted. This actor model distinguishes between three 'moral domains' of reasoning: *homo economicus* (*i.e.* rational choice), *homo honoris* and *homo communalis* (or 'ethics of care'); more detail is in De Groot (1992).[5]

Action-in-Context may be applied in many forms, in full or partially, formal or informal, as heuristic tool for guiding fieldwork or as a static model. See, for instance, De Groot and Kamminga (1995), Van den Top (1998) and Cleuren (2001) for qualitative applications on tropical deforestation. In the present chapter, we will especially use a quantified version of the decision model of the deeper analysis. The decision model (Figure 3.1) will be used as a structured model describing land use decision-making, which will be quantified and tested in full. In this respect the approach is deductive since the model and its parameters are determined using the pre-defined decision-making structure and the prediction is derived from this model, after which the result is tested against observations. Therefore, the decision model, depicted in Figure 3.1, is described in detail below.

In Figure 3.1, all arrows stand for causal relations. The top layers of the figure only repeat the core element of AiC, be it that the options are now specified as "implementable" options and that the motivations are the options' merits ("advantage and appropriateness") "as perceived and valued" by the actor. These specifications facilitate the definitions on the next lower level, which is the one of most interest here. At this level the implementable options are seen to result from the actor's "potential options" and "autonomy". Potential options are defined as everything the actor could do if he were infinitely autonomous (rich, powerful). Basically, they are all options that the actor knows to be possible. In land use issues, the typical role of agronomic research and extension is to bring more potential options to farmers (the former are then secondary actors). This is not to say, of course, that these options will also be implemented (*i.e.* become an action in Figure 3.1); farmers should also have the capacity (autonomy) to be able to implement them, as well as the motivation. "Autonomy" is capacity-to-implement, and is defined as the sum of all resources the actor can access (economic capital, private social capital, cultural capital, entitlements to common goods, etc.), taking into account possible restrictions (e.g. zoning regulations). Potential options and autonomy together determine the implementable options. Figure 3.1 does not specify the structure of this joint causality but we may note that it can not be some simple form of addition; just adding potential options does not automatically add to implementable options (let alone change actions), and neither does just adding to the actor's autonomy ('empowerment'). The case study of this chapter shows one way of modelling this.

---

[5] *In land use studies, it seems logical to assume that many choices will have a rational choice character. Deeper reflections may sometimes be needed, however, for instance to explain why people may stop planting trees once they are offered a financial incentive. It may be that planting trees was done in the moral domain of ethics of care or ethics of honour (we do it for each other, we do it for the children, we do it for the pride of the village), but flipped into the domain of homo economicus reasoning, triggered by the sudden association with monetary gain. And then of course, we do not plant trees for so little money.*

Motivations are the merits of the options. In Figure 3.1, the motivations "as perceived and valued" are separated into "objectified motivations" and their "interpretations". Objectified motivations are all those that may easily be quantified, such as economic cost and benefits, time expenditure, risk probabilities, caloric value of food and so on – in short, all these factors that micro-economists and farming system analysts feel at home with. Interpretations, on the other hand, are all those factors that give weight, coherence, shape and colour to the objectified motivations. Note that this way the interpretations are set as somehow multipliers of the objectified motivations rather than a 'filter' between actor and reality; psychology and culture add life to the actor, so to speak. Deeper down in the figure (but without causal arrows, indicating that the relationship is difficult to quantify) these interpretations are supposed to rise out of broader "interpretative frames" and "self-image / worldview". One example is the actor's image of what it is to be a good farmer (Zuiderwijk, 1998).
In Figure 3.1 furthermore, the third-layer elements are supposed to arise out of the actor's micro-structure (defined as all structures, social and physical, where the actor makes a difference) and macro-structure. Since these relationships do not play a role in our quantified model, we do not go into them here.
Overall, Figure 3.1 is obviously not something special as is AiC's actors field but rather designed as the reverse. It aims to overarch and coherently integrate all elements of broad rational choice theory, including cultural elements, the 'capitals' of Bebbington (1999) and so on, and remain close to the models of social psychology (albeit dropping the cumbersome intervening variable of 'attitudes'). Roughly then, many disciplinary focal points are included in the model: the options of agronomy and forestry, the objectified motivations of economics, the culture of anthropology, the capitals (autonomy) of development studies, the environment of geographers, and so on. Thus, the model facilitates explanatory work without preoccupation towards any specific discipline.

3.4 Material and methods

3.4.1 Study area

The study area is situated in Cagayan Valley in the northeastern part of the island Luzon, the Philippines (Figure 3.2). The study area includes 16 villages (*barangays*) in the municipality of San Mariano, in the province of Isabela, and comprises approximately 260 km2. It is situated between the town of San Mariano in the west and the forested mountains of the Sierra Madre in the east.
The population is approximately 16,500 persons (about 3,150 households) of various ethnic groups, among whom the Ilocano, Ibanag and Ifugao, who are all migrants or descendents of migrants that came to the area from the 1900s onwards, and the Kalinga and Agta, who are the indigenous inhabitants. Before immigration started, the area was completely forested with tropical lowland forest. At present, the study area shows a clear land use gradient ranging from intensive agriculture, with wet rice and yellow corn, near San Mariano via a scattered pattern of wet rice, yellow corn, banana, grasses, and (fruit) trees in the foothills to residual and primary forest in the eastern part.

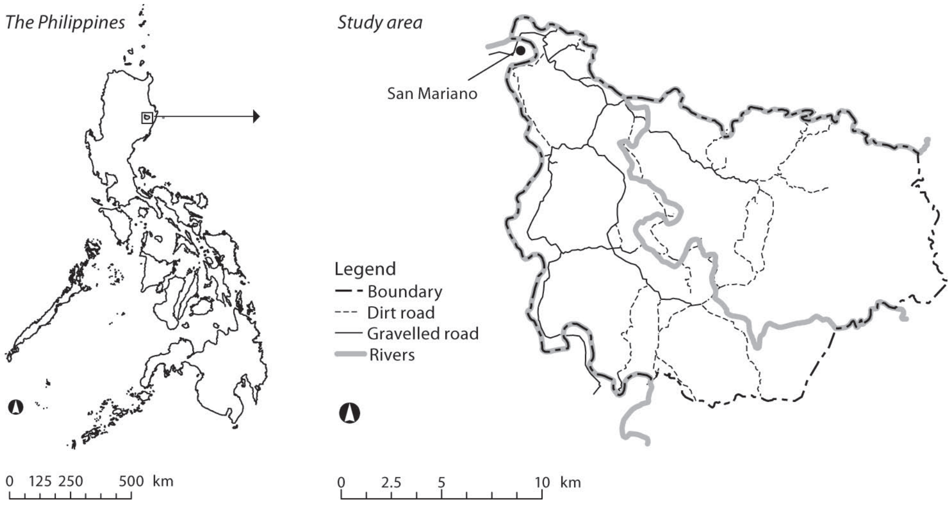


Figure 3.2: Location of the study area in the Philippines (left) and topographic features of the area (right)

3.4.2 Data collection

To collect the household-level data we conducted an interview campaign between June and November 2002 in 13 of the 16 villages, using a structured questionnaire to collect data on land use and its explanatory variables. Selection of households to be interviewed was based on systematic random sampling using population data available per village. In all villages every twentieth household was selected through systematic random sampling. A total of 151 households were interviewed.

The household questionnaire was structured in a nested hierarchy (see also Overmars and Verburg, 2005 (Chapter 2)) with the household level at the top and the plot level and the field level underneath. The plot is defined as a piece of land owned or used by the household, and a field is a specific part of the plot used for one of the land use types. On average a household owns 2.07 plots, with 1.15 different crops per plot. Variables were collected on their appropriate levels, e.g. soil characteristics at the field level, travel time to plot at the plot level and ethnicity at the household level. Records containing missing values and households without any land were excluded from the dataset. In total 114 households were included in the analysis. These households are the managers of 236 plots with 272 fields. This data was used to fit the multinomial regression model of the inductive approach and to validate both the inductive and the deductive model. Table 3.1 shows a selection of the data set consisting of those variables that turned out to be relevant in this case study.

Besides the questionnaires, semi-structured interviews were held with farmers and key-actors like heads of the villages and the elderly. If possible, these interviews were held with a group of people to enable discussion and verification. These interviews dealt, *inter alia*, with the motivations of the people to grow one crop or another. People were asked, for instance, to compare the different options for a specific field or to compare crops in general. People were also asked hypothetical questions about what they would change (or

Table 3.1: Description and descriptive statistics of the variables of the household survey (n = 272)

Variable name	**Description**	**Min.**	**Max.**	**Mean**	**St. dev.**	**Incl. in ded. model**	**Incl. in ind. model**
Dependent variables							
Yellow corn	1 if yellow corn, 0 otherwise	0	1	0.58		Y	N
Wet rice	1 if wet rice, 0 otherwise	0	1	0.13		Y	N
Banana	1 if banana, 0 otherwise	0	1	0.24		Y	N
Fruit trees	1 if fruit trees, 0 otherwise	0	1	0.05		Y	N
Land use	Yellow corn (4), wet rice (1), banana (2) and fruit (3) (nominal)	1	4			N	Y
Independent variables							
Slope1	1 if slope category is flat, 0 otherwise	0	1	0.38		Y	N
Slope2	1 if slope category is flat to rolling/moderate, 0 otherwise	0	1	0.23		Y	N
Slope3	1 if slope category is rolling/moderate, 0 otherwise	0	1	0.28		Y	N
Slope4	1 if slope category is rolling/'moderate to steep/hilly, 0 otherwise	0	1	0.08		Y	N
Slope5	1 if slope category is steep/hilly, 0 otherwise	0	1	0.03		Y	N
Slope	1 if flat to 5 if steep (ordinal)	1	5			N	Y
Ethnicity Ifugao	1 if male household head is Ifugao, 0 otherwise	0	1	0.10		Y	Y
Ethnicity Ibanag	1 if male household head is Ibanag, 0 otherwise	0	1	0.30		Y	N
Ethnicity Ilocano	1 if male household head is Ilocano (or Tagalog speaking), 0 otherwise	0	1	0.55		Y	Y
Ethnicity Kalinga	1 if male household head is Kalinga, 0 otherwise	0	1	0.03		Y	N
Ethnicity Other	1 if male household head is other than Ifugao, Ibanag, Ilocano or Kalinga	0	1	0.02		Y	N
Municipality of origin	1 if male or female head is born in San Mariano, 2 if both, 0 otherwise	0	2	1.17		Y	Y
Creek	1 if there is a creek or spring trough or bordering the plot, 0 otherwise	0	1	0.58		Y	Y
Plot distance	Minutes walking to the plot (min)	0	240	23.81	33.50	Y	Y
Transportation cost	Cost to transport a bag of corn from the house to San Mariano (pesos)	7	45	23.85	12.49	Y	Y
Average age	Average age of household heads (years)	21	78	42.71	13.07	Y	Y
Tenure tenant	1 if the household is tenant of the plot, 0 otherwise	0	1	0.19		Y	Y
Tenure SIFMA	1 if the plot is has SIFMA tenure, 0 otherwise	0	1	0.06		Y	Y

not) in their land use practices if certain conditions would change. These semi-structured interviews were guided by the AiC framework, which was used for progressive contextualization (Vayda, 1983) in this stage. The information from these interviews was used to qualitatively describe the processes in the area as well as to quantify the decision model and to determine the calculi to relate the elements of this model.

3.4.3 Analysis

The inductive (regression) model

For the inductive approach to predict the land use on a field we applied multinomial logistic regression, which is regularly used in land use analysis (e.g. Nelson *et al*., 2001; Müller and Zeller, 2002). Multinomial logistic regression extends the possibilities of logistic regression by allowing for more than two categories in the dependent variable. In this case four land use categories are included. The parameter estimates are calculated simultaneously and the probabilities of the different land use categories add up to one. The final prediction is the land use with the highest probability. In the multinomial model, the estimated parameters are to be interpreted in relation to one of the categories of the dependent variable, which serves as reference category. Under the assumption that all error terms are mutually independent and have a log Weibull distribution the multinomial logistic regression model can be expressed as follows:

$$P(y_i = j) = \frac{\exp(x'_{ij}\beta)}{1 + \exp(x'_{i2}\beta) + \ldots + \exp(x'_{iM}\beta)}, j = 1,2,\ldots, M \qquad (3.1)$$

where y is the dependent variable (land use type), j are the alternatives of M land use types, i is the i^{th} field, x are the explanatory variables, and $ß$ is a vector of regression coefficients. In this equation ($x'_{i1}ß$) is normalized and set to zero, (exp(0)=1), so in this case alternative 1 (yellow corn) is used as the reference category (Verbeek, 2000).

The probability ratio (odds ratio) for the normalized model is given by:

$$\frac{P(y_i = 2)}{P(y_i = 1)} = \exp(x'_{i2}\beta) \qquad (3.2)$$

This probability ratio is used to interpret the effect of a change of the values of one of the explanatory variables. Equation 3.2 shows that the probability ratio is only dependent on the reference alternative and the alternative under study and does not depend on the nature of any of the other alternatives (Verbeek, 2000). An increase in the probability of a land use relative to the reference land use may have no significance on the probability when compared to other possible land uses (Nelson *et al*., 2001).

In the multinomial regression the following variables were taken into account: slope, ethnicity, municipality of origin, presence of creek, plot distance, transportation cost, average age of the household heads and tenure (see also Table 3.1). The approach follows the rung 2 level of the deduction/inductive ladder presented previously: selection of the factors was inspired by several (disciplinary) land use theories, prior studies and field knowledge.[6]

---

[6] *The original 'rung 2' study upon which the multinomial regression model in this chapter was based is described in Overmars and Verburg (2005) (Chapter 2).*

These theories are not tested as such, but their factors are used in a regression analysis. The variable slope was taken into account as if it were a continuous variable with a ratio scale (flat slopes were given the value 1 up to steeps slopes with the value 5). Doing this we suggest a linear relation between the slope categories and the land use types. Including four of the slope categories as nominal variables, as we do in the deductive model, was not possible because, due to (quasi) complete separation, the maximum likelihood of that multinomial regression model was impossible to calculate. For the same reason the variable ethnicity Ibanag male was not included in the regression analysis. All variables were tested for multi-collinearity by regressing each of the independent variables upon the remaining independent variables. This test revealed no multi-collinearity.

The causal model for the deductive (AiC) approach
The causal model we applied is a quantification of the decision model of the Action-in-Context framework (the deeper analysis).[7] As actor model we use broad rational choice, saying that the actor chooses for the implementable option of maximum merit. In the next section, a qualitative description of the case is provided following the structure of the framework. A detailed description of the actions is given and the primary actors are identified and described. Then, the potential options and autonomy are described in detail, followed by the objectified motivations and interpretations. The subsequent section is devoted to the quantification of this causal decision model. This model was used to predict the probability of the occurrence of the land use types. In explaining and quantifying the model we used the same set of explanatory factors as in the multinomial model, together with some additional constants like, for example, maximum benefit and investments.

Comparison
Since the inductive (regression) model and the deductive (AiC) model describe the land use system in the same area and use the same variables, the performance of the models can easily be compared. The performance was calculated using cross-tables (also called prediction matrix). The outcomes are a goodness-of-fit and a goodness-of-prediction for the inductive and deductive model, respectively. The cross-tables show the observed category against the modelled category of land use. Subsequently, the diagonal of the table contains the correct predictions. Besides the number of correctly predicted fields also the relative quantity of the predicted categories will be taken into account while assessing the model performance. In this application we consider not only the total score of good predictions to be important, but we also want that the correct predictions do not have an extreme bias towards only a few of the categories.

3.5 Qualitative description of the deductive model

3.5.1 Actors

Ninety percent of the households in the study area have a piece of land to cultivate. Farming is the main source of income of 80 percent of the households and the second source of

---

[7] *The actors field of land use choices has not been part of the model. The actors field explaining yellow corn production in the area and the actors field of forest migration are described by Van den Top (1998).*

income for 10 percent of the households. Besides this, most people work also as a labourer for other farmers, which provides them with additional income. The actors considered in the analysis are those households that have control over a piece of land that they can possibly cultivate. They will be called farmers in the remainder of the chapter.

3.5.2 Actions

The analysis focuses on the decision-making on agricultural land use types. However, the possibilities of making a living in the area are broader than agriculture alone. Besides options that do not involve land (e.g. carpenter, storekeeper, driver), the people in the area also have other land use options such as small-scale logging and collecting non-timber forest products. Both these option categories are not considered in this study. Hence, the research question is why farmers cultivate a certain crop at a certain location. The area that the farmer cultivates per crop is not subject of analysis. This study is restricted to the explanation of the occurrence of agricultural land use types on existing fields.

3.5.3 Potential options

To construct a list of potential land use options we could include, for example, all crops grown in the region over the last 30 years. Based on data from 1971 onwards, this would include tobacco, peanut, white corn, vegetables and several other crops, besides the four most important crops at present, which are wet rice, yellow corn, banana and fruit trees. Currently, these land use types account for 92 percent of the observed fields. Considering our research objective, which is a methodological comparison rather than location specific detail, we chose to include only these four land use types. Methodologically, it is important to know that these four potential options are not all of them implementable options at all locations, as we will see. Some more detail on these four crops is supplied below.
Yellow corn is the most important cash crop in the area. To cultivate yellow corn, the farmers use hybrid seeds, often apply large quantities of fertilizer and most of them use pesticides. Most farmers get these agrochemical inputs, which are indispensable to cultivate yellow corn, on credit from traders. Often, these traders also provide the farmers with consumptive credit. The farmers are obliged to sell the harvest to the same traders, who charge a high interest rate, to pay back their debts. This reduces the farmers' freedom to get the highest price on the market. Due to the risks inherent to corn production and due to the credit system farmers end up in a strong debt bondage with the traders. Yet, many farmers continue to seek the patronage relationship with the traders because corn is in fact is their only access to credit and the traders may also help out in times of need (Van den Top, 1998). Yellow corn is cultivated twice per year. The main risks for yellow corn, as reported by the farmers, are pests like rats, insects and birds and climatic conditions like droughts, floods and typhoons.
The cultivation of wet rice is for consumption by the household and is only marketed if there is a rare surplus. Water sources in the area, necessary to cultivate wet rice, are rainfall, natural irrigation by creeks and springs, and a large irrigation scheme. Most farmers transplant the rice, though some use the system of direct seeding (broadcasting), which is less time-consuming, but requires some skills. Fertilizer and pesticides are used, but less than in yellow corn production. If sufficient water is available for irrigation two crops are cultivated per year. The most reported risk is crop damage by rats, insects, birds and snails.

Banana is largely a cash crop, but the input level of fertilizers and pesticides is low. The most important varieties that are used in the area are Damilig, which is a cooking banana for industrial use (banana chips and banana ketchup), and Lakatan and Turdan, which are dessert bananas. These three have different characteristics regarding spacing, harvest and price. Damilig is normally sold at a lower price, but the plantation has to be renewed less often and is more resistant to diseases than Turdan and Lakatan. The time between planting and the first harvest is 1 to 1.5 years (Sterken, 2004). The main risks for banana are typhoons, Banana Bunchy Top Virus and wild pigs. Newly planted banana fields are often intercropped with yellow corn or upland rice for the first one or two years. In some areas bananas are cultivated in rows between fields.

Fruit trees are not grown on a large scale in the area. Recently, a number of farmers participated in a program called SIFMA (Socialized Industrial Forest Management Agreement), which awarded them with 25 years of tenure rights provided that they plant a certain area with (fruit) trees (mainly mango, citrus and coconut), which were provided for free by an NGO (General, 1999). This land use type was included in the analysis, because it might become an important land use type in the future. However, marketing of fruit is still underdeveloped in the area and also maintenance of the plantations is often lacking, which prevents fruit tree plantations from being successful so far. Fruit trees are often intercropped with yellow corn, especially when the fruit trees are small and not bearing fruits yet. Others plant the fruit trees on the borders of their fields. Only a few farmers have fruit trees as their only crop (Klein, 2003). The most important risks for fruit trees are typhoons and fire.

3.5.4 Autonomy

The level of autonomy determines if a potential option can be implemented or not. Autonomy consists of two elements: resources and (absence of) restrictions. The autonomy of the farmer is restricted by the variables tenancy, creek and ethnicity. If the farmer is a tenant of the land he cultivates, the landowner often decides what the tenant should cultivate, which is most often yellow corn. So, the tenant cannot make an autonomous decision about what to cultivate. For the cultivation of wet rice two restrictions were added: presence of a creek and the farmers' ethnicity. Wet rice cultivation is restricted to Ifugao and Ilocano because, generally spoken, they have better skills and knowledge in constructing rice fields and rice terraces and cultivating wet rice. Ibanag people, who are the original lowlanders of the Cagayan valley, have a long tradition in corn cultivation. Formerly, they produced white corn as staple food because growing white corn could be combined with tobacco, which was an important crop in the region during the Spanish time (Van den Top, 1998). The assumption is that many Ibanag farmers do not know (or know to a lesser extent) how to cultivate wet rice because it was not part of their tradition (Romero, pers. comm.). The presence of a creek on or near the plot is important for the cultivation of wet rice, since it needs a water source. The source of water could be a pump or an artificial irrigation system, though in most cases this is a small river or stream that is diverted towards the rice field. This stream should be close to the rice field. So, a creek nearby is considered to be a prerequisite to cultivate wet rice.

The other element of autonomy is the resources of a farmer. In this case study, resources are considered to be necessary to do initial investments to start a new land use type, like clearing a forested area for corn cultivation or constructing a rice terrace. If the resources

are sufficient to do the investment the land use type is an implementable option. So, the initial investments function as a threshold. They are built up of two components: basic investments and, for rice only, additional investments dependent on slope. In our model, the resources are composed of the 'level of assistance', the possibility to obtain credit to buy inputs for a crop and participation in the SIFMA program, which together should be sufficient to do the initial investments for a specific land use type.
The resource 'assistance' is composed of the factor municipality of origin and a factor proportional to the average age of the household heads. The municipality of origin of the household heads is considered to be indicative for the size of a household's social network (roughly: social capital). The assumption is that people who are born in San Mariano have more relatives and friends nearby than people coming from outside the municipality. This social network is necessary for farmers to organize a group of people to do the work at relatively low costs. In many places in the area it is a custom to help one another by working in a large group to do the larger jobs like cleaning, planting and harvesting (Moonen, 2002). Ifugao were considered to have assistance from relatives even when they are not from San Mariano, because often they migrate after invitation of relatives or friends and cluster together. Also a higher age is considered to be indicative for a larger network to organize labour (children, relatives).
Another way to meet the necessary investments is to borrow money. In the research area credit is almost exclusively provided for yellow corn. Other sources of capital to make investments for other crops are hardly available, which actually restricts farmers in their options.
The last resource is participation in the SIFMA program, which provides tenural security and assistance in starting an agro-forestry plantation and therefore applies to the land use type fruit trees. In the study area, land titles can only be obtained for the so-called A&D (alienable and disposable) lands, which are the flat areas. Sloping lands are classified as forest and owned by the state and for these lands no official titles can be acquired. Nevertheless people cultivate these state-owned forest lands. Governmental as well as non-governmental organizations encourage farmers on these lands to invest in agro-forestry systems, which are considered to be more sustainable than arable farming. However, insecure property rights hamper the development of these agricultural systems because they require high investments and need a long time to become profitable (e.g. tree planting and conservation measures). Farmers do not have the money to invest and they are not sure if they can still use the land at the time the crops become profitable. Therefore, the SIFMA program allows farmers to apply for a 'stewardship contract' for 25 years while committing themselves to a more sustainable way of farming. Farmers that were awarded a SIFMA contract can receive free fruit tree seedlings to be planted on their SIFMA lot, covering a part of the high initial investment costs.

3.5.5 Objectified motivations

Motivations are composed of objectified motivations and interpretations. In this study the objectified motivations are considered to be the net economic benefit from one hectare of a land use type at the moment the product is sold in San Mariano (in case of yellow corn, banana and fruit trees) or consumed (in case of rice). The net benefit is defined as the maximum benefit under ideal climatological (no extraordinary droughts or typhoons) and biophysical conditions (flat area with a good soil) for an average price, multiplied by a yield-reducing factor depending on slope (for yellow corn) and a yield-reducing factor depending on risks lowered with the transport cost. The maximum benefit is considered to be the same at all locations in the study area.

Steeper slopes will decrease the objectified motivation towards corn because the costs are higher and the yield is lower. Ploughing is more difficult or impossible on steeper slopes, which increases the costs spent on planting the corn. On steep slopes, seeds and fertilizer are washed away during heavy rains. This will reduce the yield of such a field in comparison with flat fields. So, on steeper slopes the cultivation of corn will cost more in effort and time and will yield less because of the poorer productivity of the plot.

Bananas can grow in every landscape position, unless soil drainage is very bad (Valmayor *et al*., 1990). Many of the drawbacks that corn has on steep slopes do not apply to banana. Banana cultivation does not involve tillage, so ploughing is not required. Bananas are renewed only once every 5-15 years. The productivity of banana is the same on steep slopes and flat areas. So, slope does not influence the motivation towards growing banana. On the contrary, many farmers plant bananas to prevent soil erosion on steep slopes.

Transportation cost is the cost to transport the product from a farmer's home to the market. In this study transportation costs apply to yellow corn, banana and fruits. Rice is used for household consumption or sold in the neighbourhood. Additionally, the distance from the plot to the residence of the farmer is taken into account. If a plot is far from the farmer's village, the farmer needs to invest more effort and time in cultivating a crop on that plot. This effect will be most prevalent with yellow corn and wet rice, which need to be frequented by the farmer more often than other crops like banana. Moreover, fields that are far away have more risk to be damaged by fire, water buffalos or people.

3.5.6 Interpretations

The objectified motivations are adjusted to the interpretation of the individual land manager. In this model interpretations are simplified to crop preferences of the different ethnic groups. The traditions and cultural values of the ethnic groups are different for the crops considered. These traditions make that people feel at ease with growing certain crop or that they are proud to have it. As said before, Ifugaos and Ilocanos have a tradition in wet rice cultivation whereas the Ibanag have a tradition in corn cultivation and not in wet rice cultivation. This is reflected in their preference for corn and rice. The preference for banana and fruit trees seems to be the same for all ethnic groups. The objectified motivations are combined with the interpretations to become the motivations "as perceived and valued". This may cause people of different ethnicity to choose a different land use option even if the objectified motivations are the same for both ethnic groups.

3.6 Quantifying the deductive model

Based on the fieldwork and the qualitative analysis in the previous section, which is derived from this fieldwork, the formal model with the structure of the deeper analysis of Figure 3.1, is quantified as follows. The core of the model is that the predicted land use is the implementable land use option with the highest motivation (Equation 3.3). Starting with the options side of the model, Equation 3.4 shows that the implementable options are composed of potential options and autonomy. The potential options are yellow corn, wet rice, banana and fruit trees. Autonomy (Equation 3.5) is determined by restrictions and resources. If a restriction is 1 or the resources are 0, the autonomy is 0 and the potential option cannot be implemented.

Action = f(implementable options, motivations) (3.3)
*Implementable options = potential options * autonomy* (3.4)
*Autonomy = (1 - restrictions) * resources* (3.5)
Restrictions = f(T_TENANT[a]*, CREEK, ETHNICITY, CROP)* (3.6)
Resources = IF(assistance + credit + tenure_SIFMA – investments>=0), resources = 1, else 0 (3.7)
Assistance = (MUNICPALITY_ORG + (AGE/34))/3 (3.8)
Credit = f(CROP) (3.9)
Tenure_SIFMA = f(TENURE SIFMA, CROP) (3.10)
Investment = inv_basic + inv_slope (3.11)
inv_basic = f(CROP) (3.12)
inv_slope = f(SLOPE, CROP) (3.13)

[a] *Variables in the equations are written in capitals*

As described in the qualitative model description the restrictions in this study are a function of tenancy, creek, ethnicity and crop (Equation 3.6). If the land manager is a tenant we only consider yellow corn to be an option. So, if the variable tenant is 1, all land use types except corn were given value 1 (Table 3.2). In the Equations 3.4 and 3.5 this leads to an autonomy of zero and therefore to a zero for the implementable options calculation, meaning the land use type is no option. If the field is not cultivated by a tenant (tenure tenant = 0) all options are possible. Wet rice is only possible if a creek is nearby and if the field is cultivated by farmers of the ethnicity Ilocano or Ifugao. These restrictions are summarized in Table 3.2. Calculations run similar to the example above. These relations are intuitively determined based on field experience and the interviews and are not fitted in any way.

In the model the resources assistance, credit and tenure SIFMA should cover the investments for a land use type to make this land use implementable (Equation 3.7). The assistance depends on municipality of origin and average age of the household heads and is specified in Equation 3.8. (For Ifugao the value of the factor municipality of origin was set on 2 even if they are not born in the municipality of San Mariano). The equation is formulated in such a way that the result is centred around one for a specific age (34 yrs.). This specific parameter was optimised, since no clear theoretical idea was available to determine the influence of age. As explained in the previous section credit is 1 (possible) for yellow corn and 0 (not possible) for the other crops. The resource due to the assistance by the NGO in the SIFMA areas is 1 for fruit trees (Table 3.2). The investments consist of two parts: basic investments and investments due to slope (Equation 3.11). The basic investments are defined as the basic investments necessary to start a new field for a specific land use type. The values of the basic investments (Table 3.2) are relative to the initial investments for yellow corn, which were set on 1. This relation was estimated by the authors based on field experience. The relation between slope and the investment necessary to build a rice terrace (Table 3.3) was estimated according to the amount of labour necessary to build a terrace (Romero, pers. comm.), which was calculated as an average from field observations. The extra investment due to slope was set on 1 for the terraces on slope category 3 and the other categories were estimated calculated to this value.

Table 3.2: Factors that determine autonomy through restrictions and resources (Values in the tables are used in the model)

Variable	**Yellow corn**	**Wet rice**	**Banana**	**Fruit trees**
<u>Restrictions</u>				
Tenure tenant = 1	0	1	1	1
Tenure tenant = 0	0	0	0	0
Creek = 1	0	0	0	0
Creek = 0	0	1	0	0
Eth. Ilocano and Ifugao	0	0	0	0
Eth. Ibanag, Kalinga and Other	0	1	0	0
<u>Resources</u>				
Credit	1	0	0	0
Tenure SIFMA = 1	0	0	0	1
Tenure SIFMA = 0	0	0	0	0
Investments	1	1.2	0.3	1.5

The result of the model structure and the parameters is that corn is possible for all farmers because investments can be covered by credits, banana is also possible for all farmers because the initial investments are low and that fruit trees is possible for people that have a SIFMA lot. Initially, the calculation resulted in no possibilities for wet rice, because of too high initial investments. Since rice does occur in the area this rule was relaxed a little. This can be justified by the fact that rice fields are usually smaller than a hectare and the calculation is per hectare and therefore initial investments are smaller in reality than the calculated investments.

Table 3.3: Calculation of investment term for the construction of rice terraces

Slope category	**Days labour per ha***	**Investment term for rice**
Slope1	52	0
Slope2	292	0.36
Slope3	716	1
Slope4	2209	3.25
Slope5		4.33**

** Source: Romero (pers. comm.) (n = 28)*

*** Estimated by the authors*

The right branch of the AiC model (Figure 3.1) deals with the motivations. The motivations (as interpreted) consist of objectified motivations multiplied with a factor for the preferences (Equation 3.14). In this case the objectified motivations are expressed in Philippine Pesos and consist of the maximum benefit, a slope factor, a risk factor and transportation cost. The maximum benefit is expressed in Table 3.4. These values stem from average yields reported in interviews, except from the maximum benefit for fruit trees, which was calculated by Klein (2003). For yellow corn the maximum benefit is multiplied by a yield factor depending on slope (Table 3.5) and an average yield reducing factor depending on estimated risks for all land use types (Table 3.6). The former were derived from reported yields on fields with different slopes and the latter was derived from interpretations of damage reports in the interviews. This risks table does not include the regular pest and diseases, because these are incorporated in the estimated yields. The high typhoon risk for banana is related to the fact that the banana is not productive for 1 to 1.5 years after a typhoon, while other crops can be replanted and productive several months after destruction. The transportation costs are computed according to Verburg *et al*. (2004a). Travelling distance to the plot (variable 'plot distance') was translated into monetary costs. For wet rice transportation costs were only based on costs from the residence to the field, since the product is not marketed, and for the other crops the calculation is a combination of costs from field to residence and from residence to the town of San Mariano. The preferences (Table 3.7) based on ethnicity were quantified by the authors based on qualitative descriptions by the farmers. It may be noted that in this model the effect of the preference for wet rice cultivation is cancelled out by the much higher net benefit of wet rice compared to the other crops, so differences in preference do not change the prediction of rice.

*Motivations = objectified motivations * preferences* (3.14)

*Objectified motivations (net benefit) = max_benefit * slope_fact * (1-risk) - tr_costs* (3.15)

Max_benefit = f(crop) (3.16)

Slope_fact = f(slope, crop) (3.17)

Risk = f(crop) (3.18)

Tr_costs = f(tr_cost, plot_distance, crop) (3.19)

Preferences = f(ethnicity, crop) (3.20)

The objectified motivations with the interpretations combine into the motivations (as perceived and valued) for each field for all four crops. These motivations (as perceived and valued) are summarized in Table 3.8. Cultivating wet rice is by far the most profitable followed by fruit trees. The benefits from corn and banana are very similar.

Table 3.4: Maximum benefit (in Ph. Pesos, calculated from field data) per land use type

Crop	Max_benefit
Yellow corn	22435*
Wet rice	42000*
Banana	21213*
Fruit trees	32230**

** Source: field data*

*** Source: Klein (2003)*

Table 3.5: Calculation of slope factor for yellow corn

Slope category	Average yield (kg/ha)	Slope_factor
Slope1	3581	1.00
Slope2	3829	1.07
Slope3	3070	0.86
Slope4	no data	0.50*
Slope5	no data	0.20*

** Estimated by the authors; other data based on field observations (n = 37)*

Table 3.6: Risk factors of crops

Risk/Crop	Yellow corn	Wet rice	Banana	Fruit trees
Typhoon	0.10	0.05	0.20	0.10
Drought	0.10	0.05	0.02	0.02
BBTV	0.00	0.00	0.10	0.00
Risk total	0.20	0.10	0.32	0.12

N.B. All numbers are estimated by the authors

Table 3.7: Preference factors based on ethnicity

Ethnicity/crop	Yellow corn	Wet rice	Banana	Fruit trees
Ifugao	0.9	1.2	1	1
Ibanag	1.2	0.9	1	1
Ilocano	1	1.2	1	1
Kalinga	1	1	1	1
Other	1	1	1	1

N.B. All numbers are estimated by the authors

Table 3.8: Summary of the motivational value for all fields per land use type

Land use	Average (Peso/ha)	St.dev
Yellow corn	14239	3834
Wet rice	40694	5181
Banana	12474	672
Fruit trees	23313	1739

3.7 Model Results

3.7.1 The inductive (multinomial regression) model

The inductive model (Table 3.9) shows the estimated parameters of wet rice, banana and fruit trees in relation to yellow corn, which is the reference category. The estimated coefficients should be interpreted relative to this category. For example, one unit increase in the explanatory variable creek will increase the $ln(P_{wet\ rice} / P_{yellow\ corn})$ with 1.988. Positive coefficients result in an increase of the probability relative to the reference category and negative coefficients in a decrease. In multinomial regression analysis the interpretation of the estimated coefficients is not completely straightforward, because the coefficients only tell us the relation between one land use category and the reference category. This complicates direct comparison of the inductive model with the deductive model.

The cross-tabulation (Table 3.10A) shows the number of observations that is modelled correctly (the bold diagonal figures) and if not, in which category. The right column shows the percentage of the observations that was fitted right. Especially yellow corn was fitted very well (91 percent), banana was fitted reasonably well (66 percent) and wet rice (50 percent) and fruit trees (43 percent) were fitted somewhat weakly. In total, the multinomial regression model fitted a total of 209 out of 272 (77 percent) observations correctly. A test was performed to what extent the observed and modelled land use distributions are alike. The Chi-square statistic of this test is significant at the 0.0001 level. The kappa statistic, which indicates the proportion of agreement after chance has been excluded, is 0.579.

Table 3.9: The multinomial regression model

	Wet rice		***Banana***		***Fruit trees***	
Variables	**b**	**s.e.**	**b**	**s.e.**	**b**	**s.e.**
Intercept	-3.182	1.764	-9.936***	1.958	-11.420***	3.215
Slope	-1.302**	0.408	2.224***	0.333	1.628***	0.489
Ethnicity Ifugao male	2.631*	1.073	-0.295	1.243	-1.588	1.661
Ethnicity Ilocano male	1.678*	0.705	0.380	0.509	-0.131	0.965
Municipality of origin	-0.668	0.359	-0.097	0.342	-0.402	0.601
Creek	1.988***	0.554	0.013	0.505	0.502	1.001
Plot distance	-0.008	0.014	0.008	0.007	0.016*	0.008
Transportation cost	0.051*	0.023	0.065**	0.022	0.050	0.038
Average age	0.011	0.021	0.037	0.021	0.067	0.036
Tenure tenant	0.084	0.594	-0.921	0.678	0.159	1.261
Tenure SIFMA	-0.908	1.363	0.749	1.297	3.931*	1.661

** p<0.05, ** p<0.01, *** p<0.001*

Table 3.10: Observed land use vs. modelled land use of the inductive (regression) model (A) and deductive (AiC) model (B)

A.	*Predicted land use type*					
Inductive (regression) model	**Wet rice**	**Banana**	**Fruit trees**	**Yellow corn**	**Total**	**% Correct**
Observed land use type						
Wet rice	**18**	1	1	16	36	50.0
Banana	0	**42**	3	19	64	65.6
Fruit trees	0	5	**6**	3	14	42.9
Yellow corn	4	11	0	**143**	158	90.5
Total	22	59	10	181	272	76.8

B.	*Predicted land use type*					
Deductive (AiC) model	**Wet rice**	**Banana**	**Fruit trees**	**Yellow corn**	**Total**	**% Correct**
Observed land use type						
Wet rice	**21**	1	1	13	36	58.3
Banana	2	**31**	5	26	64	48.4
Fruit trees	1	3	**5**	5	14	35.7
Yellow corn	18	5	1	**134**	158	84.8
Total	42	40	12	178	272	70.2

3.7.2 The deductive (AiC) model

The results of the deductive model (Table 3.10B) are largely the same as the results of the inductive model. Wet rice is predicted better than in the multinomial model and for the other land use types the deductive model performed slightly less. The model was able to predict 70 percent of the occurring land uses of a dataset of 272 fields. The Chi-square statistic is significant at the 0.0001 level. For this model the kappa statistic is 0.471. The kappa statistics of the two models are not significantly different ($p < 0.05$) (Couto, 2003). So, based on the kappa statistic it cannot be shown that the inductive model performs better than the deductive model.

3.8 Discussion and Conclusions

Following the objectives of the chapter, this section will discuss some of the case study outcomes, but pays special attention to the differences between inductive and deductive research approaches and especially those presented in this study.

3.8.1 Factors of land use change

The AiC framework is designed to incorporate relevant factors from all scientific disciplines in a balanced manner. Using the deeper analysis of the AiC framework as a template for the deductive model, we were able to incorporate variables from various different disciplines, including geographic (e.g. slope, presence of creek), economic (e.g. investments, net benefit), social (age, municipality of origin), anthropologic (ethnicity), and policy (the tenural instrument SIFMA). The same factors are incorporated in the inductive model and in that respect both models are equally multi-disciplinary, 'integrated' models. The factors comprise a good many of those listed in the recent overviews of driving factors by Geist and Lambin (2002) and Lambin *et al*. (2003), even though we have focused only on simple crop choices. Since we have not compared land use in two or more points in time, our factors are explanatory factors rather than dynamic 'drivers' of land us change, formally. Predictions of the effect of incremental changes in factors may be derived from both models, however (as in Nelson *et al*. 2001, for example). In the sense of factors and predictions, therefore, the present study is comparable to mainstream land use studies.

3.8.2 Field-level conclusions

The inductive approach (of type 2 in terms of the deductive/inductive ladder) has been to fit all factors to the actual land use in a multinomial regression, thus generating a structure of land use depending on $ln(P_{category\ M} / P_{reference\ category}) = ß_1x_1 + ß_2x_2 + \ldots + ß_nx_n$. The deductive approach (of type 5 of the deductive/inductive ladder) has been to use the factors as elements in a model of the decisions of the land users, to quantify the model on the basis of field knowledge and then test the whole causal structure against the actual land use data. The results of the two approaches look much alike, on the surface. The inductive model fitted 77 percent of the observations correctly, and the deductive model predicted 70 percent of the observations correctly. The performance on yellow corn was good in both models (90 and 85 percent, respectively). Both models overestimated the total amount of yellow corn and underestimated the amount of banana fields. In both models some of the observed banana fields were classified as yellow corn. This is possibly due to the fact that the maximum benefit of both crops is almost the same and both crops have hardly any restrictions (corn is supported with credit, banana has low initial investments). Therefore, small imperfections in the model can cause the prediction to go wrong. Another cause of poor prediction of banana could be the existence of a time lag between changes in context and changes in the land use observed. Most bananas are cultivated for 5 to 15 years and may still be standing even when benefits are low. The predictive capability for wet rice was low for both models. The deductive model performed a little better than the inductive model. In the deductive model, the occurrence of wet rice is completely determined by the restrictions (absence of a creek, 'non-rice ethnicity' and tenancy) and these restrictions may be too rigid. In both models the prediction of fruit trees is weak. This is due to the fact that only five percent of the observations are fruit trees and that the mechanisms behind the planting fruit trees in this area are not completely understood yet.

3.8.3 Differences between inductive and deductive approaches

The real and important differences between the two models lie on a deeper level. As said in the second section, deductive approaches, because they start out from theory and maintain theoretical structures, better feed back into theory development than inductive studies tend to do. The present study, for instance, is a true verification of broad rational choice theory expressed in the structure of Action-in-Context's decision model. The Popperian critique here would be, of course, that this is no surprise because one should aim to verify *un*likely structures, or to falsify the likely ones. Nevertheless, an entry to the theory level it is and once there, progress may be pursued in many directions, including the testing of less likely actor models or spatially explicit neo-Thünian theory of moving land use frontiers (e.g. De Groot, 2003). The model structure may also be expanded easily to include multi-level causal influences on the region's land use, for instance through AiC's actors field, see Verburg *et al*. (2003). Using the actors field, we arrive at a multi-agent modelling of land use.

A second advantage of deductive work is that, as it tests full causal structures rather than separate causal factors, a much better grip on causality is established. Two specific aspects of this characteristic stand out from the present study. (1) Any inductive model, working as it does from the data 'upward', can only fit for variables that vary across the dataset. The influence of all factors that are constant across the dataset, such as in our case the market price of corn, end up, implicitly, in the intercept (β_0). Therefore, it is less straightforward to predict the amount of land use change for a change in one of the factors accumulated in the intercept. A deductive model, however, allows the inclusion of all factors assumed to have causal influence (hence relevant for explanations as well as policy-oriented predictions). In the deductive model, for instance, changes from corn to another land use could be predicted if the corn price, and with that its net benefit, would fall. (2) For the same reason of testing full structure rather than factors, deductive models are able to handle new phenomena, assuming that they do not alter the model structure. In the study area, for instance, cassava may be an alternative cash crop. The inductive model cannot handle this, because cassava is new, hence absent from the dataset and therewith from the model. In order to make a prediction of the region's response to cassava by way of the deductive model we do need general cassava production data such as its price, productivity, position in cultural preferences, accompanying credit scheme and so on, but once we have these, a prediction is produced and various policy scenarios may be studied. (The predictions may turn out to be wrong, of course, but that is a normal issue of model quality.)

3.8.4 Reaping the benefits of combining inductive and deductive approaches

Research programs often tend toward a certain development on the deduction/induction ladder. For example, starting out with a more or less extreme data mining (rung 1), the selection and shaping of causal factors may become more led by insight (rung 2) and researchers may end up in studies more consciously and fully in discussion with theory (rung 3). In fact, quite a number of inductively acquired insights into our field research region, gathered by previous studies, informal discussions and visits, interviews and observations, underlie our own deductive model. In other words, we sojourned long on rungs 1 and 2 before our deductive jump to rung 5 in the present chapter. For the sake of clear-cut illustration we refrained here from what would in fact have been the natural follow-up of our strictly deductive approach, namely, to better calibrate and fit our model parameters on reality, *i.e.* move one level down to rung 4, searching to reach a better fit than the 70 percent of the strictly deductive model. After that, we could even have begun to adapt our whole model structure in discussion with theory and field realities, thereby arriving at rung 3.

Overall, then, the most effective way to reap the benefits of more deductive work does not seem to rigidly 'go deductive' and stay there. Rather, the message should be that research will profit most from a consciousness of the whole spectrum between the inductive and deductive extremes, an awareness of the advantages of deductive approaches versus the currently dominant inductive research routines, and then seek the most fertile sequences and interactions between inductive and deductive work. This then is the invitation meant to emanate from the present chapter to the scientific community of land use change research.

4

Multilevel modelling of land use from field to village level in the Philippines

Abstract

In land use research regression techniques are a widely used approach to explore datasets and to test hypotheses between land use variables and socio-economic, institutional and environmental variables. Within land use science researchers have argued the importance of scale and levels. Nevertheless, the incorporation of multiple scales and levels and their interactions in one analysis is often lacking. Ignoring the hierarchical data structure originating from scale effects and levels may lead to erroneous conclusions due to invalid specification of the regression model. The objective of this chapter is to apply a multilevel analysis to construct a predictive statistical model for the occurrence of land use. Multilevel modelling is a statistically sound methodology for the analysis of hierarchically structured data with regression models that explicitly takes variability at different levels into account. For a land use study in the Philippines multilevel models are presented for two land use types that incorporate the field, household and village level. The value of multilevel modelling for land use studies and the implications of multilevel modelling for data collection will be discussed. The results show that explanatory variables can account for group level variability, but in most cases a multilevel approach is necessary to construct a sound regression model. Although land use studies often show clear hierarchical structures, it is not always possible to use a multilevel approach due to the structure of most land use datasets and due to data quality. Potentially, multilevel models can address many important land use issues involving scales and levels. Therefore, it is important in land use change research to formulate hypotheses that explicitly take scale and levels into account and then collect the appropriate data to answer these questions with approaches such as multilevel analysis.

Based on: Overmars, K.P., Verburg, P.H. 2006. Multilevel modelling of land use from field to village level in the Philippines. Agricultural systems 89, 435-456.

4.1 Introduction

In the past decade substantial advances has been made in land use and land cover change (LUCC) research by the development of a wide range of analytic tools to observe, explore and model LUCC (Lambin *et al*., 1999; Rindfuss *et al*., 2004, Veldkamp and Verburg, 2004). In general, LUCC is considered to be the result of the interplay between socio-economic, institutional and environmental factors, the so-called 'driving forces' of land use change. These driving forces are often subdivided into proximate causes and underlying causes. Proximate causes are the activities and actions that directly affect land use. Underlying causes are the fundamental processes that underpin the proximate causes, including demographic, economic, technological, institutional and cultural factors (Geist and Lambin, 2002). A widely used approach to explore the relations between land use (changes) and the underlying causes are regression techniques of various kinds (e.g. Nelson *et al*., 2001; Chomitz and Thomas, 2003; Perz and Skole, 2003; Verburg *et al*., 2004b). The approach in this chapter makes use of a regression technique that explicitly can deal with issues of scale and levels, which are characteristic for land use studies.

Within the LUCC discipline as a whole and in reference to regression approaches in particular, LUCC scientists have argued the importance of scale and levels (e.g. McConnell and Moran, 2001; Veldkamp and Lambin, 2001; Walsh *et al*., 2001; Nelson, 2002; Rindfuss *et al*., 2004). Gibson *et al*. (2000) state that scale is the spatial, temporal, quantitative, or analytical dimension used by scientists to measure and study objects and processes and level refers to specific locations along a scale. For this chapter the following definitions are used: Levels refer to organisational levels originating from social context, for example, household level, village level and municipality level and scale is used for artificial resolution and extent originating from a geographic representation of reality in maps. The following issues regarding scales and levels that are important in land use (change) analysis can be identified (Gibson *et al*., 2000; Verburg *et al*., 2004d). First, land use is the result of processes that act at different scales and levels, which ideally would be addressed simultaneously. The choices that are made in a study about the extent and the unit of analysis determine to a large extent what patterns will be observed and which correlation will be found. Often, these choices are different between disciplines (Verburg *et al*., 2003). Second, scale and levels are important in identifying relations, but the fact that a relation occurs at a certain scale or level does not explain the phenomenon. Therefore, causal statements between variables should be made explicit and tested. Within these causal statements scale and level are important factors, because different relations occur at different scales and levels. Moreover, causal relations can occur between different scales and levels. For example, village level variables like population or leadership capacity of the village head can influence land use at field level. Third, aggregation of processes to a higher level does not straightforwardly lead to a proper representation of these higher level processes because relations identified at the micro-level (or fine resolution) does not automatically translate into the same relation at the macro-level (or course resolution) (Robinson, 1950; Jones and Duncan, 1995; Easterling, 1997). The other way around the same phenomenon occurs: Inferences made on higher levels can often not be directly translated to lower level processes. Finally, all analyses, and therefore the insights from these analyses, are bounded by resolution or level of analysis and extent, which are determined by data structure and choices made by the researcher.

Mostly, scale and level issues are identified by comparing analyses at different resolutions and levels. Geoghegan *et al*. (2001) and Overmars and Verburg (2005) (Chapter 2) compared

an analysis of land use decisions based on a household dataset with an analysis using a spatial dataset. Walsh *et al*. (2001) and Veldkamp *et al*. (2001) analysed the relation between land use and its explanatory factors at different resolutions created by aggregating grid data. However, the incorporation of multiple scales and levels in the analysis and including interactions between levels is often lacking. So far, the statistical tools that explicitly deal with these issues are not often applied as noted by Pan and Bilsborrow (2005) and Polsky and Easterling (2001). Multilevel modelling, which is the approach used in this study, is one of the statistical tools that are potentially capable to integrate artificial scales and organisational levels and to include interactions between these scales and levels. Multilevel statistical modelling allows for the analysis of data with complex patterns of variability that originate from hierarchical structure (Snijders and Bosker, 1999).
Multilevel modelling has mainly been used in the social sciences, for example, in sociology, education, psychology, economics, criminology (Snijders and Bosker, 1999), and is becoming more popular in geographic applications (e.g. in studying transport and land values (Schwanen *et al*., 2004; Polsky and Easterling, 2001)). In most of these applications multilevel modelling is used to study the effects of social context on the individual behaviour and to study the confusion between aggregate and individual effects. Land use studies can potentially benefit much from multilevel analysis, because land use data often has a very clear hierarchical structure (e.g. administrative levels, agro-ecological divisions and subdivisions, societal levels, artificial scales). Therefore, it is remarkable that multilevel modelling is not (yet) widely applied in land use studies. Some land use studies do incorporate data from multiple levels, but only few actually use multilevel modelling (Hoshino, 2001; Pan and Bilsborrow, 2005).

This chapter aims to use multilevel analysis as the methodology to construct a predictive statistical model for the occurrence of land use that is statistically sound and which integrates different scales and levels. On the basis of a case study from a municipality in the Philippines different multilevel models will be presented that explain the occurrence of two major crops on individual fields in the area. In the discussion we explore and describe the (surplus) value of multilevel modelling for land use studies regarding the issues of scale and levels in LUCC research and describe the implications of multilevel modelling for data collection.

4.2 Multilevel analysis

In this section a short introduction of multilevel models is given in respect to land use issues in a general manner regardless of the outcome variable. Specific differences exist between models with a continuous, binary or multinomial outcome variable regarding estimation, model formation and the interpretation of coefficients. For the case study the logistic approach was adopted and the model specification is given in Section 4.3.3.
Multilevel analysis (e.g. Goldstein, 1995; Snijders en Bosker, 1999) is a methodology designed for the statistical analysis of hierarchically structured data. Multilevel regression models explicitly take the variability at different levels into account. Therefore, it is potentially a valuable tool in dealing with scaling issues in land use analysis. Multilevel modelling can address the scales and levels that are important to the land use system simultaneously, it can test hypothesis between scales and the modeller is not forced to aggregate or disag-

gregate data to one unit of analysis. Multilevel modelling can deal with nested data, such as hierarchically structured administrative units (e.g. farms in municipalities), as well as handle cases with observations that are structured differently, like lower level observations that are member of several groups at the higher level (e.g. farmers that have several buyers for their products).

Fundamental to multilevel modelling is "that the outcome variable Y has an individual as well as a group aspect" (Snijders and Bosker, 1999). This is reflected in the model by including explanatory variables at the individual level and at the group level, as well as in the way unexplained variation is modelled. Both unexplained variation within groups and unexplained variation between groups is conceived as random variation and is expressed in multilevel models as 'random effects'. Thus, multilevel models include an error term for every level in the model (Snijders and Bosker, 1999). Multilevel models can be constructed by including random intercepts only or by including both random intercepts and random slopes. Furthermore, variables can be added to the model to explain variability at the individual and group level, and also to explain the differences in slopes. For example, in a model with a household and a village level a random intercept can account for unobserved structural effects between villages. These structural effects may be caused by differences in technology. Including an explanatory variable "technology" could explain part of the structural effects. Random slopes actually incorporate differences between groups in the rate of change in output per unit change in the explanatory variable (*i.e.* the regression coefficients). For example, if you were predicting yields a random slope at the village level for soil fertility would account for differences between villages in the relation between soil fertility and yield, which may be caused by an unobserved difference in use of chemical fertiliser.

Multilevel models are applicable to data with hierarchical structures of various origins. Also for data that are acquired by using a multistage sampling scheme, and have therefore a hierarchical structure, a conventional regression model may be incorrect and a multilevel model would be a statistically sound method. In a multistage sampling design the selection of lower level observations depends on the choices made at higher levels. An example of a multistage sampling approach, when conducting a regional survey among land owners, is to first sample villages and than sample people within these villages. In this case the data at the lower level is not independent from the higher levels and therefore a conventional statistical approach might lead to underestimation of the standard errors (Rasbash *et al*., 2000). In any case, having some kind of hierarchy in the data, a multilevel analysis will model this hierarchy explicitly and prevent erroneous model inference.

If the multilevel structure of the data is ignored the data will inevitably be analysed at either an aggregate level or a disaggregate level. Analysing aggregated data, like in the work of Perz and Skole (2003), can only tell us something about the relation between macro-level variables. Analysing macro-micro or micro-level propositions with aggregated data may result in gross errors (Jones and Duncan, 1995) because by aggregating the data the variable changes in its meaning and cannot be used anymore to draw conclusions at the lower level. This phenomenon is called the ecological fallacy: A relation identified between macro-level does not automatically translate into the same relation at the micro-level (Robinson, 1950; Jones and Duncan, 1995; Easterling, 1997). A drawback of aggregation is that it disables the examination of cross-level relations, for example, when a micro-level relation differs by macro-level group or depends on a macro-level variable.

Disaggregation of macro-level data into micro-level data, by assigning the values of a few higher level observations to all lower level units, results in an exaggeration of the sample size. Wrongly assuming that all these observations are independent leads to an over-confidence in the estimated level of significance (due to underestimation of the standard errors), which in turn leads to elevated probabilities of a type I error when studying between group differences (type I errors: concluding there is a relation while in reality there is none). When studying within group differences it can result in failing to detect a relation (Snijders and Bosker, 1999; Rasbash *et al*., 2000; Polsky and Easterling, 2001).

4.3 Material and methods

4.3.1 Study area

The study area is situated in Cagayan Valley in the northeastern part of the island Luzon, the Philippines (Figure 4.1). The study area includes 20 villages (*barangays*) in the municipality of San Mariano, in the province of Isabela, and comprises approximately 480 km2. It is situated between the town of San Mariano in the west and the forested mountains of the Sierra Madre in the east. The population is approximately 20,000 persons (about 4,000 households) of various ethnic groups, among whom the Ilocano, Ibanag and Ifugao, who are all migrants or descendents of migrants that came to the area from the 1900s onwards, and the Kalinga and Agta, who are the indigenous inhabitants. Before immigration started, the area was completely forested with tropical lowland forest. At present, the study area shows a clear land use gradient ranging from intensive agriculture (mainly wet rice and yellow corn) near San Mariano via a scattered pattern of wet rice, yellow corn, banana, grasses, and (fruit) trees in the foothills to residual and primary forest in the eastern part. In the area a village unit actually consists of a group of settlements (*sitios*). The people live in

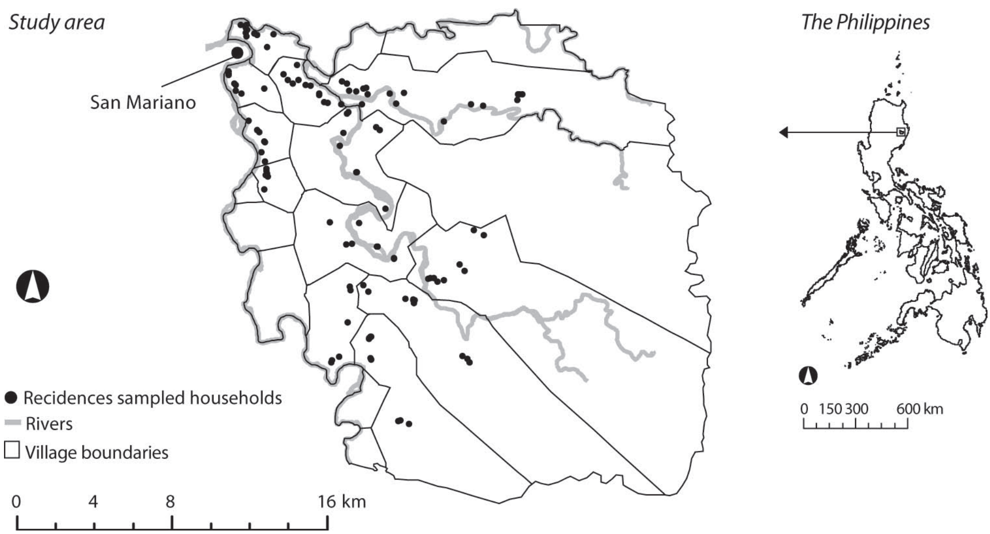


Figure 4.1: Location of the study area in the Philippines and the location of the households' homes within the area

these settlements, while their fields are often located in the surroundings of the settlement at an average distance of about 30 minutes walking.

4.3.2 Data

Data were collected in 13 of the 20 villages between June and November 2002 by interviewing households about their land use practices and household characteristics using a structured questionnaire. The questionnaire was designed to create an exhaustive list of variables that might explain land use decisions. This list of variables was based on literature, theories from a range of disciplines and expert knowledge of the area (see Overmars and Verburg, 2005 (Chapter 2) for more information). For the analysis in this study a subset of variables was used.

The selection of households to be interviewed was based on systematic random sampling using population data available at the POPMAT (POPulation Manipulation Action Team) member in the village. In all 13 villages every twentieth household was selected (systematic random sampling with sampling interval 20) from the POPMAT's list. From a total of approximately 3150 households in the 13 villages, 151 households were interviewed. The number of interviews per village ranges from 6 in the least populated village to 20 in the most populated. For the selected households the relevant characteristics were recorded for all fields (where a field is defined as a piece of land of a single owner used for one crop type). A household often owns or uses a number of fields at different locations and which are cultivated with different crops.

The most detailed (nested) hierarchy in the area, relevant to the land use system, could be constructed as follows (from the lowest level to the higher level): fields - plots (where a plot consists of a number of adjacent fields from the same owner) - households - *sitios* (the settlements) - villages - municipality. For the analysis only the field, household and village level were used (see Figure 4.2). This is the most functional grouping, because the plots consist mostly of only one field and the dataset does not contain enough observations to discriminate between *sitio* and household level. Most *sitios* have only one or two households within the sample, which is insufficient for a proper multilevel analysis. Each of the variables was collected at its corresponding level, e.g. soil characteristics and slope at field level and household structure at the household level. Village level variables were derived from census data of 1997 (data about ethnicity and the percentage of the population that is born in the municipality of San Mariano).

Records with missing data were omitted from the dataset. Table 4.1 presents the dataset as it was used in the analysis, which is a subset of the original dataset and includes the most relevant variables based on preceding research and field experience (Overmars and Verburg, 2005 (Chapter 2)).

4.3.3 Multilevel model specification

Multilevel models can be constructed in various forms with different levels of complexity. In this section we start with the description of a simple model to explain how we arrive at the model that we will use to explain the occurrence of land use. The description of the models is based on Snijders and Bosker (1999).

Since we will estimate a binary response variable (land use choice) we start with a conventional multiple logistic regression model (Equation 4.1).

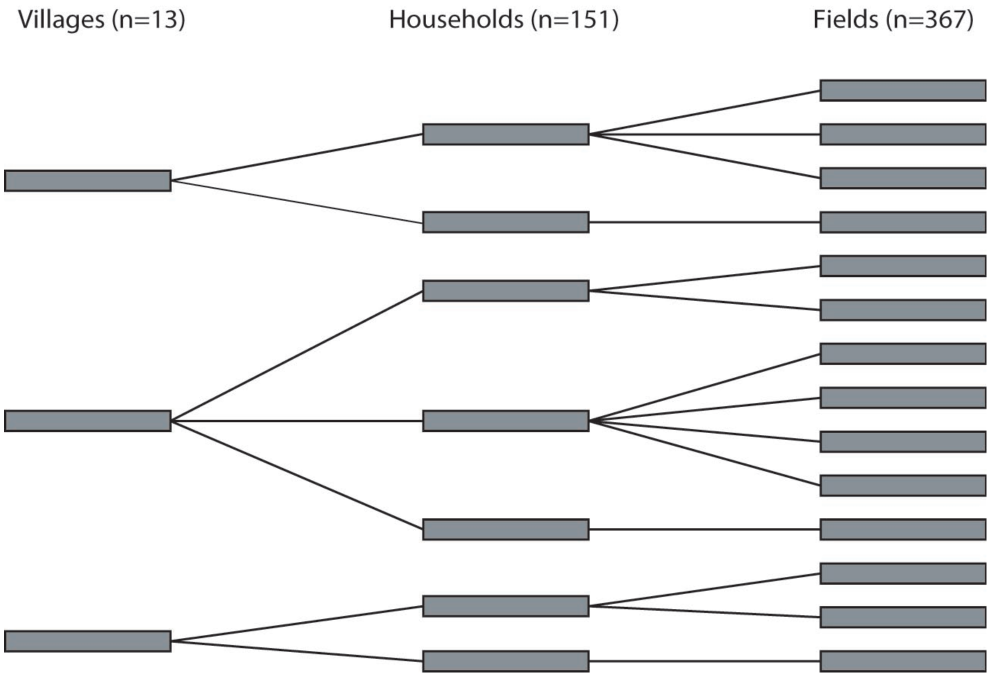


Figure 4.2: Schematic representation of the hierarchical structure of the dataset

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \quad (4.1)$$

In this model p is the probability for the occurrence of the event, which in this study is the occurrence of a land use type on a field, β_0 is an intercept, β_n are regression coefficients to be estimated, and the x_n are exogenous explanatory variables.
The simplest imaginable way to incorporate levels would be to identify explanatory variables at the lower and the higher level (Equation 4.2).

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n + \alpha_1 z_1 + \ldots + \alpha_m z_m \quad (4.2)$$

Here, β_n and α_m are regression coefficients to be estimated, and the x_n are exogenous explanatory variables at the lower level (e.g. field) and z_m are explanatory variables at the higher level (e.g. a household).
Actually, many studies apply this approach by including variables from different levels in the regression model (e.g. Müller and Zeller, 2002; Overmars and Verburg, 2005 (Chapter 2)) but do not report this explicitly. This model is typically called a fixed effect model since it lacks the random effects corresponding to higher levels in a multilevel model (Snijders and Bosker, 1999). The assumptions that belong to this model are that the residuals are mutually independent and have a zero mean. An additional assumption that is often made

Table 4.1: Description of the variables in the dataset used in this study

Variable name	**Description**	**Min.**	**Max.**	**Mean**	**St. dev.**
Dependent variables (field level: level 1, n=297)					
Yellow corn	1 if yellow corn, 0 otherwise	0	1	0.532	
Banana	1 if banana, 0 otherwise	0	1	0.215	
Independent variables at field level (level 1, n=297)					
Slope1	1 if slope category is flat, 0 otherwise	0	1	0.380	
Slope2	1 if slope category is flat to rolling/moderate, 0 otherwise	0	1	0.229	
Slope3	1 if slope category is rolling/moderate, 0 otherwise	0	1	0.283	
Slope4	1 if slope category is rolling/'moderate to steep/hilly, 0 otherwise	0	1	0.081	
Slope5	1 if slope category is steep/hilly, 0 otherwise	0	1	0.027	
Creek	1 if there is a creek or spring trough or bordering the plot, 0 otherwise	0	1	0.593	
Plot distance	Hours walking from the residence of the household to the plot (hrs)	0	10	0.511	
Independent variables at household level (level 2, n=115)					
Ethnicity Ilocano	1 if male household head is Ilocano (or Tagalog speaking), 0 otherwise	0	1	0.539	
Ethnicity Ifugao	1 if male household head is Ifugao, 0 otherwise	0	1	0.087	
Ethnicity rest	0 if ethnicity is Ilocano or Ifugao, 1 otherwise	0	1	0.374	
Transportation cost	Cost to transport a bag of corn from the residence to San Mariano (pesos)	7	45	22.652	12.214
Municipality of origin 0	1 if both male and female were not born in San Mariano, 0 otherwise	0	1	0.244	
Municipality of origin 1	1 if male or female head is born in San Mariano, 0 otherwise	0	1	0.322	
Municipality of origin 2	1 if both male and female were born in San Mariano, 1 otherwise	0	1	0.435	
Independent variables at village level (level 3, n=12)					
Ethnicity Ilocano (village)	Fraction of the population of the village that is Ilocano (or Tagalog speaking)	0.021	0.900	0.573	0.259
Ethnicity Ifugao (village)	Fraction of the population of the village that is Ifugao	0.000	0.404	0.076	0.147
Municipality of origin (village)	% of the population of the village born in San Mariano	64.899	99.007	84.479	9.748

is that all groups have the same variances (homoskedasticity assumption). Implicitly the assumption is made that all group structure is represented by the explanatory variables. If this is not the case the residuals will be heteroskedastic. A second problem with this approach is that the higher level data is often disaggregated to the lowest level. As said before, this will lead to type I errors. The following models describe how the effects of the different levels can be incorporated into the regression model. With these models the assumptions stated above can be tested.

The model in Equation 4.3 incorporates group effects but as yet without any explanatory variables. Besides the general intercept a random term U_{0j} is introduced, which is a group dependent intercept, in other words, an error term at the group level. With this random term the variance that exists between groups is modelled explicitly. The effect of being a 'member' of a specific group is taken into account. Introducing this term will help to prevent the residuals from being heteroskedastic.
For reasons of clarity indices mark the different levels: *i* for level 1, *j* for level 2 (and *k* for level 3) and a zero indicates that a parameter is not variable at that level.

$$\log\left(\frac{p_j}{1-p_j}\right) = \gamma_{00} + U_{0j} \tag{4.3}$$

In Equation 4.3 γ_{00} is the general intercept and U_{0j} is the group dependent deviation. The deviations U_{0j} are assumed to be independent and normally distributed with a zero mean and a variance of $\tau_0^{\,2}$ (Snijders and Bosker, 1999).
This model is called the 'pure random effects model', 'empty model' or 'unconditional model'. The empty model is a random intercept model without explanatory variables. With this model the variance of the dependent variable can be decomposed in a part caused by the individual level and a part caused by the group level (Snijders and Bosker, 1999; Polsky and Easterling, 2001). We will use this model as the base model to estimate if the group level variance in the dependent variable is significant. In the case study this is called model 1.

Including explanatory variables leads to the following model:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \gamma_{00} + \gamma_{10}x_{1ij} + \ldots + \gamma_{q0}x_{qij} + \gamma_{01}z_{1j} + \ldots + \gamma_{0r}z_{rj} + U_{0j} \tag{4.4}$$

where x_{qij} are q explanatory level-1 variables and z_{rj} are r explanatory level-2 variables. Again, the deviations U_{0j} are assumed to have zero mean (given the values of the explanatory variables) and a variance of $\tau_0^{\,2}$ (Snijders and Bosker, 1999). This model (Equation 4.4) is a random intercept model: a model where the intercept varies randomly between groups. The first part, $\gamma_{00} + \gamma_{10}x_{1ij} +\ldots+ \gamma_{q0}x_{qij} + \gamma_{01}z_{1j} +\ldots+ \gamma_{0r}z_{rj}$, is called the fixed part of the model and the second part, U_{0j}, is the random part of the model. In the case study analysis model 2, 3, 4 and 5 are based on this model (note that in the case study the model is extended to a model with 3 levels).
The interpretation of the regression coefficients is similar to ordinary logistic regression and is facilitated by the odds ratio (*exp*(γ)). The odds ratio can be interpreted as the change in odds for the considered event upon an increase of one unit in the corresponding factor,

while the other factors are considered to be unchanged. This means that the odds, *p/(p-1)*, are multiplied by *exp(γ)* for every unit increase of the variable corresponding to γ (Neter *et al*., 1996).
Starting from the empty model variables can be added at all levels. Variables at the individual level can explain part of the individual level variability as well as part of the group level variability, in the case when the values of the level one variable are consistently higher or lower than the general mean. For example, the slopes of the fields can be consistently higher in some of the villages and lower in others. Incorporating this field level variable can account for village level variability detected with the empty model.
Variables at the higher level(s) can be grouped in higher level variables that can only be observed at the higher level (e.g. the presence of a secondary school in a village) and aggregates of lower level variables (e.g. the average income of the inhabitants, which is an aggregate of observations at a lower level). Including these aggregates allows for the separation of the effect at the individual level and the effect at group level, which gives insight in the way a variable influences the outcome. In a model with only the level 1 data of that variable included the effect at both levels is forced to be equal (Snijders and Bosker, 1999). This difference is important while interpreting the regression coefficients. As described in Section 1 processes at the aggregate level can be substantially different from processes at the individual level. The village level variables in this study are of the aggregated type. Although they were calculated from census data, they have their equivalent at the household level.

The random intercept model (Equation 4.4) can be expanded by introducing random slopes. In a model with random slopes the regression coefficients (γ_{q0}) that act on the explanatory (level 1) variables are subdivided in a fixed and a random part. The addition of random slopes allows specific variables to differ by group. Even more complexity can be modelled by introducing level 2 variables in these slopes to explain (part of) the differences in slope. This is actually the same as a cross-product with an explanatory variable from level 1 and an explanatory variable from level 2. In multilevel modelling this cross-product is called cross-level interaction (Snijders and Bosker, 1999). In this study random slopes and cross-level interactions were not included in the models. This will be explained in greater detail in the final discussion.

In the case study models with three levels were applied (Equation 4.5), which is just an expansion of the model in Equation 4.4. The first model in the analysis is a pure random intercept model (empty model) with 3 levels. The subsequent models (models 2, 3, 4 and 5) are random effect models with three levels (Equation 4.5).

$$\log\left(\frac{p}{1-p}\right) = \gamma_{000} + \gamma_{100}x_{1ijk} + \ldots + \gamma_{q00}x_{qijk} + \gamma_{010}z_{1jk} + \ldots + \gamma_{0r0}z_{rjk} + \gamma_{001}a_{1k} + \ldots + \gamma_{00s}a_{sk} + R_{0jk} + U_{00k} \tag{4.5}$$

In Equation 4.5 the a_{sk} are s explanatory level 3 variables, the R_{0jk} is the level 2 random part and U_{00k} is the level 3 random part. In this model fields are the unit of analysis at level 1, level 2 consists of households and level 3 are the villages. The dependent variable Y is land use. If Y = 1 the land use occurs, if Y = 0 the land use does not occur and p is the probability that the land use is found on that field.

Two analyses will be presented: one explaining the occurrence of yellow corn and one explaining the occurrence of banana. These are the most dominant crops in the study area (53 % of the fields were cultivated with corn and 22 % with banana). In the analysis we present five different random intercept models per land use type. The first model is the empty model, which informs about the variability at the different levels. In the subsequent models variables will be added per level to see the influence of these groups of variables on the variance component of the higher levels.
The variables included were selected by studying prior analyses (Overmars and Verburg, 2005 (Chapter 2), Overmars *et al*., 2006 (Chapter 3)) and field experience. For the corn model variables from the following list were added in different compositions: slope, creek and plot distance at the field level; transportation cost, ethnicity and municipality of origin at the household level; and averages of municipality of origin and ethnicity at the municipal level. For the banana model the same variables were used except for presence of creeks, because this was considered to be of no influence to the occurrence of banana.
The analysis is performed with HLM software (Raudenbush *et al*., 2004). All models were estimated using the PQL (Penalized Quasi likelihood) routine. In HLM6 all 3-level hierarchical generalised linear models are estimated by full PQL by default (Snijders and Bosker, 1999; Raudenbush *et al*., 2004).

To indicate the proportion of variance that is accounted for by the group level the intraclass correlation coefficients (ρ_R and ρ_U, for the household and village level, respectively) are calculated. Equation 4.6 shows the calculation of the intraclass correlation coefficient for the household level. (Snijders and Bosker, 1999; Browne *et al*., 2005).

$$\rho_R = \mathrm{var}(R_{0jk}) / (\mathrm{var}(R_{0jk}) + \mathrm{var}(U_{00k}) + \pi^2 / 3) \qquad (4.6)$$

Where ρ_R is the intraclass coefficient for the household level, $var(R_{0jk})$ is the variance of the random intercept at household level and $var(U_{00k})$ is the variance of the random intercept at village level. A logistic distribution for the level one residual implies a variance of $\pi/3$, which appears as the level 1 variance in Equation 4.6 (Snijders and Bosker, 1999). In an linear multilevel model this would be the level 1 variance σ^2.
To assess the goodness-of-fit of the models the ROC (Relative Operating Characteristic) (Swets, 1988) was used . This measure is capable to assess the quality of the predictor and can be compared between different models. The ROC summarises the performance of a logistic regression model over a range of cut-off values classifying the probabilities. The value of the ROC is defined as the area under the curve linking the relation between the proportion of true positives versus the proportion of false positives for an infinite number of cut-off values. The ROC statistic varies between 0.5 (completely random) and 1 (perfect discrimination).

4.4 Results

4.4.1 Corn models

This section presents various multilevel models, with different sets of explanatory variables, predicting the occurrence of yellow corn on a field. Model 1 is the empty model, which

does not include any explanatory variables, but only includes random effects at the higher levels. Model 1 (Table 4.2) shows that the variance is significant (*p<0.05*) at both level 2 and 3. The intraclass correlation coefficients (ρ_R and ρ_U, Table 4.2) indicate that 10 percent of the variance can be attributed to the household level and 4 percent to the village level. The remaining variance is in level 1, which is fixed in this modelling approach to $\pi/3$. Thus, both the households and the villages show significant clustering of the occurrence of corn. The variance detected in this model might be accounted for by explanatory variables. This is studied with the models 2, 3, 4 and 5.
Model 2 introduces a set of geographic and biophysical variables that are known explanatory variables for the occurrence of corn in the study area. These are slope, presence of a creek, hours walking from the residence of the household to the plot, and the cost to transport a bag of corn from the residence to San Mariano.
Table 4.2 (model 2) shows that almost all explanatory variables (the fixed effects) have significant coefficients. Corn is more likely to occur on field that are flatter, not close to a creek, close to the household's residence and close to the market town of San Mariano. The random part of level 3 turns out to be lower. So, these variables explain some of the variance at the village level detected in the empty model. This might be caused, for example, by the fact the transportation costs vary on average per village because the villages are situated at different distance from the market place. By introducing this variable (or perhaps one of the other variables) the variability disappeared from the village level. Although the level 3 random part is not significant and, theoretically, level 3 could be excluded, the structure with three levels is maintained in order to study the level 3 behaviour in the following models. At the household level the variance component is still significant and similar to the variance component of the empty model. Thus, the variables included do not account for any of the household level variability.
Model 3 adds household variables to model 2. Additional to the relations in model 2 corn turned out the be negatively related with households where both the male and female are born outside the municipality and negatively with people of Ilocano origin. After including the household level variables the random part of level 2 (the household level) decreased substantially. Apparently, the variance at level 2 is captured by the included variables. The geographical/biophysical variables are still significant. The level 3 variance increased in comparison with model 2.
Model 4 adds the village level variables to the model 3 configuration. This model investigates if there is a fixed effect of the village level variables besides the variables included in model 3. For example, one can imagine that a village dominated by one ethnic group has an extra village level effect besides the effect of ethnicity at household level for the whole study area. The village level variables are aggregated values of variables at level 2 (ethnicity and municipality of origin). Instead of using the survey data to derive these level 3 variables census data of the complete population was used. Table 4.2 shows that there are no significant effects for the variables at village level. Including these variables results in a similar random part at the village level as model 3. Thus the level 3 variables did not explain any of the variance in level 3.
Model 5 was constructed to see if including the household variables at village level instead of at the household level would be a good alternative. This would be convenient because census data at village level is often more easily available then household level data. However, like in model 4, none of the village level variables are significant in model 5. Besides that, the variance component at household level is the same at in the models 1 and 2. This

Table 4.2: Multilevel models for yellow corn

Yellow corn	Model 1	Model 2	Model 3	Model 4	Model 5
Fixed effects					
Level 1					
Intercept	0.201	-0.546	-0.217	0.330	-1.083
Slope1		3.108**	3.688**	3.572**	3.113**
Slope2		3.446**	4.091**	4.007**	3.397**
Slope3		2.274	2.768*	2.657*	2.281
Slope4		-0.469	-0.560	-0.629	-0.603
Creek		-0.833**	-0.759*	-0.742*	-0.843**
Plot distance		-0.586*	-0.599*	-0.616*	-0.569*
Level 2					
Transportation cost		-0.050**	-0.050**	-0.055**	-0.038*
Ethnicity Ilocano			-0.929*	-0.973*	
Ethnicity Ifugao			-0.997	-1.094	
Municipality of origin 0			-1.313**	-1.347**	
Municipality of origin 1			0.233	0.239	
Level 3					
Municipality of origin village				-0.744	0.845
Ethnicity Ilocano village				0.314	-0.384
Ethnicity Ifugao village				0.283	-0.781
Random effects					
Level 2					
var (R_{0jk})	0.395*	0.441**	0.001***	0.004***	0.494**
ρ_R	0.103	0.115	0.000	0.001	0.130
Level 3					
var (U_{00k})	0.143*	0.103	0.187*	0.177*	0.018
ρ_U	0.037	0.027	0.054	0.051	0.005
ROC	0.855	0.881	0.864	0.863	0.882

** p<0.05, ** p<0.01, *** p<0.001*

shows that the aggregated variables do not capture any of the variability at the household level. As theory suggests (Robinson, 1950; Jones and Duncan, 1995) the effect of the aggregate variable is quite different than that of its lower level equivalent.

The ROC value of the corn model 1 is 0.855. The ROC value of model 2 is about the same as model 1. This indicates that including the variables at field level does not lead to better predictions, because then the ROC would be higher if they did explain field level variance. However, the variables included in model 2 do explain part of the village level variance, which is showed by a lower variance component at the village level and significant regression coefficients.

4.4.2 Banana models

The analysis of the occurrence of banana shows a different result (Table 4.3). In the empty model (model 1) there are no signs of significant between-group variances. Model 2, which incorporates the geographical/biophysical variables, shows a significant relation between slope of a field and the choice to cultivate banana and a significant random effect at the village level. Thus, including variables results in a large and significant random part at the village level. This could not be explained completely. Part of the explanation is that in general changes in the fixed effects part can cause big changes in the random part while changes in the random part usually do not cause big changes in the fixed effect part. The mean structure the model can change the stationarity of the mean causing a shift in variance and making the random effects part significant.

Model 3 introduces household level variables ethnicity and municipality of origin, model 4 includes also village averages of these variables and model 5 includes only the village level and field level variables. All the coefficients of these variables do not differ significantly from zero, neither do they influence the level 2 and level 3 random parts significantly. Therefore, we conclude that these variables do not influence banana cultivation significantly and that this is predominantly determined by slope. To find out what process might cause the differences between villages the random intercepts of the village level were examined. This did not show a clear pattern. Furthermore, models with additional explanatory variables and models with random slopes were tested , but this did not result in a satisfying explanation of the variability at the village level in model 2.

The ROC of banana model 1 is 0.694. Model 2 has an ROC of 0.906. This indicates that the slope of the fields does explain part of the variance at the field level. Including variables in model 3, 4 and 5 does not produce a higher ROC than model 2, which is obvious, because in the model 2 the random part is included in the predicted values and no additional level 1 variables are included. The two random parts accounts for all variance at level 2 and 3. The difference between model 2 and models 3 ,4 and 5 is that the variables at household and village level can explain part of the variance. However, in this model the explanatory factors at household and village level are not significant and the variance of the random part is similar throughout models 2, 3, 4 and 5.

4.5 Discussion and conclusions

4.5.1 Multilevel statistics for land use studies

In this section the main findings of the multilevel analysis are discussed for the case study. Then, these findings are used to evaluate the advantages and disadvantages of multilevel analysis for land use studies in general.

For corn cultivation the empty model indicated significant between-group variability at two higher levels (household and village). Explanatory variables at the household level turned out to account for that variability at that level (Table 4.2, model 3). Replacing some of the household level variables with their village level aggregates did show a significant variance component at the household level. From this it can be concluded that the household level variables cannot be substituted by village level aggregates in this case. The explanatory variables at the household level can explain a significant part of the occurrence of corn at

Table 4.3: Multilevel models for banana

Banana	**model1**	**model2**	**model3**	**model4**	**model5**
Fixed effects					
Level 1					
Intercept	-1.289***	-3.430***	-3.260**	-4.050	-4.120
Slope3		2.389***	2.435***	2.432***	2.397***
Slope4		5.022***	5.006***	4.975***	4.970***
Slope5		5.634***	5.971***	5.765***	5.461***
Plot distance		0.006	0.007	0.023	0.020
Level 2					
Transportation cost		0.015	0.012	0.017	0.019
Ethnicity Ilocano			-0.297	-0.236	
Ethnicity Ifugao			-0.613	-0.500	
Municipality of origin 0			0.532	0.528	
Municipality of origin 1			-0.242	-0.247	
Level 3					
Municipality of origin village				1.253	1.198
Ethnicity Ilocano village				-0.380	-0.414
Ethnicity Ifugao village				-0.055	0.114
Random effects					
Level 2					
var (R_{0jk})	0.003	0.006	0.003	0.004	0.006
ρ_R	0.001	0.002	0.001	0.001	0.002
Level 3					
var (U_{00k})	0.107	0.546**	0.724**	0.637***	0.487**
ρ_U	0.031	0.142	0.180	0.162	0.129
ROC	0.694	0.906	0.909	0.908	0.903

** p<0.05, ** p<0.01, ***p<0.001*

field level. The exploratory procedure applied has revealed at which levels important variables explain land use decisions. Model 3 and 4 show that within these models a significant part of variability is left at the village level, which is left unexplained in these models.
The empty model with banana as dependent variable did not show any significant variance component. However, after introducing the variables slope and transportation costs as explanatory variables (Table 4.3, model 2) the village level variance component is significant ($p<0.01$). The variability at village level could not be accounted for by any of the explanatory variables used in this study. The question what causes the village differences in this model will therefore remain unanswered. The village level variability might be caused by differences in soils or geomorphology, though these variables were not included in this study. The results for both the corn and banana models indicate that a conventional regression model would not be correct because the residuals would be heteroskedastic. The multilevel structure accounts for the unobserved effects between villages and provides a statistically correct model.

The multilevel analysis of the land use system in the study area provided additional information to previous analyses based on conventional regression models (Overmars and Verburg, 2005 (Chapter 2)). In the case of corn the analysis confirmed the hypothesis that the household level plays an important role. In this analysis the municipality of origin of the household (which is proxy for migration history) in combination with the ethnicity variables turned out to be a significant explanatory variables that account for the variability at the household level. For the case of banana the analysis confirmed the idea that household level characteristics did not play an important role in the decision to cultivate bananas. Bananas occur mostly on sites that are less productive or too steep for arable crops like corn, rice or vegetables. Significant village level variance indicates the importance of village level conditions in explaining the decision to cultivate banana. The ROC values of the analyses of Chapter 2 and this analysis cannot be compared straightforwardly because the multilevel analysis incorporates a random part that contributes to the value of the ROC without actually explaining the dependent variable since this random part is fitted.
Like in any other statistical analysis, drawing conclusions about the causality of the relations from the regression analysis should be done with care. For example, the positive relation between slope of a field and bananas results from the fact that the flatter areas are devoted to arable crops, not because bananas perform better on the steep slopes. Like any other regression analysis multilevel models can only reveal associations between variables and partition variance. Additional research is needed to study the causality of the relations. An example of such a method for the case study area is described in Overmars *et al*. (2006) (Chapter 3).
In this chapter random slopes were not incorporated in any of the models. Although there were no strong arguments suggesting that the coefficients for the explanatory variables were different, some experiments were carried out to study the behaviour of models that include random slopes. This resulted in either insignificant random slopes or models that did not converge. Most likely the data structure and the amount of observations made the estimation of the random slopes complicated. The number of observations (fields) per household is low and this complicates the determination of the random slopes.

The results indicate that the household level can be crucial in explaining land use at the field level. However, in many studies household level data are not available because in many regional studies the analysis is based on remote sensing, maps and census data (e.g. Nelson *et al*., 2001; Walsh *et al*., 2001; Müller and Zeller, 2002). As this study shows simply substituting household level variables with their village level equivalents, which can be calculated from widely available census data, will most often not account for the household level variability because of errors due to aggregation. Disaggregating higher level variables to the level of analysis can lead to erroneous conclusions. In any case, disregarding the household level variables while explaining land use at field level ignores the conclusion of Rindfuss *et al*. (2003) that the household level is the central level to be included in explanations of land use.
Data availability and data structure play an important role in land use studies. As illustrated in this chapter, data availability determines at what level land use can be studied, and therefore at what level one can draw conclusions. If the hierarchical structure of the data is important to the land use system under study and the research questions that arise from this, this structure should be considered in the sample design to take full advantage of the multilevel modelling technique. Ideally, at every level a sample is drawn that is

representative for the population at that level. For the highest level, one should keep in mind that a small sample size cause the same difficulties as an ordinary regression with that sample size (Snijders and Bosker, 1999), *i.e.* small sample sizes have less power than larger samples. For the lowest level, which is the unit of analysis, the number of observations per group (e.g. the number of fields per household) should be enough to estimate the parameters that are included in the model.

Datasets that were not designed for multilevel modelling often appear to be inadequate. This is a serious constraint for applying multilevel modelling in land use studies, because many studies use available datasets. In studies with levels other than farmers and fields, for example including country and sub-country level, the data structure can be more favourable to multilevel modelling.

In the dataset used in this study the number of observations (fields) per household was very low, but this is inherent to the structure of the land use system, because the farmers have only a few fields. At the village level only 12 observations were present, but this is the complete population in the study area (*i.e.* one village was kept out of the analysis due to missing data). This data structure provides relatively few degrees of freedom for multilevel modelling and may have hampered the estimation of random slopes, which were therefore not included in the models presented. Polsky and Easterling (2001) have a similar experience in estimating a multilevel model based on 446 counties nested within 57 districts. To deal with small sample sizes one might consider to use bootstrap or MCMC (Markov Chain Monte Carlo) approaches, which are available in MLwiN (Rasbash *et al*., 2000), for example.

Verburg *et al*. (2004d) emphasise the importance of multi-scale approaches and cross-scale dynamics and name multilevel modelling as a potential approach that can deal with scale issues in land use studies. Multilevel modelling can address a variety of these issues. First of all, the multilevel approach explicitly includes different levels. These levels can be, for example, organisational levels of the land use system or nested administrative units, but can also be artificial aggregations of a grid. Where in other studies the effects of scale on the observed relations between land use and driving factors were studied by the separate analysis at different organisational levels or by (dis)aggregating grids to one level of analysis (e.g. Verburg and Chen, 2000; Walsh *et al*., 2001; Overmars and Verburg, 2005 (Chapter 2)), the multilevel approach is capable of incorporating different levels of aggregation within one model and exploring the contributions of the various levels.

Secondly, within the multilevel approach cross-scale dynamics can be modelled as cross-level interactions. A cross-level interaction can be defined as dependence of a relation between two micro variables on a macro-level variable (Snijders en Bosker, 1999). A difference with conventional models is that when including the cross-level interaction the slopes parameters also have a random effect. An additional option in a multilevel approach is to include group level aggregates of variables. This clearly separates level 1 effects from higher level effects, which can be completely different.

Another important aspect to consider in land use studies is spatial dependency, which refers to the geographic law that nearby things are more related then distant things (Tobler, 1970). Spatial dependency in land use patterns can be caused by dependence of the land use pattern on an explanatory factor that is spatially structured (trend) or a spatial interaction process of the land use variable itself, like competition or imitation (Anselin, 1988; Irwin and Geoghegan, 2001; Overmars *et al*., 2003; Polsky, 2004). Both Polsky and Easterling

(2001) and Pan and Bilsborrow (2005) mention that multilevel modelling can partly reduce the effect of spatial autocorrelation when neighbouring observations are nested within one group. If the spatial dependency is only related to the nested hierarchy this might even correct for all spatial autocorrelation. However, often spatial dependency is structured differently than the nested hierarchy of the dataset. In this case the neighbourhood effects can be incorporated in the multilevel model as cross random effects (where lower level observation can be member of different groups at the higher level). For example, each observation can be part of a group with all its neighbours. This approach would correct for spatial autocorrelation but is not yet studied in land use research. In this study this approach was not applied because the observed fields are relatively far apart due to the relatively small sample size and spatial autocorrelation is therefore assumed to be minimal.

4.5.2 Conclusions

The case study has shown that multilevel analysis can be applied statistically to model the occurrence of land use. We consider multilevel modelling to be a relevant tool for land use studies because organisational levels and spatial and temporal scale dependencies are characteristic for land use data. Multilevel modelling offers a method to study the influence of these levels and scales as well as great flexibility in testing hypothesis on explanatory variables and their cross-level interactions and spatial dependencies. Multilevel regression modelling is considered to be a statistically sound method to create regression models when analysing hierarchically structured data. Including random parts in the model ensures correct estimates of the regression parameters and their significance levels. However, so far, few scholars have applied this approach in land use studies. This might have to do with data quality and data availability. Another cause can be that the methodology is only recently developed. Currently, multilevel software is becoming more generally available (see Centre for Multilevel Modelling (2005) for a detailed review) which might promote the use of multilevel models in land use change studies.

In recent LUCC literature many have advocated for explicit attention for scale issues in LUCC research (e.g. McConnell and Moran, 2001; Veldkamp and Lambin, 2001; Rindfuss *et al.*, 2004). From this study it can be concluded that it is indeed important explicitly to identify and report on the levels that are present in the study. Levels that are crucial in explaining the land use system should be included in modelling exercises. Moreover, the propositions that are studied should indicate more explicitly to which scales and levels they apply. Potentially, a multitude of propositions can be formulated that involve scale and level, like micro-micro, macro-macro, micro-macro, macro-micro and multi-level propositions. To be able to test these hypotheses it is important to collect adequate data to enable the application of a multilevel approach in order to answer questions that are inherently hierarchical in reference to land use studies. Multilevel modelling is a useful addition to the land use research toolbox that allows the exploration of a number of cross scale propositions.

5

Comparison of a deductive and an inductive approach to specify land suitability in a spatially explicit land use model

Abstract

In this chapter, two research approaches to specify the relation between land use types and their explanatory factors are applied to the same modelling framework. The two approaches are used to construct land suitability maps, which are used as inputs in two model applications. The first is an inductive approach that uses regression analysis. The second applies a theoretical, actor decision framework to derive relations deductively using detailed field data. Broadly speaking, this classification coincides with the distinction between empirical and theoretical models and the distinction between deriving process from pattern and pattern from process. The two modelling approaches are illustrated by a scenario analysis for a case study in a municipality in the Philippines. Goodness-of-fit of the deductive approach in predicting current land use is slightly lower compared to the inductive approach. Resulting land use projections from the modelling exercise for the two approaches differ in 15 percent of the cells, which is caused by differences in the specification of the suitability maps. The chapter discusses the assumptions underlying the two approaches as well as the implications for the applicability of the models in policy-oriented research. The deductive approach describes processes explicitly and can therefore better handle discontinuities in land use processes. This approach allows the user to evaluate a wide range of scenarios, which can also include new land use types. The inductive approach is easily reproducible by others but cannot guarantee causality. Therefore, the inductive approach is less suitable to handle discontinuities or additional land use types, but is well able to rapidly identify hotspots of land use change. It is concluded that both approaches have their advantages and drawbacks for different purposes. Generally speaking, the inductive approach is applicable in situations with relatively small land use changes, without introduction of new land use types, whereas the deductive approach is more flexible. The choice of modelling approach should therefore be based on the research and policy questions for which it is used.

Based on: Overmars, K.P., Verburg, P.H., Veldkamp, A. 2006. Comparison of a deductive and an inductive approach to specify land suitability in a spatially explicit land use model. Land Use Policy (Accepted).

5.1 Introduction

Within LUCC (land use and land cover change) research much attention has been paid to the development of models (Briassoulis, 2000; Veldkamp and Lambin, 2001; Parker *et al*., 2003). Land use models are used as a tool to combine different aspects of the complex land use system and therefore enable researchers to study the dynamics of this system. Furthermore, land use change models are applied to evaluate scenarios to inform policy makers (Brown *et al*., 2004; Solecki and Oliveri, 2004).
In reviewing land use models many criteria have been used to classify models: for example, whether a model is economic or non-economic, spatially explicit or not or whether the model is statistical/empirical, mathematical or rule-based (Briassoulis, 2000; Brown *et al*., 2004; Verburg *et al*., 2004d). Most of the current land use models have in common that they all try to combine human and natural processes, which implies the involvement of various disciplines (Couclelis, 2001). In this chapter we will use the broad distinction between deductive and inductive approaches of modelling (e.g. Laney, 2004; Overmars *et al*., 2006 (Chapter 3)). Broadly speaking, this classification coincides with the distinction between theoretical and empirical models and the distinction between deriving pattern from process and process from pattern.
Overmars *et al*. (2006) (Chapter 3) identify six types of modelling, which vary from completely deductive to completely inductive. In this study two of these types will be used to specify the relation between land use and its explanatory factors, which will be implemented in two applications of a spatially explicit land use model in the same region. The first approach can be classified as 'unstructured factors induction'. In this approach a conceptual framework is used to define the dependent variable and the independent variables but then leave it to the procedures of statistical inference to find correlations between these variables. Theories are used to construct hypotheses about the relation between land use and its explanatory factors, but the structure of these theories is not used or tested (e.g. Serneels and Lambin, 2001; Nelson *et al*., 2004). The second, more deductive approach used in this chapter is called 'imposed theory'. In this approach a land use theory is specified for a real world case in terms of both structure and parameters, without any fitting to empirical data, and used to predict land use.
The two approaches to quantify the relation between driving factors and land use, resulting in a land 'suitability' estimate, will be implemented in two applications of CLUE-S, which is a dynamic land use model, to simulate scenarios of LUCC in a study area in the municipality of San Mariano in the northern part of the Philippines. The remainder of the model setting will be kept the same for the two modelling approaches to be able purely to assess the effect of having different methods to specify land suitability.

The aim of this chapter is to compare the differences between the two model applications, which have different specifications of land suitability as input. The difference in outcome of two model applications as well as the different assumptions underlying the two model specifications will be discussed. Furthermore, the chapter describes the implications for the applicability of the approaches for different research and policy questions.

5.2 Study area and data collection

5.2.1 Study area

The study area is situated in Cagayan Valley in the northeastern part of the island Luzon, the Philippines (Figure 5.1). The study area includes 16 *barangays* (villages) in the municipality of San Mariano, in the province of Isabela, and its size is approximately 26,000 ha. It is situated between the town of San Mariano in the west and the forested mountains of the Sierra Madre mountain range in the east. The area is inhabited by approximately 16,500 people (about 3,150 households) of various ethnic groups, among whom the Ilocano, Ibanag and Ifugao, who are migrants or descendents of migrants that came to the area from the 1900s onwards, and the Kalinga and Agta, who are the indigenous inhabitants of the area. At present, the study area shows a clear land use gradient ranging from intensive agriculture, with mainly rice and yellow corn, near San Mariano to a scattered pattern of rice, yellow corn, banana, grasses and trees to residual and primary forest in the eastern part of the study area. Before immigration started the area was completely covered with tropical lowland forest. About 76 percent of the population has farming as their main source of livelihood and another 12 percent is involved in working on other people's farms.

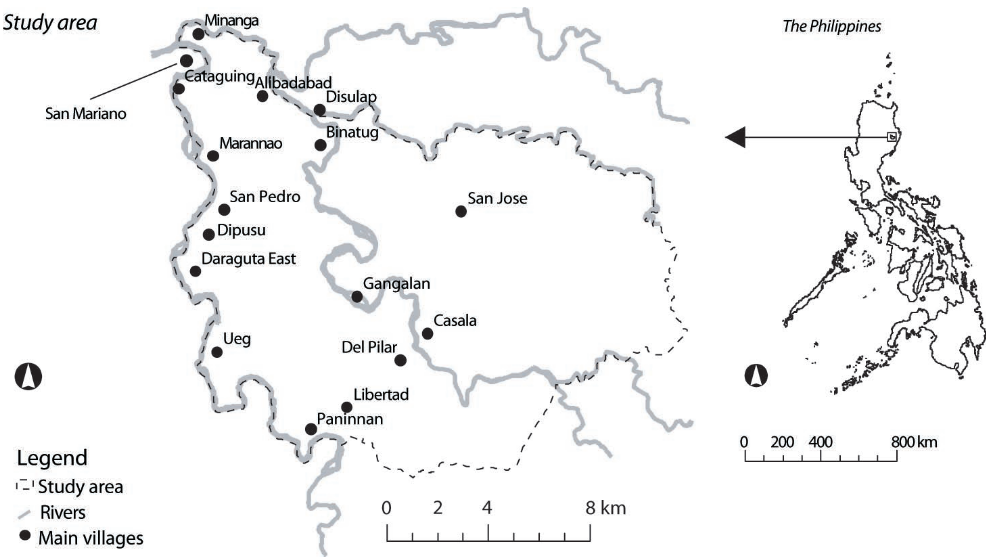


Figure 5.1: Study area

5.2.2 Data

Land use data were interpreted from two remote sensing images, a Landsat ETM+ image (http://www.landsat.org) from June 2001 and an ASTER image from March 2002. First, unsupervised classifications of both images were made for the study area. Second, the classes of the unsupervised classifications were recoded into land use types according to a set of 96 field observations. Finally, the land use map was constructed by combining the classifications of the two images. In this procedure the ASTER image was first resampled from 15m resolution to the same grid as the Landsat image (30 by 30 m). Then, all land use classes of the two images were put in separate layers. In a GIS (Geographical Information System) these layers were combined, using overlay, where the delineation of the top layers overrule those of the layers underneath, in order to obtain the best fit with the field observations. This way the best elements of two images were combined and the best overall land use classification was created. To improve the identification of wet rice fields an extra SPOT image from July 2001 was used. Finally, the image was resampled (aggregated) to a 50 by 50 m grid that coincides with the other data. Resampling was performed by taking the value of the original map under the centre point of the newly created grid. Classification accuracy of the land use map is 68 percent, which was calculated using an independent sample of 76 field observations (Verburg *et al*. 2004a).

In creating the land use map from remote sensing images banana plantations and low-density forest types (secondary forest) were difficult to separate from each other, because banana cultivation is quite extensive and often many trees grow in between the bananas, which results in a similar spectral reflection as secondary forest. Therefore, a class that included both banana and secondary forest was manually divided based on field observations into a part with predominantly banana and a part with predominantly secondary forest. The western half of the area was identified as an area in which this class can be considered to contain almost exclusively extensive banana plantations. In the eastern part the same class is considered to be predominantly secondary forest. The resulting land use map is depicted in Figure 5.5A.

The set of explanatory variables is based on a previous analysis (Overmars *et al*., 2006 (Chapter 3)) and includes slope, ethnicity variables, accessibility variables, potential for rice and a reforestation policy. The slope map was derived from a 1:50,000 topographic map of the area (NAMRIA, unknown). This slope map was reclassified into five slope classes that correspond with classes in the survey held amongst farmers in the area, which was used in the deductive approach. It was not possible to obtain a map that depicts the ethnicity of the individual landowners, because no data were available that link all land managers to their individual parcels. Instead, maps of the percentage of every tribe per village were created based on census information of the National Statistics Office. The two accessibility measures in this study are based on an in-depth study on accessibility in the study area (Verburg *et al*., 2004a). The time farmers have to travel from their homes to their fields is calculated with a cost distance algorithm. In this calculation different travel speeds were attributed to different types of roads and off road and these were used to calculate the minimum travel time. Transportation costs are calculated by assigning the transportation costs (to the market place in San Mariano) of the nearest village, based on the travel time calculation, to all locations. A map with the possibilities for irrigation to cultivate wet rice was constructed from a map indicating the area within 200 m distance to a creek (excluding big rivers) and a map indicating the land that can potentially be served by a NIA (National Irrigation Agency) project that was established in the area. These rules were combined to a map containing location with and locations without the possibility of cultivating rice. The final data source is a map delineating an area which is targeted by a policy called SIFMA (Socialized Industrial Forest Management Agreement), which promotes the planting of trees (DENR-CENRO, 1998). Within this policy farmers were offered 25 years of tenure rights on the condition that they plant a certain area with (fruit) trees (mainly mango, citrus and coconut). In the study area this policy was especially promoted by an NGO that provided free seedlings and assisted the farmers in obtaining the tenure documents (General, 1999).

Table 5.1: Descriptive statistics based on a 5 % sample of the complete dataset, n = 5002

Variable name	Description	Min.	Max.	Mean	St. dev.
Land use variables					
Wet rice	1 if wet rice, 0 otherwise	0	1	0.02	
Yellow corn	1 if yellow corn (and 10% other arable crops), 0 otherwise	0	1	0.21	
Banana	1 if banana, 0 otherwise	0	1	0.17	
Grass	1 if grass 0 otherwise	0	1	0.30	
Sec. forest	1 if secondary forest 0 otherwise	0	1	0.17	
Forest	1 if forest 0 otherwise	0	1	0.09	
Water bodies	1 if lake or river, 0 otherwise	0	1	0.02	
Explanatory variables					
Slope1	1 if slope < 2.5 degrees	0	1	0.22	
Slope2	1 if 2.5 ≤ slope < 6.5 degrees	0	1	0.25	
Slope3	1 if 6.5 ≤ slope < 12.5 degrees	0	1	0.34	
Slope4	1 if 12.5 ≤ slope < 20.5 degrees	0	1	0.16	
Slope5	1 if slope ≥ 20.5 degrees	0	1	0.04	
Dist. to small river	1 if distance to a small river < 200 m or part of NIA irrigation project	0	1	0.26	
Plot distance	Minutes walking to the plot (min.)	0	405	76.63	73.02
Transportation cost	Cost to transport a bag of corn from the house to San Mariano (pesos)	0	45	25.50	10.18
Ethnicity Ilocano	% in the barangay that is Ilocano	2.09	96.6	67.95	22.53
Ethnicity Ifugao	% in the barangay that is Ifugao	0	40.42	7.07	12.72
Ethnicity Ibanag	% in the barangay that is Ibanag	0	89.47	13.06	16.67

5.3 Methods

5.3.1 Overview

In this section the inductive and deductive approach to derive the relation between land use and its explanatory factors (*i.e.* 'suitability' maps) are presented. The inductive model, using logistic regression analysis, is rather straightforward. The deductive approach uses an actor decision framework. This approach is less known than the inductive approach and is therefore described in more detail. Both approaches make use of the dataset described in Table 5.1. So, differences between the results of the two approaches cannot arise from differences in the specification of variables. However, as will be explained, the two approaches differ in their model specification, for example, they use a different selection of variables from this dataset and different model parameters. Moreover, the deductive approach additionally includes variables that are constant over the area (e.g. prices and investments levels). This will result in different outcomes of the two approaches.
The resulting suitability maps of the two approaches are input to two different applications

of CLUE-S. CLUE-S is a spatially explicit and dynamic land use model, which is described below. The suitability maps produced with either the inductive or the deductive approach provides only one of the mechanisms that are responsible for land use distribution in the CLUE-S model. The other mechanisms and their inputs are modelled the same in both model applications.

Finally, we describe two scenarios that are used to illustrate the two modelling applications. One scenario is used for both model applications to compare differences. A second scenario introduces a new land use type and is only applied in the model with the deductive approach.

5.3.2 Approaches to determine the relation between land use and its explanatory factors

Inductive approach

In the inductive approach the suitability of a location for a land use type is determined in an empirical way by using logistic regression analysis. This regression model describes the relation between the occurrence of a land use type and the set of explanatory variables (Table 5.1) that are considered to influence land use allocation. The current land use is assumed to reflect the influence that these explanatory variables have exerted on the land use.

The dependent variables in the analysis are binary maps where the land use type under study has a value 1 and all other land use types have value 0. The variables that were inserted in the regression models were selected with a forward stepwise regression procedure (with probability levels of 0.01 for entry in the model and 0.02 for removal from the model). Originally, the data stems from fewer observations than a representation as a grid would suggest and all cells would be considered to be observations. Therefore, a five percent sample was drawn from the original dataset of 99,863 cells to reduce spatial autocorrelation in the analysis. Sampling from a grid is a commonly used method in analysing land use patterns and will minimise spatial autocorrelation to a level that it will not affect the results (Serneels and Lambin, 2001; Stolle *et al*., 2003). Based on the logistic regression analysis the probability of finding the land use type at each location can be determined. These probabilities are assumed to indicate the relative suitability of that location.

Deductive approach

Action-in-Context (AiC) (De Groot, 1992) is a methodology for problem-oriented research that puts activities of actors, for example land use, into context to gain insight in the causes of the activities. Based on Vayda (1983), the research sequence of the AiC methodology is to start with the actions under study, to identify the decision-making social entities directly behind these actions, and then to study the range of options available to the actors and the motivations attached to these (Verburg *et al*., 2003). One of the elements of the AiC methodology is the 'deeper analysis', which ties the options and motivations of the primary actor to underlying cultural and structural factors. The structural framework of the deeper analysis will be used as an actor-model to study the decision-making process of farmers, who are the primary land managers in the study area.

The structural framework of the deeper analysis is depicted in Figure 5.2 (De Groot, 1992) where the arrows show the direction of the causal relations. The first layer in Figure 5.2

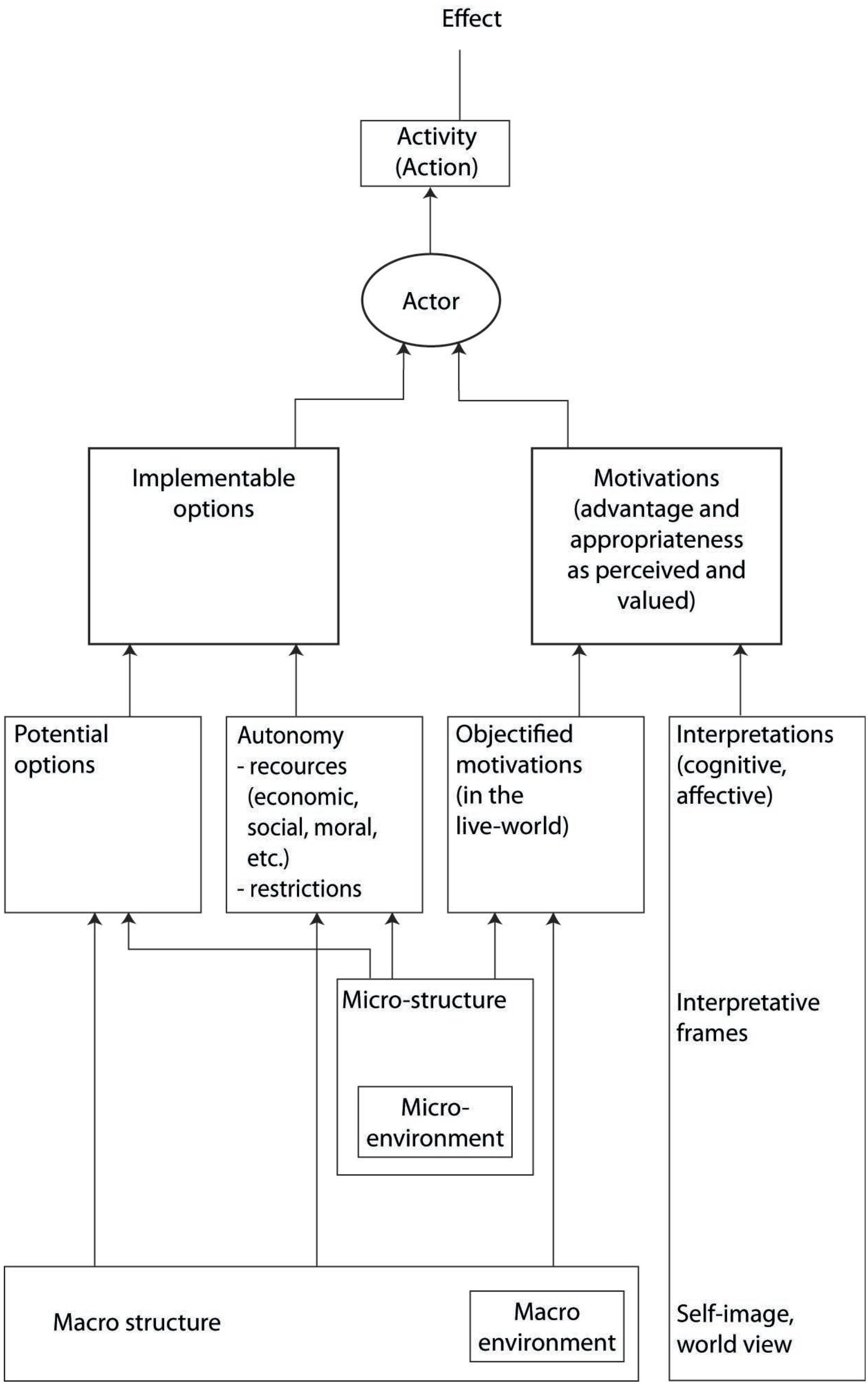


Figure 5.2: Structure of the deeper analysis of the Action-in-Context methodology (De Groot, 1992)

consists of three elements: effect, action and actor. In the case of land use, effects can be soil degradation, biodiversity loss and greenhouse gas emissions, for example. Though, in this study the land use actions rather than the effects are the subject of study. The actors are social entities that exercise a significant decision-making capacity on the activity. In this study the actors are farmers. An example of the relations in the first layer is a farmer (actor) who grows corn (activity), leading to soil degradation (effect).

The second layer consists of ‘implementable options’ and ‘motivations as interpreted’. Implementable options are built up from ‘potential options’ and ‘autonomy’ (layer 3). Potential options are all options the actor is aware of. Though, not all of these options can be implemented. The difference between the implementable options in the second layer and the potential options in the third layer is the difference between what the actor really can do as opposed to what the actor might do if he had the possibility. This difference is determined by the so-called autonomy of the actor. The autonomy consists of resources and restrictions, which together determine which options an actor can implement. Resources contribute to the actor’s capacity to implement actions they consider. The nature of these resources can be economic, social, cognitive, environmental, moral, psychological and physical. For example, for a farmer to grow corn he needs access to land, money to buy inputs and the knowledge how to cultivate corn. Restrictions are autonomy reducing factors, like prohibitions, prescriptions and standards related to environmental licences, but also include physical restrictions.

Motivations are the aspects of the options under consideration by the actor that are normatively relevant to the actor (*i.e.* that give value to the different options). In layer two the motivational factors are specified in terms of ‘advantage and appropriateness’ as interpreted by the actor. These interpreted motivations are determined by ‘objectified motivations’ and ‘interpretations’ (layer 3). The objectified motivations are easily quantifiable units, such as economic costs and benefits or caloric value of produced foods (Verburg *et al*., 2003). Interpretation is shaped by the cultural and psychological opinions and ways of looking that give weight, coherence, shape and colour to the objectified motivations. Together they form the motivations as interpreted by the actor.

In the fourth layer the factors of the third level are seen as being determined by ‘microstructure’, ‘macrostructure’ and ‘interpretative frames, self image, world views’, which is the cultural aspect of the actor’s context. A more elaborate description of the deeper analysis can be found in De Groot (1992) and Overmars *et al*. (2006) (Chapter 3).

The options and motivations of the deeper analysis are used to construct the relations between land use and the explanatory factors in a theoretical-deductive manner as opposed to the inductive method described above. Normally, the AiC approach is applied to cases in which actors or households are the objects of study (e.g. Overmars *et al*., 2006 (Chapter 3)). In the CLUE-S model locations, regular grid cells of 50 by 50 m, are the unit of analysis. Therefore, the options and motivations of land managers have to be converted into suitability maps. This conversion is not always straightforward (Overmars and Verburg, 2005 (Chapter 2)). The field characteristics from the deeper analysis can be easily represented in maps, because field characteristics are directly linked with locations. The influence of household characteristics on land use, as determined in the deeper analysis, is more difficult to incorporate in the suitability maps, because household data are not available in maps. Instead, the household variables are represented as aggregates at the village level. This aggregation may lead to aggregation problems, but the logistic regression analysis of the deductive approach revealed that the aggregated effects of the household variables (ethnicity) are also present. The interpretation of those variables should be made on village level, because the relations at village level can be different from the relations at household level.

As far as the policies and restrictions are spatial these can also be directly represented in maps. Some spatial policies restrict all land use change in a certain area, like, for example, the protection of a nature reserve which is implemented very strictly. Other land use policies restrict a single land use conversion, like the prohibition of the construction of houses in designated agricultural areas or permanent agriculture in the buffer zone of a nature reserve.

An important difference between the regression approach and the deeper analysis is the way potential options are modelled. In the deeper analysis spatial policies and restrictions (or lack of resources) reduce the autonomy of the land manager and therefore the number of options a farmer can implement. This approach really excludes land use types on certain locations, whereas in the regression analysis including variables can only leads to a reduced probability and not to a real exclusion of a land use type at a location due to the way the regression model is specified.

5.3.3 The CLUE-S modelling framework

In this section the CLUE-S model is described and for those aspects that are the same in both modelling approaches a specification is provided. The difference between the two model applications that will be presented is the way the suitability maps are computed.

The modelling framework that is used is the most recent version of the CLUE-S model (Verburg *et al*., 2002; Verburg and Veldkamp, 2004; Verburg *et al*., 2004c). This modelling framework is used to integrate different mechanisms of the land use system into a spatially explicit land use model that is capable of the dynamic simulation of competition and interactions that occur in land use systems. Incorporation of these mechanisms result in a model output that shows path dependency and non-linear behaviour, which characterises the land use system in real-world situations. Path dependency implies that model results of earlier modelling steps have their influence on later modelling steps. Path dependency is dependent on the specification of the incorporated mechanisms, for example land use history and conversion rules, and initial conditions (Brown *et al*., 2005). A general way to look at models, mentioned by Couclelis (2001), is that they are frameworks for organising knowledge. This description fits very well the CLUE-S modelling framework, which integrates different aspects of the land use change process.

The CLUE-S model consists of an allocation module and a series of inputs. The allocation module is a computer program that iteratively computes land use allocation for a number of modelling steps (a detailed description of the allocation module is provided by Verburg *et al*. (2002)). The allocation module can incorporate various mechanisms that are considered to determine the distribution of land use changes in a landscape. These mechanisms are parameterised by the inputs of the model. The quantity of land use change is also imposed to the model as an input. These land use requirements impose the quantity of land use change per modelling step for every land use type in the whole study area. This future 'land claim' can be a fictitious land use scenario based on story lines, as will be used in this study, or an external modelling procedure like macro-economic modelling. Then, the allocation module allocates the aggregated land claim year by year to the cells based on the various mechanisms in an iterative way. So, the strength of the CLUE-S model is to

allocate land use changes rather that modelling the quantity of change.
The mechanisms responsible for the land use allocations can be divided in location characteristics and conversion characteristics. The first locational characteristic is the 'suitability', which is based on the relation between land use and a broad set of biophysical and socio-economic factors. Suitability has an important influence on the allocation of land use change in the model. The basic assumption behind this mechanism is that a location changes into a certain land use in those locations where the 'suitability' is high for that land use type. Suitabilities are represented as a map with values between 0 (low suitability) and 1 (high suitability). This is where the deductive or the inductive approach to derive the land use suitability maps is inserted.
The second location characteristic allows for the incorporation of spatial policies. The suitability map can be altered at locations that a policy applies to. The suitability can be set to zero at locations where a land use type is not allowed to change, for example in a conservation area, or the suitability value can be adjusted by a certain value in areas that are under a policy that, for example, awards subsidies for a certain land use in that area. This mechanism is not used in the applications in this study.
The third location characteristic is the neighbourhood effect (Verburg *et al*., 2004c). Although several theories are available addressing the interaction between neighbouring land use types, for example trends in explanatory variables or spatial processes like imitation, this interaction was not studied extensively in this research. Neighbourhood effects can be included between land use types as well as within a land use type. Because the cell size of the application is smaller than the average parcel size, a small neighbourhood effect was implemented in the model for all land use types with themselves, simulating the clustering of land use into fields and parcels. The value of the neighbourhood effect was based on the eight closest neighbours of each cell. In the calculation of the overall suitability to be used in the model the neighbourhood function determines 20 percent and the suitability maps of the inductive and deductive approach determine the other 80 percent.
The conversion mechanisms that can be incorporated in the model are the so-called conversion elasticities and land use type specific transition sequences. Conversion elasticities can be explained as the resistance of a land use type to change location. For example, tree plantation cannot easily move to another location because the investments made to establish the tree plantation are lost when the plantation moved to another location to make room for another land use type. The conversion settings can be used to create stability in the model by assigning a large influence to the land use history (Verburg *et al*., 2002). The conversion elasticities are implemented in the model as an additional suitability for those locations that are currently under that specific land use. The user should decide on this factor based on expert knowledge or observed behaviour in the recent past or use the factor to calibrate the model. The conversion elasticities that are incorporated in this study are estimated by the authors based on field knowledge and can be motivated as follows. Grassland is easily converted and was given a low conversion elasticity. Corn has a somewhat higher elasticity value since it is relatively easy to establish a corn field. The only requirement for corn is a cleared field. For banana higher investments have to be made and it takes time to before the fruits can be harvested, therefore this land use types received a higher elasticity compared to grass and corn. For rice a considerable effort has to be made to construct a rice paddy. Therefore, rice received a high elasticity. Secondary forest was given an intermediate value and forest a high value.
The transition sequence is a set of rules that determine the possible land use conversions.

Not all land use changes are possible and many land use conversions follow a certain sequence. Sometimes these conversions include a temporal constraint. The conversion mechanisms determine to a large extent the temporal dynamics of the simulations, because they include land use history. In the model applications most land use conversions are allowed except for changes into secondary forest and forest. The only pathway allowed for changes into secondary forest is through grass or banana. The idea behind this rule is that a field must be not used for five years, and thus be grassy, to become secondary (regrowing) forest. Banana fields in the area are cultivated quite extensively and often trees are present. It is considered that if a banana plantation is not maintained for three years this banana plantation can become secondary forest. From secondary forest it takes another five years to grow into mature forest, which is the only pathway to mature forest. The time necessary to grow from one land use into another is estimated and might be subject for further research. Incorporating the effect as such does incorporate path dependency in the model, although it might be not the exact number of years. Banana is allowed to remain for a maximum of twenty years. After these twenty years the banana plantation has to change for at least one year. This rule is based on the lifespan of a banana plantation, which is about twenty years in the area. After these years the banana plant is not producing anymore and is replaced with an annual crop for a short period after which bananas can be replanted.
The various mechanisms are combined in the allocation module. The allocation of all land use types in the case study occurs at the same time in an iterative procedure. Altogether this results in the dynamic simulation of land use competition.

5.3.4 Scenarios

Two scenarios were developed to test the models. The first scenario provides a general indication of what might happen in the research area. The principles are based on the comprehensive development plan of the municipality of San Mariano (Municipality of San Mariano and Housing and Land Use Regulatory Board, 2000). The plans and prospective indicated in this document are translated into general linear trends. The quantities of change are assessments rather than detailed calculations, but will suffice for the objectives of this chapter. The second scenario introduces a new land use type.
As indicated in the planning document of the municipality, the total amount of agricultural lands will not increase substantially for two reasons. First, slope prevents agriculture to expand much further, because productivity of the steeper areas is low. Secondly, there is a necessity for more environmentally-friendly activities in these areas to prevent soil erosion and flooding and to protect natural values of the area. Forested areas will be protected and grasslands with potential for forest production will be rehabilitated and protected. To improve food security and self-sufficiency the production of rice will be promoted. Furthermore, the municipality aims at an increase in productivity per hectare to meet the necessary production and idle lands (mainly unused grasslands) should be taken into production (Municipality of San Mariano and Housing and Land Use Regulatory Board, 2000). The improvement of accessibility, which is also an important municipal goal, is not taken into account in this study.
The scenario sketch above is translated to a quantitative yearly land claim (Figure 5.3 left). The land claim is the total area per land use type for every modelling step and serves as an input to the CLUE-S model, which is specifically developed to allocate this land claim. The translation into real figures is fabricated by the authors based on the ideas of the munici-

pality assuming that the changes are predominantly linear. The scenarios are a projection of changes that might happen rather than strong predictions with a prospective value.
In Figure 5.3 (left) the rice area expands by five percent of the 2001 rice area per year to better meet self-sufficiency in rice, which is the main staple crop of the population. The corn area increases by one percent of the 2001 area per year (allowing some agricultural intensification) and banana area reduces by one percent compared to the 2001 area per year. The latter is to represent improved production of the banana area as well as a reduced interest in this crop due to diseases. Forest area remains the same to visualise the intended conservation effort. Secondary forest will grow by two percent. All increases of land use types are at the expense of the grass area. This scenario is used in the model application using the inductive approach as well as in the application with the deductive approach.
In the second scenario a new development was included (Figure 5.3 right). In a certain policy area (SIFMA) an NGO stimulated the cultivation of (fruit) trees by providing seedlings and assisting farmers to acquire a 25-year tenureship for the parcels involved. The possibility to change to fruit trees was restricted to the SIFMA area, which was identified as an area that needed reforestation. Using the deeper analysis framework an analysis was made for a this new land use type, which was not yet present in the original land use map. So, this scenario could only be modelled using the deductive (AiC) approach because in the inductive approach it is not possible to make inferences for land use types that are not yet present. The information of this analysis was also converted into a suitability map. With this information fruit tree plantations were introduced in the scenario for the deductive modelling approach. The scenario starts with a newly established area of 150 ha with a 5 ha increase in the following years. The extra area for fruit trees was introduced at the expense of the grass area.

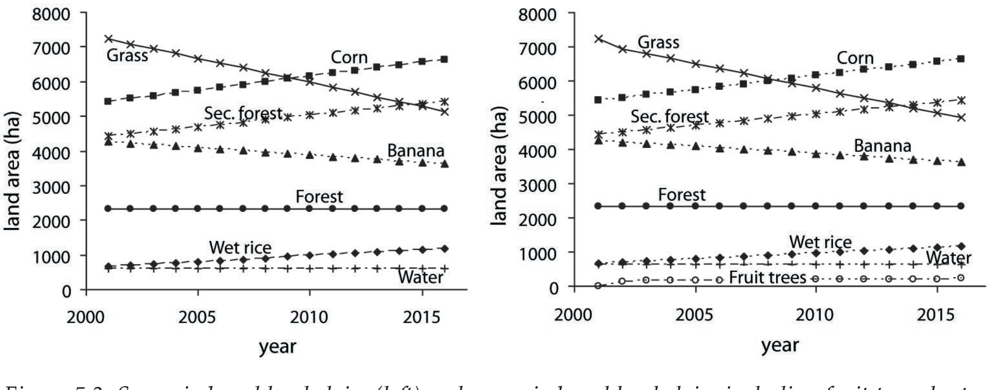


Figure 5.3: Scenario-based land claim (left) and scenario-based land claim including fruit tree plantations (right)

5.4 Results

5.4.1 Results of the inductive CLUE-S application

The results of the regression analysis that was used to determine the relation between land use and its explanatory factors are presented in Table 5.2. For rice and corn all variables were entered in the stepwise regression procedure. For the land use types banana, secondary forest and forest the ethnicity variables were excluded. Ethnicity was assumed to have no relation with these land use types. Table 5.2 shows which variables were actually included by the stepwise procedure together with their significance levels.

Table 5.2: Regression coefficients of the resulting logistic regression models

Variables	Rice	Corn	Banana	Sec. forest	Forest
Constant	-3.856***	-0.599**	-0.838***	-2.607***	-8.142***
Slope1	0.515*	1.137***	-0.655***	-0.756***	
Slope2		0.750***			
Slope3		0.399**		0.230**	0.725***
Slope4	-1.077*				1.396***
Slope5					1.551***
Creek	1.010***	-0.217*		0.313***	
Plot distance	-0.009***	-0.016***	-0.005***	0.002***	0.021***
Transportation cost		-0.026***	-0.012**	0.031***	0.084***
Ethnicity Ilocano					
Ethnicity Ifugao	0.020**				
Ethnicity Ibanag		0.011***			
ROC	0.73	0.77	0.65	0.68	0.96

* $p<0.05$, ** $p<0.01$, *** $p<0.001$

The ROC (Relative Operating Characteristic) (Swets, 1988) was used to indicate the goodness-of-fit of the regression models. The ROC summarises the performance of a logistic regression model over a range of cut-off values. The value of the ROC is defined as the area under the curve linking the relation between the proportion of true positives versus the proportion of false positives for an infinite number of cut-off values. The ROC statistic varies between 0.5 (completely random) and 1 (perfect discrimination) (see for more details Pontius and Schneider, 2001). Rice and corn have a good model fit, banana and secondary forest had a relatively poor fit and forest had a very good fit.

For grass and water no regression analyses were made. Grass is a leftover land use type that is not used and is not cultivated intentionally. It just grows on locations that are not occupied with one of the other land use types and is therefore treated as a ‘rest’ category without any specific suitability. Water is considered to be constant over the modelling period and is therefore excluded from the modelling exercise. Examples of suitability maps for banana and secondary forest, which are constructed with the relation found in the regression analysis, are shown in Figure 5.4A and 5.4C.

The resulting land use map of the inductive modelling approach after a 15 years modelling period is shown in Figure 5.5B. Under the scenario and modelling assumptions applied to this model the following major trends can be identified. First, the banana area decreases in the area marked with 1 and is relocated in area 2. The abandoned areas in area 1 are occupied with grass. In general, existing corn areas (like in the area near 3) expand throughout the area near places where corn was already present. Rice expands in the area near 4 and just below area 2. Secondary forest increases mostly near 5. Forest is stable in this scenario.

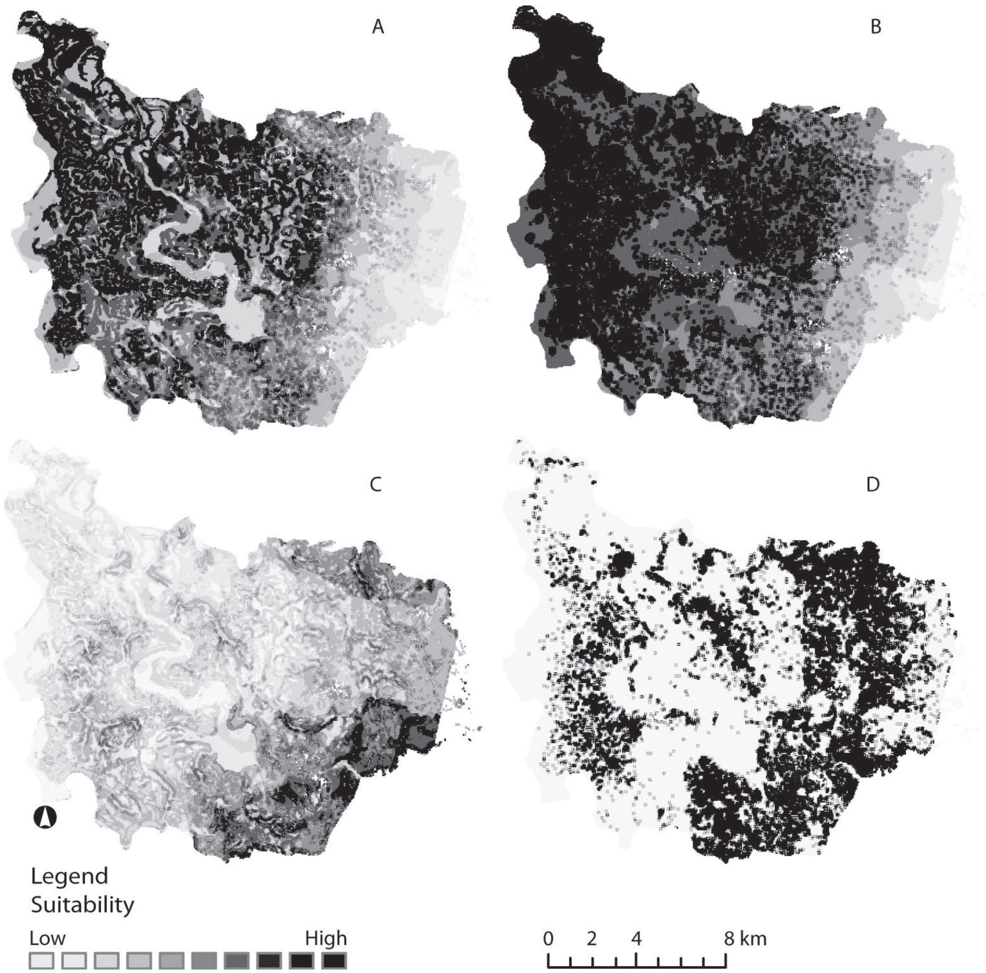


Figure 5.4: Suitability map for banana for the inductive case (A) and the deductive case (B) and suitability map for secondary forest for the inductive case (C) and the deductive case (D). The neighbourhood effect is added to these suitability maps (note that the neighbourhood effect is recalculated every time step and changes with changes in the land use). The scaling of the legends was stretched between the highest and lowest value.

5.4.2 Results of the deductive CLUE-S application

The first element of the actor decision model that is described are the motivations to cultivate a land use type, which are calculated according to Overmars *et al*. (2006) (Chapter 3). The relations between the land use and the explanatory factors are formulated in Equations 5.1 to 5.7. For every land use type this calculation is different. The parameters for this model were calculated using with field observations where possible and otherwise they were based expert knowledge and interviews with farmers. Not all parameters are provided in this chapter, but can be accessed from (Overmars *et al*., 2006 (Chapter 3)).

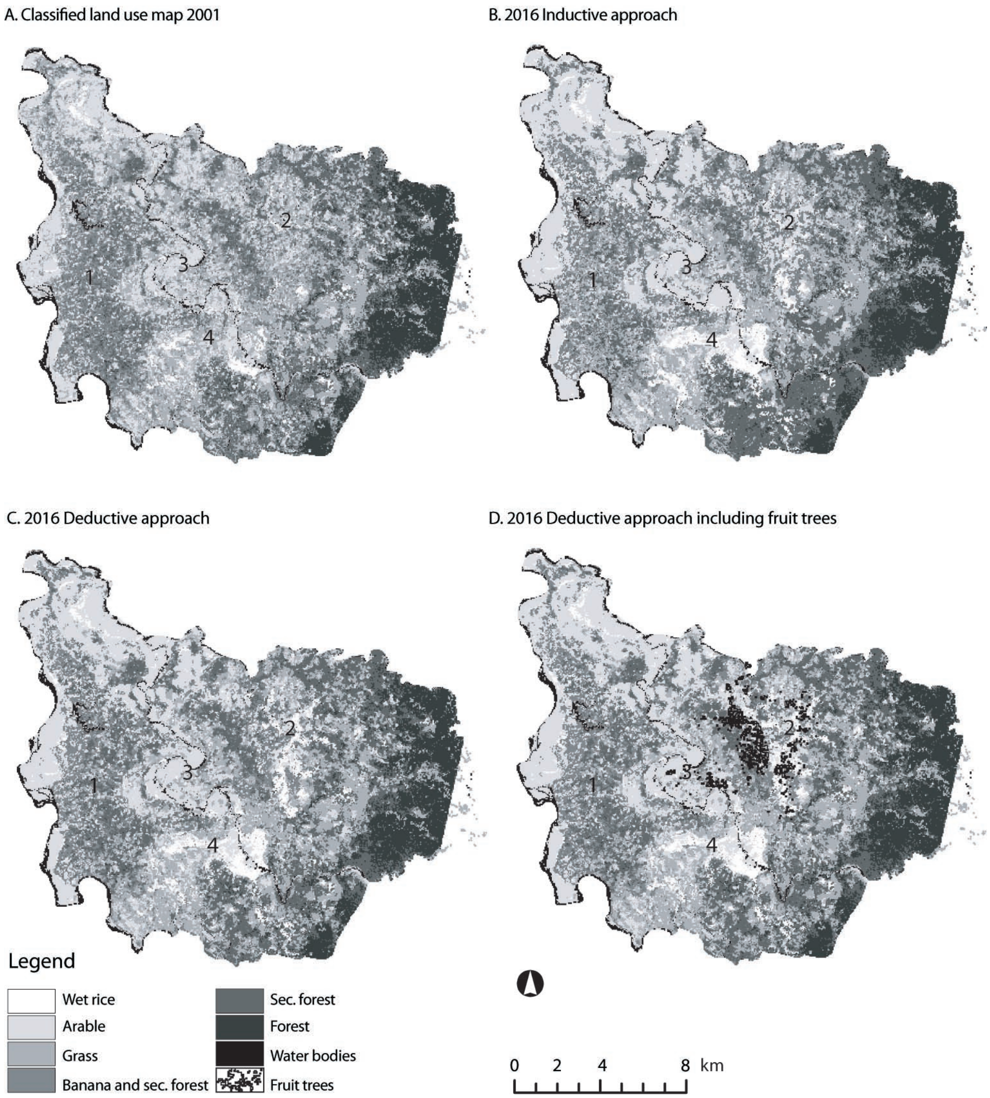


Figure 5.5: Classified land use map 2001 (A), simulated land use map in the year 2016 of the inductive modelling approach (B), the deductive modelling approach (C), and the deductive approach including fruit trees (D)

The maximum benefit is different per land use type. The slope factor applies only to corn and reflects the losses in yield due to slope. The risk is calculated from the average yearly loss due to typhoons, droughts, pests and diseases as reported by farmers. The costs to transport the harvested product are a combination of cost made by travelling from the field to the residence of the farmer and transportation costs to the market in San Mariano (the latter is not included for wet rice). Preferences for a specific land use type depend on the ethnicity of the household. Since the household information is not available this preference is calculated for the village as a whole. The ethnicity specific preferences are multiplied by the percentages of the population that belong to an ethnicity and summed for all ethnicities. The preference for corn is considered to be higher than average for ethnicity Ibanag and low for ethnicity Ifugao. For rice the preference is high for Ilocano and Ifugao and low for Ibanag (compared to the group of other ethnicities).

*Motivations = objectified motivations * preferences* (5.1)
Objectified motivations (net benefit) = (max_benefit - tr_costs) slope_fact * (1-risk)* (5.2)
Max_benefit = f(CROP) (5.3)
Slope_fact = f(SLOPE, CROP) (5.4)
Risk = f(CROP) (5.5)
Tr_costs = f(TR_COST, PLOT_DISTANCE, CROP) (5.6)
Preferences = f(ETHNICITY, CROP) (5.7)

The implementable options are also based on Overmars *et al*. (2006) (Chapter 3). The implementable options for rice were determined by slope, which should be flat or flat to moderate, and the possibilities for irrigation, close to a creek or an irrigation facility. For corn and banana no restrictions were formulated in this analysis.
The options and motivations together form the suitability maps. The deeper analysis was used to calculate suitability for rice, corn and banana. The factors that are incorporated per land use type are indicated in Table 5.3 as well as the performance of the model, which is indicated with the ROC. These ROC values were added to compare the results with the inductive approach, although the aim of the deductive approach is not to get the best fit possible, but a good representation of the processes. This subject will be further explained in the discussion for the case of banana. The remaining land use types were modelled as having an equal suitability for all locations. For grass we made the same assumption as in the inductive approach. In contrast with the case for grass, the other two land use types without suitability analysis actually do have a use. Forest is used for (illegal) logging and for this purpose accessibility plays an important role. Secondary forest can be also used as timber or firewood and therefore accessibility may also play a role. Part of these processes is covered by the neighbourhood functions that were incorporated. The argument to not include suitabilities for forest and secondary forest in these model applications is that under the scenario forest is stable and secondary forest is increasing. The suitability for increasing (regrowing) forest and secondary forest is not related to factors that determine suitability for logging and can actually be constant for all locations.
The resulting land use map of the model application with the deductive approach after fifteen years modelling period is shown in Figure 5.5C. In this application the following major trends can be identified. First, the banana area decreases in the area marked with 1 and is relocated in area 2. The abandoned areas are occupied with grass and secondary forest. Like in the inductive approach existing corn areas expand. Rice expands in the area near 4 as well as below area 2. Secondary forest increases evenly in the study area and the forest area is stable.

In the second scenario applied to the deductive modelling approach fruit tree plantations are introduced. The suitability for this land use type (Figure 5.6) is similar to that of banana, although the general profitability of the fruit tree plantations is higher than the cultivation of banana. The most important difference with the suitability of banana is that the fruit trees are restricted to an area where the SIFMA policy applies to because in the other areas

Table 5.3: Factors included in the deductive approach indicating if the factors are incorporated in the options or the motivations of the land managers

Variables	**Rice**	**Corn**	**Banana**	**Sec. forest**	**Forest**	**Fruit trees**
Slope1	options	motivations	-	-	-	-
Slope2	options	motivations	-	-	-	-
Slope3	-	motivations	-	-	-	-
Slope4	-	motivations	-	-	-	-
Slope5	-	motivations	-	-	-	-
Creek	options	-	-	-	-	-
Plot distance	motivations	motivations	motivations	-	-	motivations
Transportation cost	-	motivations	motivations	-	-	motivations
Ethnicity Ilocano	motivations	-	-	-	-	-
Ethnicity Ifugao	motivations	motivations	-	-	-	-
Ethnicity Ibanag	motivations	motivations	-	-	-	-
SIFMA	-	-	-	-	-	options
ROC	0.66	0.74	0.54			

initial investments are considered to be too high.
The land use changes are similar to those in the deductive approach (without introduction of fruit trees). The additional area for fruit trees caused other land use types (mainly banana) to move to other areas.

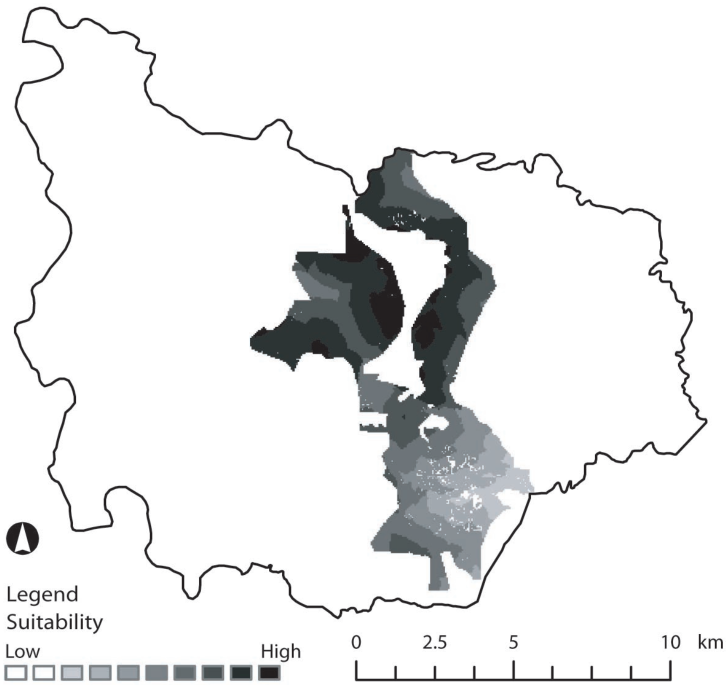


Figure 5.6: Final suitability maps for fruit trees including the restrictions

5.4.3 Comparison of the two modelling approaches

If the maps of the application using the inductive and the application using the deductive approach (Figure 5.5B and 5.5C, respectively) are compared cell by cell the maps have 85 percent in common. In total the land use changed in 25 percent of the locations after 15 years. From these changes 54 percent were exactly the same changes in both modelling approaches. If the comparison is made within larger windows (Costanza, 1989), allowing for differences in location within the window, the similarity increases (Figure 5.7).
A general observation from these scenario studies is that if the grassy areas are used both the agricultural as well as the forested areas can be improved. Agriculture is the main source of livelihood in the area and the forest can sustain the ecological function of the area. Grassland on the contrary does not contribute to production and neither has much ecological value.

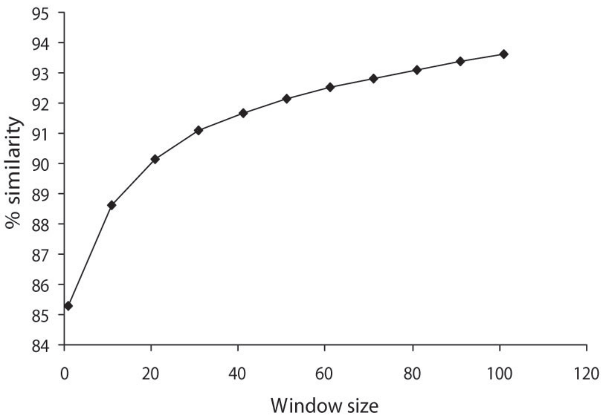


Figure 5.7: Similarity between the inductive and deductive modelling approaches with varying window size based on their 2016 maps

5.5 Discussion and conclusions

At first instance the final results of the inductive and deductive modelling approaches look quite similar (Figure 5.5B and 5.5C). If only the changes are considered 54 percent of the changes are at different locations. These differences between the maps of the two applications are caused by differences in the suitability maps, which were constructed in different ways. After all, the other inputs and model setting (*i.e.* the neighbourhood effects and the conversion mechanisms) are completely identical, so these cannot cause the differences. The deductive and inductive approach to create suitability maps for a land use type can vary because of different variables, different relations between the variables and different parameters. Some suitability maps show only local differences. These differences in suitability maps translate into differences in land use allocation at small distances, which disappear when the similarity is compared with bigger windows. A good example of local differences is the suitability of banana. In the inductive case (Figure 5.4A) slope was included while in the deductive approach (Figure 5.4B) slope was not included. As can be

seen the general pattern, caused by plot distance and transportation cost (and having the same sign), is similar, but the slope introduces differences at small distances. The suitability maps of secondary forest show more differences. The inductive approach (Figure 5.4C) detected correlation with a series of factors while in the deductive approach (Figure 5.4D) no factors were included and therefore the suitability map shows only the neighbourhood effect. These differences in suitability are reflected in the differences in the resulting land use maps even if larger windows are used in the evaluation procedure.
So, the inductive and deductive approach to specify the suitability map do not always lead to the same results, because the inductive approach is based on correlation while the deductive approach is based on processes that were observed by the authors and/or described by the respondents. In other studies suitability from both approaches may be more different leading to even more differences in the suitability maps of the two approaches.

Even more important is that the research paradigms of the two approaches are different, which has its implications for the interpretation and the use of the modelling approaches. The structural difference between the two approaches used in this study is that the deductive approach attempts to describe causality while the inductive approach to land use analysis reveals associations rather than causal relations (Serneels and Lambin, 2001; McConnell, 2002; Verburg *et al*., 2003; Verburg *et al*., 2004d).
In the inductive approach the current land use pattern is assumed to reflect the processes of land use in the past. The result of these processes, the land use pattern, is described with the regression model using correlations between the land use and its explanatory variables. The processes themselves are not described and, therefore, changes in the processes and their effect on the suitability of a location for a land use type cannot be incorporated in the modelling of future scenarios. So, using the inductive approach, the assumption that has to be made for the modelling exercise is that the processes that determine land use do not change. This approach is described by a study by Kok and Winograd (2002) where modelling of scenarios with and without the impacts of Hurricane Mitch (Honduras) results in the same land use map after ten years under the assumption that the relations between land use and its drivers was re-established after a few years. It may be true that relation do not change at short time scales, but at larger times scales different factors may become important and sudden events, like a change in political system may cause dramatic changes in behaviour. In the models presented, however, the behavioural rules are assumed to be constant. Besides this, no new land use types can be introduced, since the relation of this new land use type with the explanatory factors cannot be determined statistically. Even if the regression analysis was able to describe processes the assumption has to be made that the land use system is in equilibrium with the explanatory factors. Analysing a system that is not in equilibrium may lead to possible error in the description of the process.
The deductive AiC approach, on the other hand, describes the processes explicitly. Therefore, changes in the processes that determine land use can be incorporated in the construction of the suitability maps, which enables the introduction of discontinuities and new land use types in the scenarios that are modelled. All these issues have their consequences on the type of scenarios that can be simulated with the modelling approach. A case with a discontinuity was demonstrated to some extent by Kok and Veldkamp (2000), who used a rule-based suitability map for a new land use type to enable the incorporation of this land use type, like was done in this chapter with a sound theoretical framework. The other suitability maps in the study by Kok and Veldkamp stem from regression analysis.

A more technical difference between the two approaches is that with an inductive approach the regression analysis determines the relation between the current pattern of the land use types and its explanatory factors, whereas the AiC approach determines the potential suitability of the land use types. In the regression analysis the occurrence of a land use type, which serves as the dependent variable in the regression analysis, is not independent of the other land use types. This dependency has its consequences for the applicability of the modelling approach. To illustrate this consider the example of the bananas in this study. In general, bananas are located on 'second best' locations, because the best locations are cultivated with corn, which is (potentially) more profitable. For example, in this study bananas are correlated with high slopes, which is due to the fact that on the flatter parts corn is preferred, not because bananas grow better on steep slopes. In the inductive approach the calculated suitability for banana is high at these second best locations. If a large change in banana area would occur, for example when suddenly all corn would disappear and the banana area would expand fast (large changes may happen for example through large price changes or diseases), these new banana areas would first be allocated on areas with a high banana probability, which in fact are the second best locations. In reality the new banana would first appear where the suitability for banana is optimal. The deductive approach would allocate these where the potential suitability for banana is high. Generally speaking, the inductive approach to specify the suitability map in CLUE-S is applicable in situations with relatively small land use changes, without introduction of new land use types, whereas the deductive approach to specifying the suitability map in CLUE-S approach is more flexible in this respect.
The advantage of the empirical approach is that the procedure of the regression analysis is straightforward and easy to reproduce. Limitations of the empirical approach are that many regression models have a restricted specification of the relation between variables (e.g. linear, log-linear). Though, increasingly, statistical tools are introduced that can capture the structure, and therefore also part of the processes, of the land use system. For example, multilevel models (Pan and Bilsborrow, 2005) can incorporate a hierarchical structure and autoregressive models (Overmars *et al*., 2003) can capture spatial processes.
The AiC analysis used in the deductive approach depends on the skills and interests of the researcher. Therefore, the AiC analysis is less reproducible than the inductive analysis. The land use system does not have to be in equilibrium because the processes are observed directly rather than derived from the current land use pattern. Finally, the deductive (AiC) approach does not constrain the specification of the mathematical relations between factors in any way, giving more flexibility to the modeller.
In this respect it is regrettable that the household information and the ownership relation with the land were not spatially available for the study area. The distribution of the parcels and their ownership is an important determinant of the observed land use pattern. By not incorporating this structure the model has the tendency to allocate the land use according to the smooth patterns of the suitability maps, while the observed land use pattern shows a more irregular pattern due to land ownership. The AiC analysis of Overmars *et al*. (2006) (Chapter 3), from where the deductive approach is derived, is based on a household survey and could have been easily incorporated if this information was available.

The differences described above have their implications for the applicability of the models to answer questions in research and policy-making. To have some foothold to assess the use of the two modelling approaches for research and/or policy Couclelis (2001) provides

some qualifications for both: Besides that both must be built on good science, use good data, and should answer good questions, research models should have a higher degree of scientific rigor and should contribute original theoretical insights or technical innovation. Policy models should preferably be used, verified and validated often and should be transparent and manipulable and should include key policy variables.
As far as the qualifications for research are concerned both approaches are quite similar. The main difference is that the approaches stem from two completely different research paradigms. With respect to the policy issues the two approaches do show important differences. First of all, the inductive approach is more transparent and the CLUE-S model using this approach is validated for several cases (Kok *et al*., 2001; Pontius *et al*., 2005). The AiC approach is dependent on the judgement of the researcher and is therefore less transparent and reproducible. Secondly, the deductive (AiC) approach is more flexible (manipulable) then the inductive (regression) approach, which has to stick to more rigid model definitions. Concerning the inclusion of key policy variables the deductive approach has the advantage of the explicit description of parameters and relations between variables. Another advantage that adds to that is that it can include variables like market prices and investments. These variables are constant throughout the study area and can therefore not be included in regression analysis. They are included in the AiC analysis and can therefore be used to study the effect of changes in price (through for example subsidy policy) or changes in technology, which can be important policy variables. So, the deductive method has more options to analyse the effects of policies, which are often implemented at the macro-level. Potentially, this approach would also enable the modelling of the amount of land use changes and therefore the possibility to make the model more dynamic.
The scale to which both approaches can be applied is different. In principle the inductive approach can be applied to any scale (*i.e.* resolution and extent). However, the amount of detail and knowledge about the decision-making structure of actors involved that can be incorporated is limited. The deductive approach as presented in this study relies on detailed information about the land managers. To incorporate this information in a spatially explicit model the resolution should be comparable with the size of the decision units of the actors. Aggregating these units to larger grid cells would lead to aggregation problems. So, the deductive approach should preferably be applied to the watershed level using a fine resolution.
Currently, many efforts in land use modelling have adopted the multi-agent modelling approach (e.g. Parker *et al*., 2003), which is an agent-based approach in which actors communicate and interact. The deductive approach in this chapter is not a multi-agent model. However, the model can be regarded as an agent-based. It specifies the decisions of farmers in various circumstances, but without communication and interaction and without other actors than farmers involved. By using an actor-decision model to specify land suitability, decisions of the land manager were given a more prominent role in the modelling approach than with a statistical approach. The deductive approach provides more process-information than the inductive approach although the representation of the actors involved is simplified to one representative actor.

Both the deductive and the inductive approaches have their own origin and research paradigms and their own advantages and disadvantages as pointed out in this final section. Within the scope of this study, no qualification of the models was presented that was based on validation of the simulated land use maps. This would not have provided many new insights because the resulting maps were quite similar in this case study. A more important conclusion is that the research question and the nature of the case that is studied determine which approach is most suitable to use. The deductive approach can better handle discontinuities in land use processes and can therefore evaluate a wide range of scenarios, which can also include new land use types. The inductive approach is easily reproducible and is well able rapidly to identify hotspots of land use change. The deductive approach is better suited for smaller study areas, but needs fieldwork to implement. The inductive approach can be applied more quickly in larger areas if basic data are available.

6

Projecting land use change and its effects on endemic birds in the northern Sierra Madre, Philippines

Abstract

Fragmentation of habitat due to human induced land use changes poses an important threat to biodiversity. For conservation management it is important to understand the relation between the landscape and biodiversity and to develop indicators that provide quantitative measures of changes in biodiversity. To take appropriate conservation measures it is also important to make projections of future land use patterns and determine how these patterns can be influenced by policies. This chapter combines landscape-biodiversity relations with land use projections from state of the art land use modelling. The research is carried out in the Philippines, which is a global hotspot of biodiversity. This chapter evaluates the relation between endemic bird species richness and the landscape by taking into account both variables at the location itself as well as landscape characteristics. All variables are derived from a land use map. This facilitates the projection of species richness based on the projected land use maps using the same relations. Scenarios with high and low agricultural expansion were evaluated with two main variants for the level of forest protection. Results show that even with the same total area per land use type, the spatial pattern of the land use types can cause differences in the value of the landscape for bird conservation. Spatial policies like park protection can be used to influence the spatial pattern of land use and therefore the biodiversity. Combining the ecological studies with a land use model has additional value for nature conservation, because the land use models can incorporate the human factor, incorporate the dynamics between different land use types and can create projected maps of future landscapes. In this way one can better assess the implications of different (land use) policy variants for nature conservation.

Based on: Overmars, K.P., Van Weerd, M., Prins, M., Thijs, W. Projecting land use change and its effects on endemic birds in the northern Sierra Madre, Philippines. Journal of Applied Ecology (Submitted).

6.1 Introduction

Habitat loss due to land use changes such as deforestation, forest fragmentation and agricultural expansion has been identified as an important threat to (avian) biodiversity (e.g. Tilman *et al*. 1994; Turner 1996; Myers *et al*., 2000; Brooks *et al*. 2002; Sodhi *et al*. 2004; Henle *et al*. 2004). To assess the impact of land use changes on biodiversity it is important to understand the relation between landscape and biodiversity and quantify associations between species and habitats (e.g. Buckland and Elston, 1993; Bailey *et al*., 2002; Dauber *et al*., 2003). These relations are often described with statistical models (e.g. Buckland and Elston, 1993; Guisan and Zimmerman, 2000; Fazey *et al*.; 2005). Habitat – species relations derived with statistical models can be used to take informed decisions on conservation management (e.g. Bunnel and Huggard, 1999; MacNally *et al*., 2003; Gibson *et al*., 2004). However, the actual effect of (conservation) policies on the landscape, and therefore on biodiversity, is difficult to assess based on these statistical relations alone for two reasons. Firstly, changes in land use types are often interdependent (Verburg *et al*., 2002). A spatial measure at one location will have its influence on land use at other locations. For example, habitat protection at one location can divert pressure for agriculture to elsewhere, where it has also an influence on biodiversity. Secondly, landscape metrics are often related in a non-linear way (e.g. Hargis *et al*., 1998), for example the relation between an increase in habitat and total edge. Therefore, the effect of, for example, an increase in habitat on biodiversity cannot be predicted accurately without expressing the land use changes in a spatial explicit way. Spatially explicit land use models can offer the means to create projections of land use changes in maps based on policies and land use developments.

Currently, land use models enable land use scientists to make projections of future landscapes under different scenario conditions (e.g. Briassoulis, 2000; Veldkamp and Lambin, 2001) by simulating competition and interactions between land use types in a spatially explicit and temporally dynamic way. If the relation between biodiversity and the environment is known and future landscape maps are available through land use modelling the effect of conservation policies and land use developments on future biodiversity can be assessed more quantitatively, which can be used as a tool for conservation planning and policy decision-making.

The combination of high numbers of endemic species and severe loss of natural habitat causes the Philippines to be one of the most important conservation hotspots for biodiversity in the world (Myers *et al*., 2000). Past and current land use changes, especially deforestation in the last century (Kummer, 1992; ESSC, 1999), are the major threat to this unique biodiversity (Brooks *et al*., 2002). The Philippines has 408 resident bird species of which 187 are endemic (Kennedy *et al*., 2000; WBCP, 2004; IUCN, 2005). Of these 187 endemic bird species 175 (94 %) are forest-dependent; Fifty-three percent of these Philippine endemic forest bird species are threatened or near-threatened (IUCN, 2005), mainly as a result of deforestation (Brooks *et al*., 1997; Brooks *et al*., 2002).

When studying the impact of deforestation on birds, total species richness, where all species have equal value, might not be a good indicator of recovery or decline of forest avian biodiversity as forest species are usually replaced by edge or open area species (Johns, 1996; Ghazoul and Hellier 2000). In the case where forest is the natural habitat the number of forest dependent species, or in this chapter endemic forest bird species, provide a much more meaningful indicator that can highlight changes in the conservation value of the area. (In the remainder of the chapter endemic forest bird species are referred to as endemic bird species).

The objective of this study is to evaluate the impact of land use changes and the influence of land use policies under different scenarios on species richness of endemic forest birds in a transition zone from agriculture to forest. Firstly, the relation between landscape properties and the occurrence of endemic bird species is determined for the current situation. This information will then be used to create a map with an index for endemic species richness. Secondly, spatially explicit land use projections will be modelled for three scenarios. These scenarios vary in the implementation of conservation policies and the level of economic and population growth. Thirdly, the changed landscape characteristics are derived from these land use projections and species maps will be created for the scenarios using the relations derived from the present landscape. These projected endemic species maps will be compared with the current situation and amongst each other to determine the possible impact of the different scenario settings on endemic bird occurrence. We will discuss the value of the combination of methods presented for nature conservation management.

6.2 Material and methods

6.2.1 Study area

The study area is located in the municipality of San Mariano, Isabela Province, the Philippines, in the north-eastern part of the island Luzon (Figure 6.1). Part of the study area is situated in the Sierra Madre Mountains, which are covered with one of the largest contiguous areas of forest left in the Philippines and stretches from Quezon province to the northeastern tip of Luzon. The study area covers approximately 48,000 ha. The topography is hilly in the western part and becomes steeper towards the east where the mountain range is situated. At present, the study area has a land use gradient from intensive agriculture, with mainly rice and yellow corn, near San Mariano to a scattered pattern of rice, yellow corn, banana, grasses and trees to residual and primary forest in the eastern part of the study area. Throughout the area remnant forest patches and patches with secondary forest are present. (In this chapter the term secondary forest is used for all low-density forest types such as secondary growth and logged over forest remnants). Altogether this leads to a mosaic landscape with a large variety of land use types.

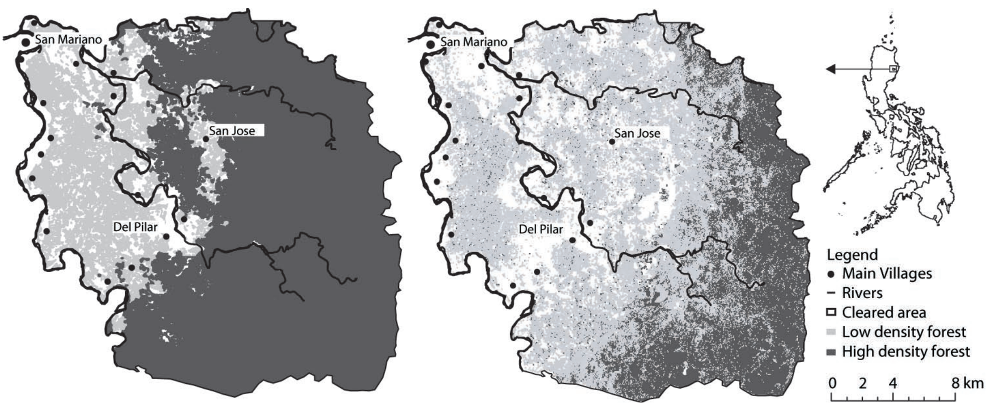


Figure 6.1: Location of the study area in the Philippines and land use maps, with a focus on forested areas, for 1972 and 2001

Before migration started about a century ago the study area was covered with tropical rain forest and few people lived in the area. Currently, most of the area has been cleared and approximately 20,000 people inhabit the area. The main cause of this deforestation in the area is commercial logging, which occurred between 1960 and 1990. Furthermore, small-scale logging and agricultural expansion have caused forest degradation. Figure 6.1 shows the state of the forest in 1972 and 2001. In this period approximately 18,000 ha of natural forest was converted in the study area. About one third of the converted area was completely cleared and two thirds of the original forest is currently residual forest, forest regrowth or extensive banana plantation. Complete clearing of land was mainly caused by farmers who clear land that commercial loggers have logged selectively. Some of the cleared areas are currently grassland. Often, these grasslands are claimed by farmers who keep those lands for their inheritors, but do not have the time to cultivate them at the moment. The grasslands are burned regularly on purpose or accidentally as a result of spreading fires. This burning obstructs regeneration of forest (Snelder, 2001).
Since a logging moratorium was enacted in 1992 commercial logging was abandoned in the area (Persoon and van der Ploeg, 2003). The moratorium made people to switch from logging based activities to agriculture. The municipality has projected population growth to be 2.86 per year based on the average growth over the period 1975-1995. This includes immigration from outside the municipality. Though illegal, selective logging still occurs, which mainly provides wood for local use and the furniture industry in the region. This small-scale logging, though less extensive than the previous commercial logging operations, poses a threat to the quality of remaining forest as a habitat for forest dependent species. Hunting and the continued removal of remnant forest for agriculture pose additional threats. In 1997, 280,000 ha of remaining forest of the Sierra Madre in Isabela Province were declared a protected area: the Northern Sierra Madre Natural Park (NSMNP). This largest protected area of the Philippines was established to protect remaining forest and its associated biodiversity, among which 119 resident lowland forest bird species of which 68 are endemic to the Philippines and 21 are classified as threatened or near-threatened (Van Weerd, 2002). The portion of the NSMNP in San Mariano is, on paper, totally protected but environmental law enforcement is weak or corrupted (Van der Ploeg and Van Weerd, 2004) and illegal logging continues on a large scale (Van Weerd *et al*., 2004).

6.2.2 Methods

Species richness mapping
To assess the landscape value for avian biodiversity in the study area it is important to determine the relation between habitat and landscape characteristics and endemic species richness. This relation is determined with regression analysis and used to create a predictive endemic species richness map for the entire study area on a 50 by 50 m grid (Gibson *et al*., 2004, Luoto *et al*., 2004). To make projections of future endemic species richness the relations from the regression analysis are applied to the modelled land use maps representing the possible future landscape.
The dependent variable in the regression analysis is the number of endemic bird species, which is determined with point counts taken over a period of 15 minutes. This variable is not the absolute number of endemics occurring on a location. Therefore, the predicted values for endemic species richness should be interpreted as an index rather than absolute species richness. The assumption is that the number of endemic species observed in a point

count is proportional to the actual number of endemic species present and their relative abundance. The index is used to quantify the relative differences of endemic bird occurrence between sites and to predict the conservation value of future landscapes for endemic bird species.

The occurrence of species should be studied at multiple levels, because habitat selection of bird species depends on factors at different scales (Atauri and Di Lucio, 2001; Dauber *et al*., 2003; Luoto *et al*., 2004). Given the extent and the aims of this study we have chosen to identify two scales to explain endemic species richness: local variables, which describe the habitat of the observation point, and landscape characteristics, which describe the structure of the surrounding landscape (Atauri and De Lucio, 2001). The structure of the surrounding landscape, e.g. the diversity and disturbance of the landscape and the proportion of different land use types, are important co-determinants of the occurrence of species (Dauber *et al*., 2003). This study does not go into the details of describing the internal structure of the vegetation. All independent variables are derived from the land use map.

The landscape is considered to be a heterogeneous and dynamic continuum of habitats. Therefore, locations (points or cells) are the unit of analysis and all land use types in the landscape mosaic are considered to be potentially suitable for endemic birds to use as their habitat or to travel through. This is in contrast to approaches that evaluate islands of suitable habitat (e.g. forest) in a sea of absolute hostile matrix (Luoto *et al*., 2004), without any specification of this matrix. Therefore, fragment size as such was not incorporated as a variable. Instead we considered the proportion of each land use type within distances of 250 and 1000m from each location to include the availability of a preferred habitat (land use) in the neighbourhood. This approach enables the assessment of the value of the mixed agriculture-nature landscape for certain species and its use for nature conservation.

In many studies, species richness is related to a multitude of variables including both biotic variables that describe habitat and landscape structure and abiotic variables based on topography, climate, soils type, etc. (e.g. Osborne *et al*., 2001). In this approach only the biotic variables are used assuming that endemic bird species richness is determined by the different available habitats (land use types) and the spatial pattern of these. The other variables (slope, climate, accessibility, etc) are considered to be determinants of this land use pattern and are used in the land use model to project the future land use map. So, these variables influence bird species richness indirectly through their influence on the landscape (land use map).

Because the dependent variable consists of counts Poisson regression is used. Count data has only positive values and Poisson regression can deal with this type of data (Crawley, 1993). Analysing count data using standard linear regression is not appropriate because the variance will not be constant and it might predict negative values. Poisson regression accounts for this by assuming a Poisson error structure and using the log link function (Crawley, 1993). Poisson regression is regularly used in ecological applications (e.g. Lobo and Martín-Piera, 2002; Mac Nally *et al*., 2003; Gibson *et al*., 2004). To select variables to incorporate in the model the variables were tested one by one on their influence on the deviance, which is a relative measure for the goodness-of-fit. In the first step the variable causing the highest change in deviation was included in the model after which the procedure was repeated to select a subsequent variable. The results based on the regression analysis with the point count data are used to construct a predictive map of the index. All independent variables are available in maps and using the loglink function predictions for all cells can be calculated.

Land use modelling

To make projections of land use changes the CLUE-S (Conversion of Land Use and its Effects at Small regional extent) modelling framework was used (Verburg *et al*., 2002; Verburg and Veldkamp, 2004; Verburg *et al*., 2004c). This model can dynamically simulate the competition and interactions between land use types. Because of its explicit focus on spatial processes this model is very well suited to produce maps of future land use patterns. The projected land use patterns will form the basis to project future occurrence of endemic birds.

The CLUE-S model consists of an allocation module and a series of inputs (Figure 6.2). The allocation module is a computer program that iteratively computes land use allocation per time step (e.g. a year) for all land use types simultaneously. To allocate land use changes in a landscape the model combines a set of mechanisms that are considered to determine the land use system, which are parameterised by the inputs of the model. The total quantity of each land use type in the study area per modelling step, the so-called 'land claim', is not modelled in the allocation module, but is imposed to the model as an input. In this study the land claim was constructed as a scenario study from fictitious story lines. (For more information see Verburg *et al*., (2002) and Chapter 5).

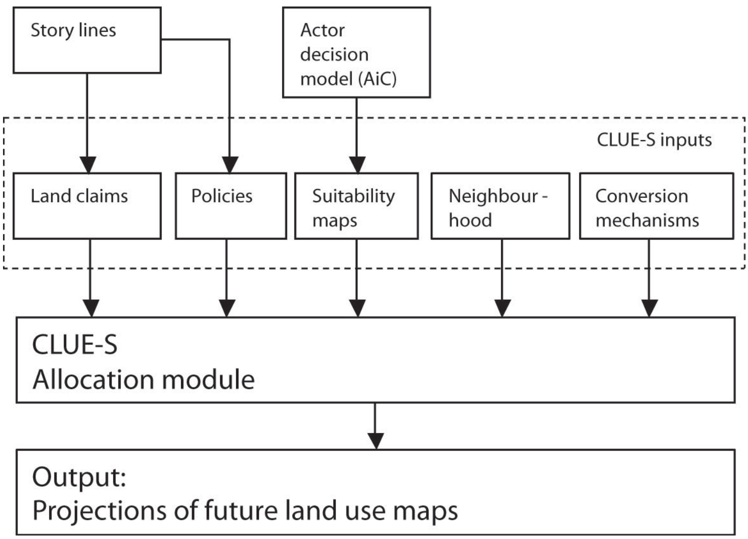


Figure 6.2: Graphical representation of the CLUE-S modelling framework

The mechanisms of land use allocation included in the model can be divided in location mechanisms and conversion mechanisms. The first location mechanism is a 'suitability' based on the relation between land use and a broad set of biophysical and socio-economic factors. The basic assumption behind this mechanism is that a location changes into a certain land use in those locations where the 'suitability' is relatively high for that land use type. The suitability maps are constructed using the Action-in-Context (AiC) framework (De Groot, 1992) as described in Overmars and Verburg (2006) (Chapter 5). The 'deeper analysis' scheme of the AiC framework was used to structure the decision-making process of actors on the basis of their options and motivations and the structural and cultural

context that determines these options and motivations.
The second location characteristic is the neighbourhood effect (Verburg *et al*., 2004c). The neighbourhood function can mimic spatial processes in land use change that occur between a location and the land use type in its neighbouring cells, for example imitation and dispersion processes. However, this was not studied in detail in this research. Because the cell size of the application is smaller than the average parcel size, a small neighbourhood effect was implemented in the model for all land use types. This simulates the clustering of land use types into fields and patches.
The conversion mechanisms that can be incorporated in the model are the so-called conversion elasticities and land use type specific transition sequences. The conversion mechanisms can be used to assign influence to the land use history and determine the temporal dynamics of the model (Verburg *et al*., 2002). The conversion elasticities are implemented in the model as an increased suitability on those locations that are currently under that specific land use. Conversion elasticities can be explained as the resistance of a land use type to change location. For example, tree plantations will not easily be moved to another location because of the costs to do so, whereas arable crops can shift quite easily. The conversion elasticities that are incorporated are estimated based on field knowledge of the authors and are partly calibrated.
The transition sequence is a set of rules that determine the possible land use conversions. Not all land use changes are possible and many land use conversions follow a certain sequence. Sometimes these conversions include a temporal constraint and they can also be applied to a specific part of the study area only. The transition rules were used to incorporate the conservation measures of the scenarios in the models. Therefore, the detailed specification of the transition rules was incorporated in the scenarios section.

6.2.3 Data

Fieldwork was carried out between February and May 2004 by two observers working closely together. One was assigned to observe avian biodiversity in forest patches and the other concentrated on the other land use types. According to a stratification of the landscape from intensive agriculture to natural forest a number of study locations was determined and at those locations a total of 193 point counts of fifteen minutes were carried out (Bibby *et al*., 1992). To check the accuracy of detecting birds with point counts, the point count data was compared with the data from a mistnetting experiment. Per site between 0 and 4.5 percent of the species was only detected with mist netting (Thijs, 2005). This showed that the observers missed only few birds with point count observations and the point count data could be used for further analysis. The locations were spaced such that double counting was avoided. For each point count all visually and acoustically detected birds were recorded. The point count location was determined using a gps receiver. Time and weather conditions were recorded since these might have an influence on the occurrence of birds (Thijs, 2005). Furthermore, the observers recorded a variety of land use and habitat characteristics. This land use information was actually used to construct the location variables for the regression analysis, because the direct observations have a higher accuracy than the land use map.
The landscape characteristics for analysis were derived from the land use map. For every land use type the percentage of cover was determined within a radius of 250m and 1000m. This measures account for matrix influences and patch size. Patch size itself was not taken

into account. Furthermore, the total edge of forest and secondary forest combined was determined within a radius of 50, 250 and 1000 m. Shannon's diversity index of the surrounding landscape was determined for 50, 250 and 1000 radii (e.g. Dauber *et al*., 2003). All variables are described in Table 6.1.
Land use data were interpreted from three remote sensing images: a Landsat ETM+ image (http://www.landsat.org) from June 2001, an ASTER image from March 2002 and a SPOT image of July 2001. The first two images were first divided into a large number of classes by unsupervised classification. Subsequently, these classes were reclassified into land use classes using a set of 96 field observations. After this the SPOT image was used to improve the classification of wet rice fields. The classification of banana fields and secondary forest was improved using the NDVI of the SPOT image. Finally, the image was resampled to a 50 by 50 m grid that coincides with the other data (Figure 6.4 upper left). The accuracy of this map at the pixel level is 68 percent. A detailed description of the explanatory variables used for the suitability maps of the land use model is in Overmars and Verburg (2006) (Chapter 5).

6.2.4 Scenarios

The land use model is used to make projections of the landscape in maps under different scenario conditions for the period 2001-2015. The scenarios were constructed by combining information from policy documents concerning planning at the village, municipal and regional level, unstructured interviews with local stakeholders and field knowledge of the authors. In this chapter scenario refers to a story line and its quantification to an aggregated land claim that serves as an input for the CLUE-S model. The scenarios are projections of the future rather than predictions or forecasts (Rotmans *et al*., 2000). Quantification of the scenarios is based on interpretations of the storylines by the authors and basic data available to make general calculations. Four scenarios were developed: Scenario 1 assumes high agricultural expansion and scenario 2 a low expansion. For both scenarios two variants (A and B) were developed: one that has low forest protection and one with a high level of forest protection (Figure 6.3). For scenario 2 the two variants led to the same results. Therefore only scenario 2B is described in this chapter.

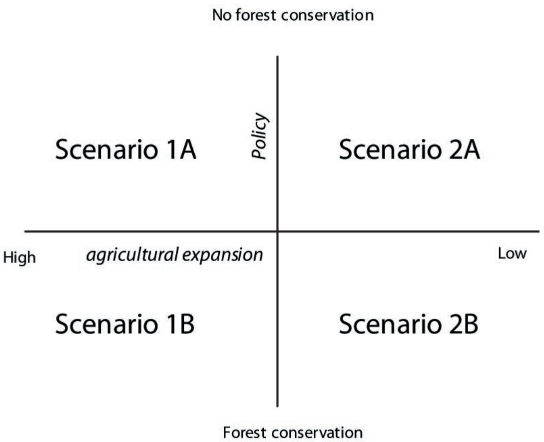


Figure 6.3: Four scenarios based on two levels of agricultural expansion and two levels of forest conservation

Table 6.1: Descriptives of the variables included in the endemic species richness analysis (n = 193)

Variable	Description	Min.	Max.	Mean	St.dev.
Dependent variable					
Endemics	Number of endemic species observed in 15min point count	0	11	1.91	2.46
Independent variables (local)					
Arable	Location has land use type arable	0	1	0.16	
Rice	Location has land use type rice	0	1	0.11	
Banana	Location has land use type banana	0	1	0.04	
Grass	Location has land use type grass	0	1	0.23	
Sec. forest	Location has land use type secondary forest	0	1	0.43	
Forest	Location has land use type forest	0	1	0.03	
Independent variables (landscape characteristics)					
Arable250	Fraction arable within 250m	0.00	0.86	0.36	0.23
Banana250	Fraction banana within 250m	0.00	0.59	0.12	0.12
Grass250	Fraction grass within 250m	0.00	0.77	0.20	0.12
Secondary250	Fraction secondary forest within 250m	0.00	0.70	0.14	0.18
Forest250	Fraction forest within 250m	0.00	0.91	0.09	0.22
Forsec250	Fraction forest and secondary forest within 250m	0.00	1.00	0.23	0.31
Water250	Fraction water within 250m	0.00	0.46	0.03	0.07
Arable1000	Fraction arable within 1000m	0.00	0.71	0.36	0.21
Banana1000	Fraction banana within 1000m	0.00	0.49	0.16	0.13
Grass1000	Fraction grass within 1000m	0.09	0.37	0.20	0.06
Secondary1000	Fraction secondary forest within 1000m	0.00	0.31	0.10	0.09
Forest1000	Fraction forest within 1000m	0.00	0.68	0.10	0.23
Forsec1000	Fraction forest and secondary forest within 1000m	0.00	0.91	0.20	0.29
Water1000	Fraction water within 1000m	0.00	0.20	0.03	0.04
SHDI50	Shannon's diversity index of the landscape within 50m	-0.61	1.54	0.62	0.42
SHDI250	Shannon's diversity index of the landscape within 250m	0.29	1.56	1.12	0.24
SHDI1000	Shannon's diversity index of the landscape within 1000m	0.88	1.72	1.28	0.24
TE50	Total edge within 50 m (km)	0.00	0.50	0.09	0.13
TE250	Total edge within 250 m (km)	0.00	3.20	0.98	0.95
TE1000	Total edge within 1000 m (km)	0.00	35.75	12.80	9.09

Scenario 1A: High agricultural expansion without forest conservation measures

Scenario 1 is a 'business as usual' scenario and projects the continuation of the current situation. Population growth will remain high and is projected to be three percent per year due to natural growth and in-migration. People in the area will remain highly dependent on agriculture. Out-migration as well as off-farm employment will not increase. The area used for the production of rice and arable combined is assumed to increase with 2.7 percent per year. The production of rice is determined to be self-sufficient for the inhabitants of the study area by 2010. The remaining projected increase in agricultural area will be realized as growth in arable land. Banana area is assumed to decrease yearly by 100 ha, because of problems with marketability and diseases. Forest is projected to decrease by 200 ha per year due to use of small-scale logging. The quantity of secondary forest is stable but dynamic because of clearing and regrowth. The remainder of the area is grassland. The land use claim is depicted in Figure 6.4.

No new land use policies are considered in this scenario and those that are present are considered to be ineffective. Thus, in this scenario land use can change without any policy constraints. All land use transitions that are biophysically possible are allowed (Figure 6.5A). Using the land use claim as model input does not control for the modelling of an increase and a decrease simultaneously (e.g. logging and regrowth of forest at different locations), because these cancel each other out. To include the typical land use dynamics of the area some land use conversions were forced to occur in the model (Figure 6.5A).

Scenario 1B: High agricultural expansion with forest conservation measures
This scenario has largely the same storyline as scenario 1A and will use the same land claim. The difference between the two scenarios is that in scenario 1B the park boundary will be fully respected, which prevents agriculture and clearing of forest in the park (see Figure 6.5B). This will result in additional pressure outside the park, because all changes to agriculture will be realized outside the park. Furthermore, the forced transition from secondary forest to arable (slash and burn agriculture, *kaingin*) was left out of the model to project improved forest management throughout the area.

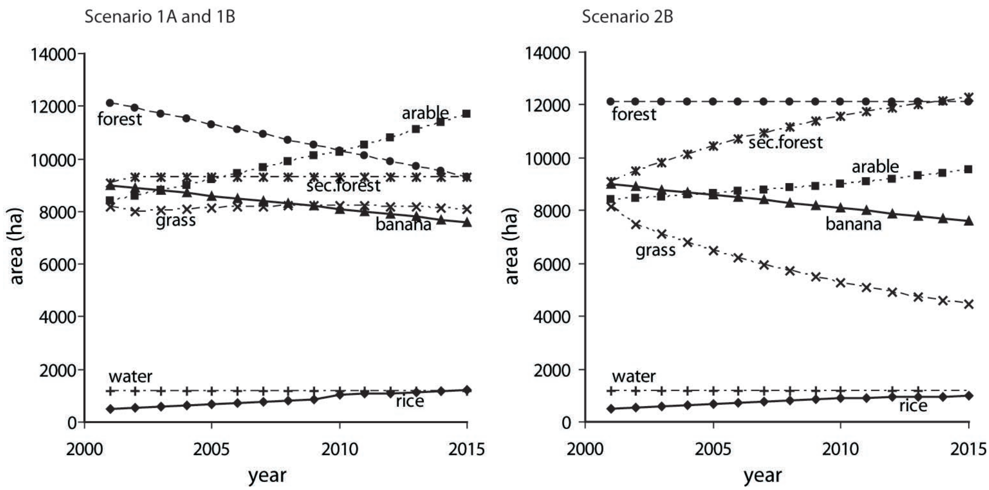


Figure 6.4: Land use claims for the scenarios 1A and 1B (left) and scenario 2B (right)

Scenario 2B: Low agricultural expansion with forest conservation measures
Scenario 2B has a different focus on the economic changes and population development. This scenario describes a situation with less agricultural expansion and improved forest conservation measures. Population growth is projected to be less than the current rate due to population control measures and a stricter immigration policy. People in densely populated areas that have a small piece of land will shift to agricultural systems that are more productive. In general, people become less dependent on agriculture. In scenario 2B total population growth is considered to be 1.5 percent resulting in an increase of the agricultural area with 1.2 percent per year. Like in the other two scenarios, self-sufficiency in rice is accomplished from 2010 onwards and the remaining agricultural growth is in arable crops and banana will decrease with 100 ha per year. In this scenario secondary forest will increase due to significant regrowth of secondary forest on grasslands. This

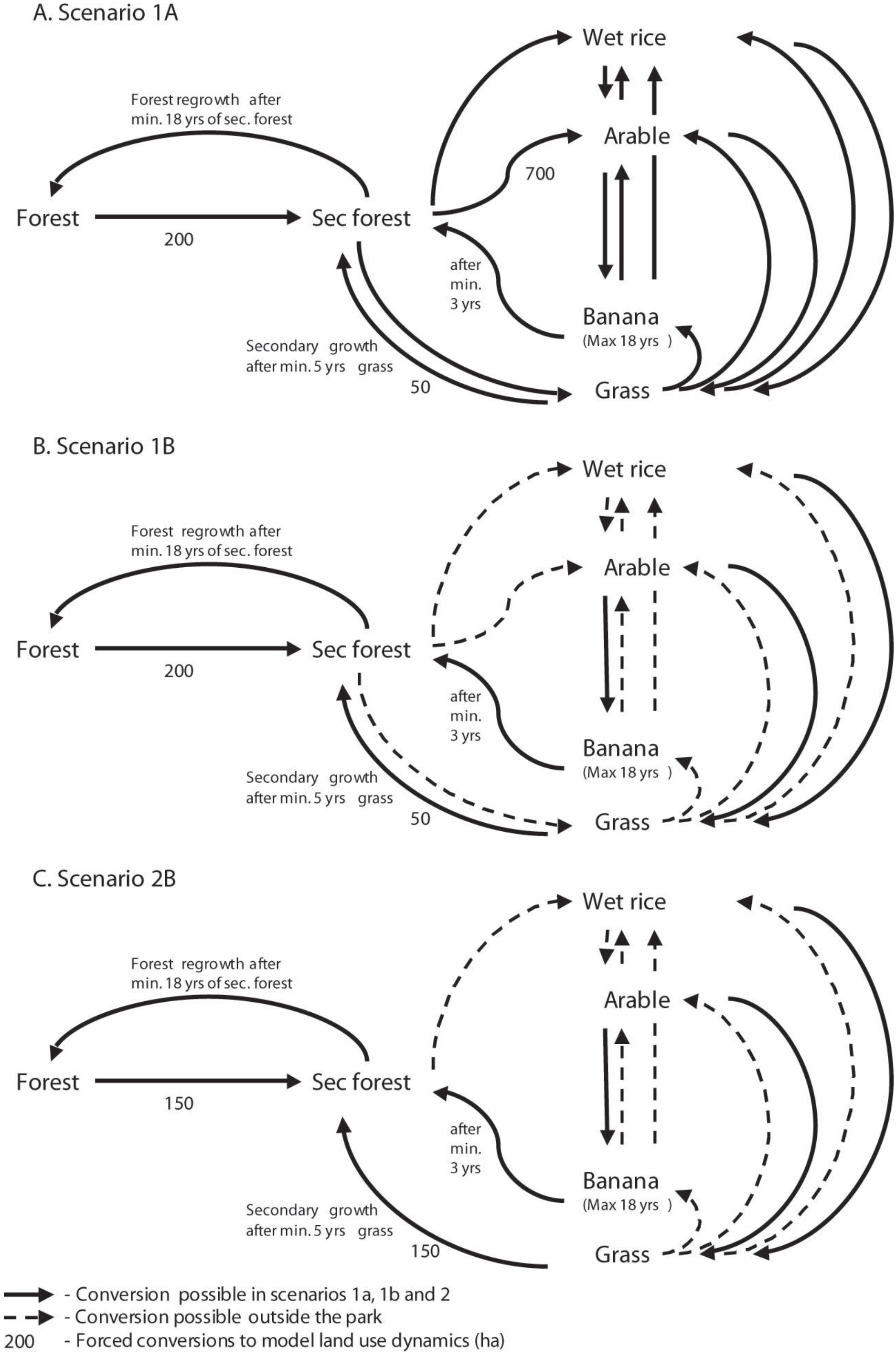


Figure 6.5A,B,C: Conversion rules included in the CLUE-S model

is possible because this scenario assumes that uncontrolled burning of grassland areas is prevented and that people reforest the grasslands, which is speeding up the regrowth of forest. Secondary forest is assumed to increase proportionally to the amount of grass with an average of 230 ha per year. Grassland makes up the remaining area. The land claim is in Figure 6.2.
As in scenario 1B the park boundary will be fully respected: agriculture is not allowed in the park. Although this scenario is positive for the environment still some illegal logging is included (Figure 6.5C). The area of forest is projected to be stable due to regrowth from secondary into 'mature' forest. Opening up secondary forest for arable agriculture, is strictly prohibited and enforced not only within the park (Scenario 1B), but also in areas under secondary forest outside the park (Figure 6.5C). So, in principle development of new agricultural area is only allowed in the idle grasslands. Conversions from secondary forest to grassland are also prohibited in the model.

6.3 Results

6.3.1 Endemic bird species richness

Statistical relations between landscape characteristics and endemic bird species richness are presented in Table 6.2. The best single predictor was the amount of forest and secondary forest within 250m of the location. The next best predictors were all local variables representing the land use type. The land use variables forest, secondary forest, banana and grass were included. The interpretation of the regression coefficients of the land use variables is therefore relative to the agricultural land use types arable and rice. Including more predictors with the stepwise procedure would result in models with collinearity between variables. The residuals were tested for correlations with observed cloud cover, precipitation and time of observation. These factors might have influenced the observations. However, no correlations between these factors and the residuals were found.
To test for overdispersion the ratio of the deviance and the degrees of freedom was calculated. This ratio should be close to 1 (Crawley, 1993; Gibson *et al.*, 2004). If not, the model is overdispersed and the assumption of Poisson errors is not valid. In the model presented this ratio is 1.09, which is relatively low and we consider the assumption of Poisson distributed errors to be justified.

Table 6.2: Poisson regression results

Variables	b	sig.
Intercept	-2.89	0.000
Forsec250	1.43	0.000
Sec. forest	3.52	0.000
Forest	3.62	0.000
Banana	3.64	0.000
Grass	1.82	0.003
dev	177.96	

6.3.2 Land use projections

In Figure 6.6 the land use map of 2001 and the projected land use maps for the three scenarios are depicted. Comparing scenarios 1A and 1B (Figure 6.6 lower left and lower right, respectively) it is clear that the inclusion of the restrictions in the nature park influences the spatial distribution of the forested area. The total area per land use type is exactly

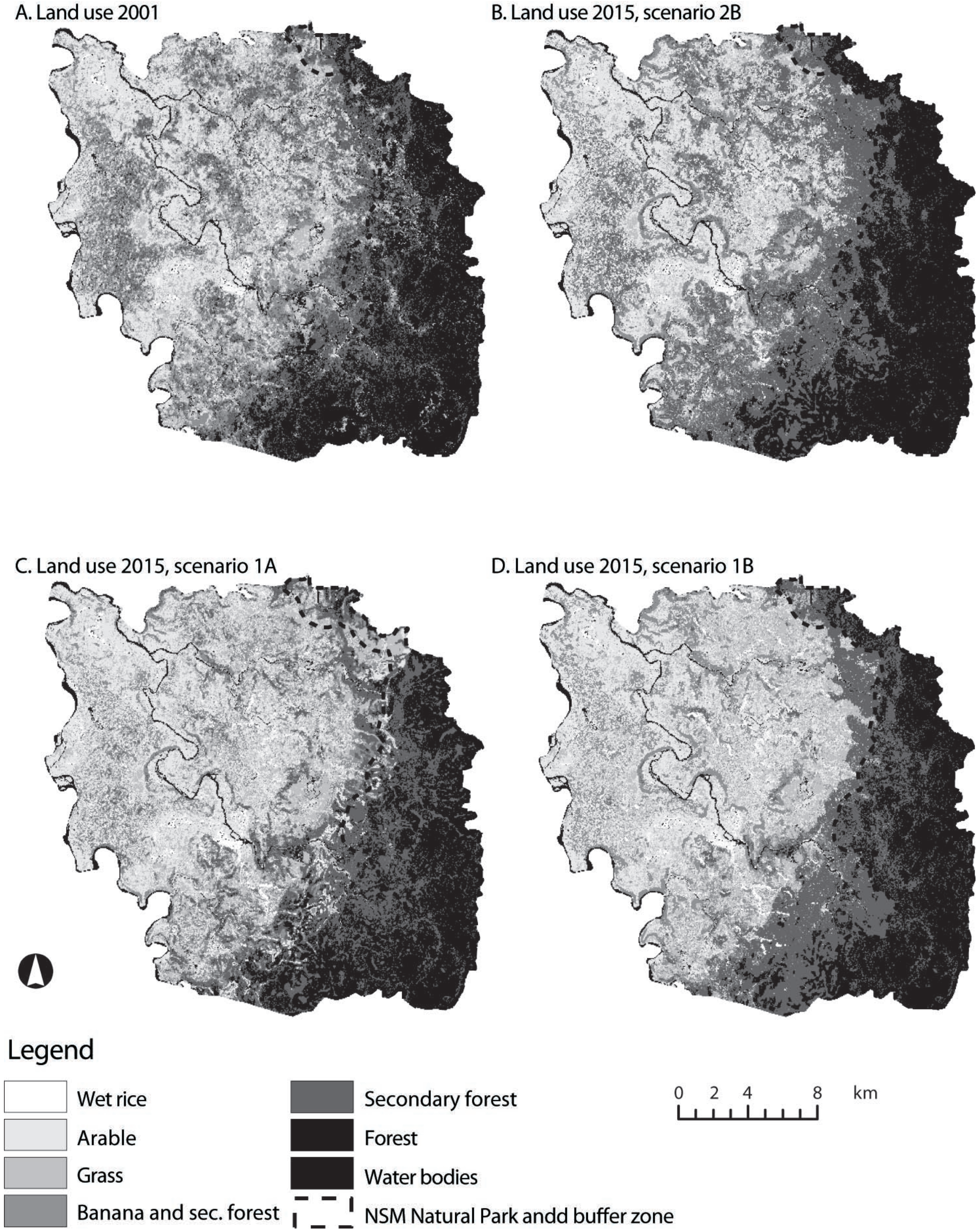


Figure 6.6: Land use map of 2001 and land use projections for the three scenarios

the same in these two scenarios; only the allocation of these land use types is different. Without park policy the northeastern part of the forested area is under threat and the forest edge shows a lot of disturbance. Regarding agriculture both scenarios 1A and 1B clearly show that extending activities in agriculture will occur mainly in the western and central part of the study area. This is caused by the accessibility of the market, but also by the unfavourable physical conditions in the eastern part. In this respect logging is more of a threat to the forest than is expansion of agriculture, although these can go hand in hand. Scenario 1B shows that the park policy diverts the pressure for agricultural use even more to the central and western part. In scenario 1B all secondary forest disappears from this area, whereas in scenario 1B secondary forest patches remain in this area. Scenario 2B shows some intensification of the agricultural areas at the expense of grasslands. Many areas show regeneration of grass to secondary forest. The park is protected and forest in the park is increasing. In the central area in the south some forest is converted to secondary forest, because the scenario includes some illegal logging. On other parts secondary forest turns into 'mature' forest. Both removal of forest by logging and regrowth of forest from secondary forest occur in the model, but the total forest area was projected to be stable. So, the net effect is that forest is only changing location.

6.3.3 Avian endemic species projections

Spatial projections of endemic bird richness were created by combining the land use projections with the relations between endemic species richness and the landscape derived through the regression analysis. A map with the endemic species richness index for 2001 is presented in Figure 6.7A. This figure shows that endemic species richness is highest in the eastern part, which is covered with closed forest. In the cultivated landscape in the west and central part of the study area endemic species do occur in secondary forest and banana/secondary forest patches, but at these locations the endemic species richness is less than in dense forest. To better visualize the differences in endemic species richness between 2001 and the year 2015 for the three scenarios, difference maps are presented instead of the index maps (Figure 6.7B, C and D). Scenario 1A shows a decrease in endemic species richness throughout the area with a larger decrease near the forest fringe. Especially in the northeastern part of the area this scenario projects major land use changes from forest into agriculture and grassland, which has a large impact on the biodiversity on this location. Scenario 1B, including the forest conservation policies, shows a larger decrease of species richness in the cultivated area (west and north in the study area) compared to scenario 1A and less near the forest fringe due to protection of the natural park. Scenario 2B generally shows an increase in species richness, due to the projected regrowth of forest and secondary forest. However, in the south central area a slight decrease of species richness occurs due to some small-scale logging that was projected in this scenario.
The difference maps are useful to make a general assessment of the relative changes in endemic bird species richness in the area. For conservation purposes it can be important to look more specifically to certain group of species or to one species only. The index presented can be used for this purpose. In the case of endemic bird species in this study area the species found in locations with a low index (the western part) are species that are also observed in the areas with a high index. The extra species one will observe in an area with a high number of endemic species are not present areas area with a low index. These species are more dependent on a specific habitat (*i.e.* the forest), and may even be threatened, and

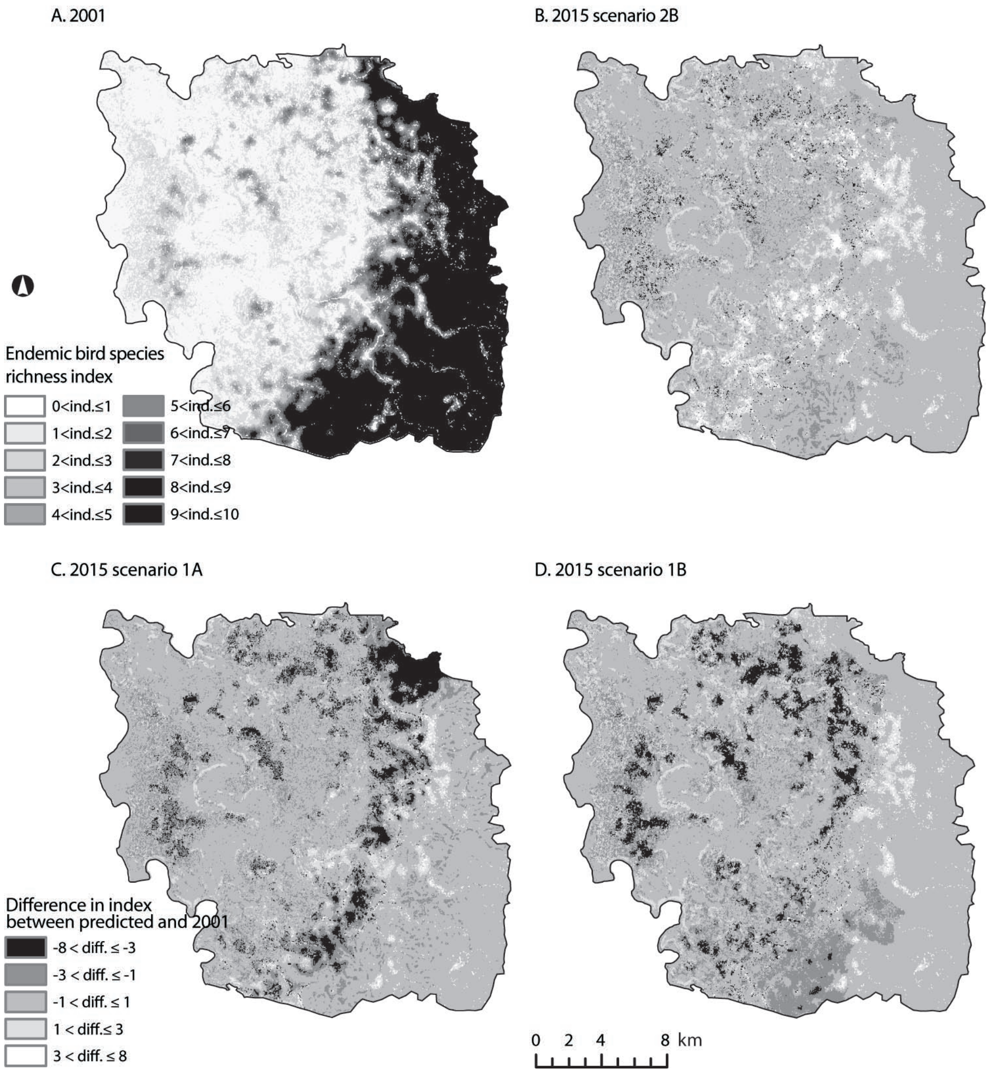


Figure 6.7: Endemic bird richness projection for 2001 and changes for the three scenarios

therefore it is important to conserve landscapes that can support these birds. To assess the effects for the areas with a high number of species the maps were reclassified into areas with index greater than or equal to 6 and with an index less than six. The changes from one category to the other were calculated for the scenarios 1A and 1B (Figure 6.8) as an illustration to analyse the spatial output for a species group of a high conservation importance. This analysis shows that the two scenarios are very different for the changes of the class with index six and higher. Scenario 1A has a much larger area that changed from class six and higher to less than six than scenario 1B. Even though the total effect looks similar (Figure 6.7C and D), from a conservation perspective it is very important where this species loss occurs since this determines which species are lost.

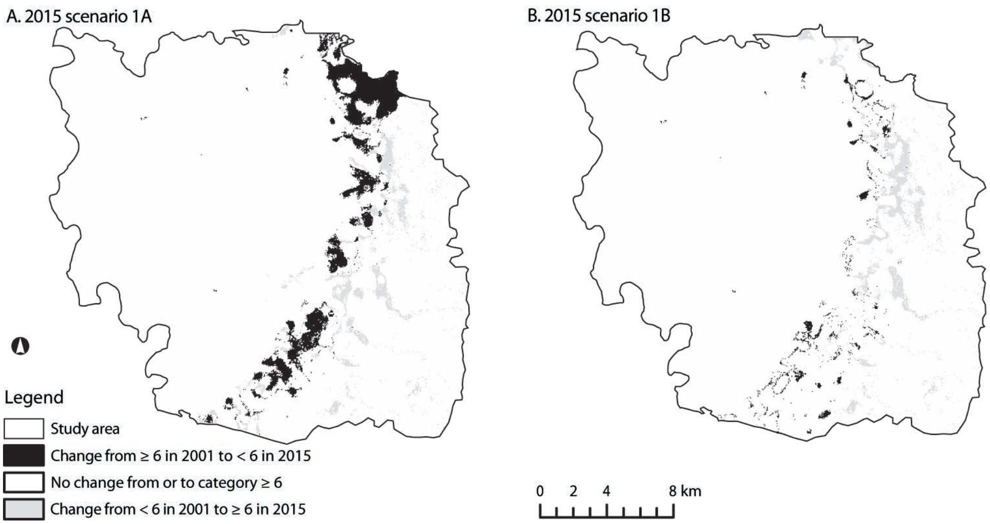


Figure 6.8: Changes from = 6 in 2001 to < 6 in 2015 (black) and vice versa (gray) for scenario 1A and scenario 1B

Table 6.3 shows a summary of the changes for the different scenarios. In this table the index for species richness is classified into 9 classes to facilitate comparison. The drawback of this table compared to the maps of Figure 6.7 is that does not reveal shifts in the location of biodiversity. The table provides a net effect per class (Table 6.3A). Considering that the higher classes are more important than the lower classes one can estimate the effect for nature conservation. The effect is even clearer if the areas per class are summed (Table 6.3B) starting at class 8-9. The accumulated areas show what the effect is for the habitats of birds that fall into a specific class under the assumption that birds in a lower class always occur at to locations with a higher class (*i.e.* a bird that occurs in location with class 3-4 will also occur in classes 2-3, 1-2 and 0-1). For example, the suitable habitat for birds that occur in class 7-8 increases from 13054 to15672 hectares under scenario 2B.

Table 6.3: Area per class of endemic species richness for 2001 and the three scenarios (ha) and accumulated area (starting with class 9) representing the total area per class where the bird species of this class can be found

	A. Area (ha per class)				**B. Accumulated area (ha)**			
Index class	**2001**	**Sc.1A 2015**	**Sc.1B 2015**	**Sc.2B 2015**	**2001**	**Sc.1A 2015**	**Sc.1B 2015**	**Sc.2B 2015**
0 < ind. ≤ 1	17970	22148	22157	16111	48416	48416	48416	48416
1 < ind. ≤ 2	335	80	25	97	30446	26268	26259	32305
2 < ind. ≤ 3	7814	7004	6836	5802	30110	26189	26234	32208
3 < ind. ≤ 4	3550	1999	1033	4044	22296	19185	19397	26406
4 < ind. ≤ 5	1707	1496	926	2615	18747	17186	18365	22363
5 < ind. ≤ 6	1706	1434	1224	1958	17040	15690	17439	19748
6 < ind. ≤ 7	2280	1388	2560	2118	15334	14257	16215	17790
7 < ind. ≤ 8	4234	4695	5488	4636	13054	12868	13655	15672
8 < ind. ≤ 9	8820	8173	8167	11036	8820	8173	8167	11036

6.4 Discussion and conclusions

The relation between the occurrence of endemic species and the landscape was explained with a model that incorporates variables of land use (habitat) at the location itself and the fraction of forest and secondary forest within 250 m of the location. Besides the last mentioned relation the analysis did not show any other clear relations between endemic species occurrence and the spatial landscape characteristics. An explanation for this is that a number of spatial landscape characteristics are correlated. Therefore, these variables cannot be included together in the same model because this would introduce collinearity. On the other hand, land use (habitat) itself is an important determinant of species occurrence and already explains a large part of the variability in the occurrence of endemic birds. A number of the landscape characteristics were simply not significant in a model that already included the land use variables. A similar conclusion was drawn by Fairbanks (2004); this study, though at a much larger scale, also concluded that most relations were explained by land use and spatial metrics played a secondary role.

This study had its focus on a transition zone from intensive agriculture to the forest fringe and the results are valid for that area. In the entire contiguous forest, which is largely outside the study area, a larger number of endemic species occurs (Van Weerd, 2002). A predictive model including more observations in the forest could be quite different, since the forest exhibits also differences in the occurrence of endemic birds. These differences did not show in this study because the model only includes the forest fringe. However, even from this study it is clear that forest protection is of critical importance to conserve this group of species. The mosaic landscape of agriculture and natural habitat has relatively low contribution to the conservation of endemic forest species in the study area.

In this study one specific indicator was used to create a map of avian biodiversity, but one could think of many others. Examples of alternative approaches are to model total species richness of birds or to model one target species. Another approach is to create species habitat maps for a number of target species and combine this into one map (Store and Jokimäki, 2003). This approach could also include species from different taxa (Dauber *et al*., 2003). Others (e.g. Luoto *et al*., 2004) argue to combine species richness with species distribution models.

It is important to use appropriate indicators in assessing biodiversity in landscapes for policy-making. Conclusions regarding biodiversity can be very different or even opposite for different indicators. Therefore, it is important to choose the right indicator(s) for optimal conservation and land use management. For example, increased habitat diversity will often enlarge the total number of species (Atauri and de Lucio, 2001; Steiner and Koehler, 2003; Luoto *et al*., 2004), but threatened species might need a completely different landscape, like large patches of forest habitat (Luoto *et al*., 2004). Furthermore, species from different taxa may even be more different in their preference for habitat and landscape conditions. Dauber *et al*. (2003) did not find any correlation between species richness for the three taxa they included in their study.

This study had its focus on endemic species richness of birds because on the one hand it specifies a group of species that is of interest for nature conservation and on the other hand is more general than having one target species. However, the case study results should be interpreted with care regarding their use to conserve biodiversity in general. The results indicate the effects for endemic bird species richness. To assess the actual impact of management strategies based on this study these policies should also be evaluated for other species.

The land use model used in this study is an appropriate tool to make projections of future landscapes under different scenarios. The strength of the CLUE-S land use model is to allocate a predefined land claim. This land claim is specified outside the model. Thus, the effect of the policy measures on the quantities of land use change is not calculated in the model. In the case of scenario 1A and 1B, with and without forest protection respectively, the total amount of land use changes is the same. The analysis shows solely the effect of a different allocation. Including the effect on the land use claim would require an extra analysis preceding the CLUE-S analysis, in which the change in land claim due to park protection would be determined.
The scenarios are a projection of events that may happen. Many more scenarios could be created according to insights of stakeholders and policy-makers. However, in this respect it is important to realize that this scenario study is not a visualization of plans. In that case the maps of the development plans could have been assessed on their value for biodiversity. The scenario study is a visualization of a possible future and assumes that a large part of the land use system functions autonomously and cannot be planned.
The result from this case study show that forest conservation leads to a different land use pattern and that these differences are relevant for biodiversity conservation. This effect was most prominent in the scenarios 1A and 1B that project a high rate of agricultural expansion. The scenarios 2A and 2B were almost the same because these scenarios did not project high pressure on the (secondary) forest. This scenario study shows that in those areas that are currently under pressure (scenarios 1) conservation policy can result in a more favourable landscape even if the total areas per land use types are the same.

The use of this combined methodology is that it can quantify effects of land use change for biodiversity. Understanding of the land use change process including its human component is important to be able to exert influence to this system and to change its course for the benefit of biodiversity conservation (Henle *et al*., 2004). Land use modelling can visualize the future landscape pattern-based on a set of demographic, economic, technological, cultural and policy drivers (Geist and Lambin, 2002) and can incorporate the influence of conservation management policies. The land use modelling approach can identify specific hotspots of land use change; the combination with a biodiversity assessment can identify specific hotspots of biodiversity change. To pinpoint these priority areas it is not only necessary to know where certain species occur, but also if this area is under threat, which can be provided by the land use model. The identification of hotspots can be used to prioritise nature conservation efforts by introducing policies that influence land use changes in these hotspots. The analysis depicted in Figure 6.8 shows that the hotspots of change in biodiversity, which are relevant for nature conservation, can be different from the pattern of land use change, because this is co-determined by the occurrence of certain species.
For nature conservation in practice it is important is to translate research results into rules and guidelines for applied research and practical tools (Henle *et al*., 2004). Coupling of land use models and biodiversity research can be used as a tool to support management decisions in nature conservation in practice. Several applications that couple land use models to biodiversity assessments have been carried out (Menon *et al*., 2001; Eppink *et al*., 2004; Jepsen *et al*., 2005). Eppink *et al*. (2004) and Jepsen *et al*. (2005) provide biodiversity values for the landscape as a whole, but do not present their biodiversity indicators in a spatial explicit way. Menon *et al*. (2001) use a land use model to identify priority areas for nature conservation based on threats to natural habitat and the current protection status.

The strength of the methodology presented in this chapter is that the land use changes are projected in a dynamic, spatially explicit way including competition of land use and the effects of land use policies and that subsequently these land use changes are translated into changes in biodiversity (for a specific indicator) in a quantitative, spatially explicit way. With this method different conservation policies can be quantitatively evaluated and spatial policies for optimal conservation management can be tailored to the area of interest.

7

General discussion and conclusions

7.1 Land use in San Mariano

This dissertation has a strong methodological focus. Nevertheless, some substantive conclusions for the local case in San Mariano can be drawn from the analyses. The analyses of the explanatory factors of land use change (Chapters 2, 3 and 4) revealed the main processes that determine the distribution of various land use types in the area. These explanatory factors can be categorised in three groups: accessibility, origin of the people and biophysical constraints.

For cash crops, for example yellow corn and banana, accessibility to the market is an important factor because it determines the transportation costs. Furthermore, lack of accessibility in the wet season can impose restrictions to the cultivation of cash crops. During the wet season transportation is difficult or even impossible and storage of the yield is often not an option because appropriate storage facilities are lacking. The distance between the field and the place of residence is also an important determinant of crop choice because people prefer to live close to their fields (Verburg *et al*. 2004a). Improvements of accessibility and reduction of transportation costs can improve the livelihood of people in the area because it would increase profitability of the crops they produce. However, better accessibility also increases the access to the forest for illegal logging activities. Improved accessibility will also attract more people, because locations further from the market become accessible to start a farm.

Other important determinants of what crops people cultivate are ethnicity and migration background. Based on preferences and habits some ethnic groups prefer to produce corn and other prefer to produce rice. Generally speaking, migrants, defined as people born outside the municipality, are more involved in growing rice and less in growing corn than people that were born in San Mariano. This relation is partly caused by the fact that migrants settle in places far away form the market town where growing corn is less profitable. Another reason is that the people that currently migrate to the area are of different ethnic background than the people that have a longer history in the area. These new migrants often prefer to cultivate rice rather than corn. A third reason is that the migrants often do not have capital to invest in inputs for cash crops and therefore start with rice production to be self-sufficient for their families.

For agricultural production the flat and rolling areas are the most suitable. Especially arable crops are preferably cultivated on the flat parts. Perennial crops, for example banana, are also cultivated in the rolling parts. For the production of irrigated rice an extra requirement is that it needs a source of water. This can be a creek or an irrigation facility.
If one compares the allocation of the two most important cash crops in the area, which are corn and banana, in relation to their profitability alone it seems that a relatively larger area is planted with corn than would be expected (Verburg *et al*., 2004a). This has two reasons. One reason is that approximately once in five years the area is hit by a large typhoon. The damage of these typhoons is such that the bananas will not bear fruits for a year. If this damage is taken into account the calculated yearly profits from banana are actually lower. Another important reason is that corn production in the area is most often financed with credits from traders that provide the seeds and agrochemical inputs. On the one hand many farmers are actually happy to have access to a source of money and use the credits also for consumptive use. On the other hand to farmers are often indebted for a longer time and therefore forced to grow corn (Van den Top, 1998). Therefore, more corn is grown in the area than could be expected from a narrow, purely economic point of view.
From scenario analysis in Chapter 6 the following can be concluded. The forested area in the northeastern part of the study area, which is part of the natural park, is most under threat under a high growth scenario without forest protection. Most of the remaining forest, however, has a kind of natural defence, because the slopes are steep and not suitable for arable production. The scenario study shows that it is possible to increase agricultural production by making use of the grasslands in the hilly part, where land is still available. In that case the forested parts may be spared. A negative scenario, which already can be observed in other parts of the Philippines, is that all forested areas will eventually be cleared for agriculture. However, taking these areas into production often provides only temporary solutions for the smallholders because without appropriate conservation measures the soils in the mountainous areas are prone to erosion, which eventually will lead to less productivity.
The key to a sustainable future in San Mariano, apart from larger economic and political developments in the region, is migration control and a strict implementation of environmental legislation and land use policies. As can be seen from scenario analyses (Chapter 6) the area that is already cleared from forest and which is currently under grass can be used to facilitate a large part of the increase in agricultural production. To create viable opportunities to cultivate on these marginal grassland the people should be assisted to implement sustainable agricultural practices. Even more important, these measures should go hand in hand with regulations to prevent the agricultural area from expanding into areas that are currently under forest. For example, improvements of the accessibility of the area will increase the opportunities to make a living for the current population, but will also attract people that will enter the marginal areas that have become more accessible. Therefore, improvements in accessibility should come with regulations to prevent extension of the cleared area and creation of new settlements. A constraint in taking the grasslands into production is that most of these areas are claimed by local farmers. This factor was not included in the model and may actually prevent these areas from being taken into production by others than the owners. Therefore, most of the immigrants currently open up their own areas to claim land. To enable immigrants to make use of the grasslands the land has to be made available to them at low cost. An alternative is to stop immigration and the creation of new settlements. In that case it is to be expected that agricultural expansion

will mainly take place in the grassy areas because these areas are owned by the families that live there and no new land has to be cleared. The municipality of San Mariano could benefit from clear decisions on land use planning by assigning some areas to agriculture and investing in these areas to enable people to make a good living, and assigning other areas for the conservation of the sparse forested areas left. However, this approach also requires means to invest in sustainable agriculture as well as implementation of the laws and policies that officially are already operative.

7.2 Methodologies for land use science

In this section general conclusions are presented regarding methodology for land use science. Differences between land use analyses in the same area by using different disciplinary approaches (Chapter 2) stem from differences in unit of analysis, differences in sample design, differences in the themes included and differences in the methods that were used to collect the data. Analyses from a specific disciplinary perspective can help to understand part of the land use system, but cannot explain the complete system. Therefore, propositions should be stated carefully and clearly and should provide explicit information about the unit of analysis, the sample design and the included variables. In Chapter 2 the field is included in the household analysis to have a similar unit of analysis as the geographical approach. Thematically, the variables included in both models are the same. Nevertheless, the results are different due to differences in sample design and data collection.
Land use research covers a range of approaches between inductive and deductive. Inductive, often statistical, approaches can provide correlations between land use and explanatory factors. These statistical analyses can be used to fit the data as well as possible. Deductive theoretical frameworks can specify relations between explanatory variables and the subject to be explained in a flexible way and are capable of representing causal relations. By validating theoretical frameworks (Chapter 3) the full causal structure is tested, which leads to a better understanding of causality and supports theory building.
Multilevel analysis (Chapter 4) is a valuable tool in land use research to unify different scales and levels in one statistical analysis. Disciplinary approaches often have different units of analysis, which hampers the comparison or combination of various methods. Because multilevel analysis allows the incorporation of variables at different levels, different units of analysis can be combined. Aggregating or disaggregating variables to the unit of analysis, which has statistical disadvantages, is therefore not necessary. This method allows a multitude of propositions between higher and lower levels to be made and testing of the relations between levels and scales. The case study revealed the importance of the household level in explaining land use at a detailed level in the study area. Generalising this, it can be concluded that all organisational levels between the resolution and the extent should be examined for their importance in explaining land use.
A wide category of land use studies apply an inductive pattern-based method to identify drivers of land use and, subsequently, to use these drivers in spatially explicit land use models. As such this is a valid method, though the approach brings about a number of restrictions. Inductive approaches are weak in the description of causality and processes. This restricts models that are based on an inductive analysis for modelling large changes in processes, for example the introduction of a new land use type. An alternative is to use a theoretical, process-based approach, for example an actor-decision framework (Chapter

5), to derive and describe relations between land use and its explanatory factors. If such an approach is used the models can be made more flexible regarding the introduction of new land use types and changes in processes during the modelling period. This approach is most valuable in small study areas where detailed actor research can be carried out and can be used for detailed policy analysis. Large-scale studies can best be carried out with an inductive approach, for example based on statistical analysis relating land use patterns to a number of explanatory variables available in maps. Their use is more in modelling general trends, for example in the identification of hotspots of land use change. The two different approaches to specify spatially explicit land use models each have their consequences for use in policy-making. The choice of one approach or the other depends on (1) the research question and the policy context of the study (Chapter 5) and (2) the wider use of the study in scientific programmes and theory building (Chapter 3).

In order to use land use models in policy-making effectively the projections of future land use patterns should be translated into normative indicators that describe consequences for biodiversity, agricultural production and watershed properties, for example. To do this additional studies have to be carried out that link land use changes to their effects. The pattern of the effects of land use changes may be different than the locations where the land use changes occur for two reasons. Land use changes may have off-site effects, which implies that land use changes at a certain location has an impact on other locations than itself, for example downstream. Secondly, land use changes may have different effects in different locations. For example, a land use change in a location with a low value for biodiversity has less impact for biodiversity conservation that the same land use change in a locations with a high conservation value. In general, the locations of land use change do not necessarily have to be the same locations as where the effects take place. Moreover, the effects of land use changes may in themselves influence land use decisions. Therefore it is important to assess the effects of land use changes and incorporate feedbacks of land use change into land use studies.

7.3 Value of the combination of approaches in the presented study

In this thesis it is the combination of approaches that have led to a greater understanding of the land use system in the study area. A summary of the key methodological components is in Figure 7.1. The empirical data can be categorised in three parts: Qualitative data from unstructured interviews, quantitative data from a household survey and a spatial dataset. The approaches for the collection of these three datasets have their origin in different research paradigms and cover the fields of qualitative gamma sciences, quantitative gamma sciences and geography respectively. Furthermore, the horizontal dashed lines (Figure 7.1) indicate the three research categories indicated in Chapter 1: Observation and monitoring of land use change (top), identification of the drivers of land use change (middle) and dynamic modelling of land use change (bottom).

The first analyses were carried out in a rather disciplinary way, but aimed at facilitating comparison and exchange of information. The study started with a statistical analysis of the household data (Chapter 2). By using the field as the unit of analysis this household analysis could be used to inform the statistical analysis of the spatial data by including the themes that were important explanatory factors at the household level. The household analysis was also used to inform the descriptive Action-in-Context (AiC) analysis (Chapter

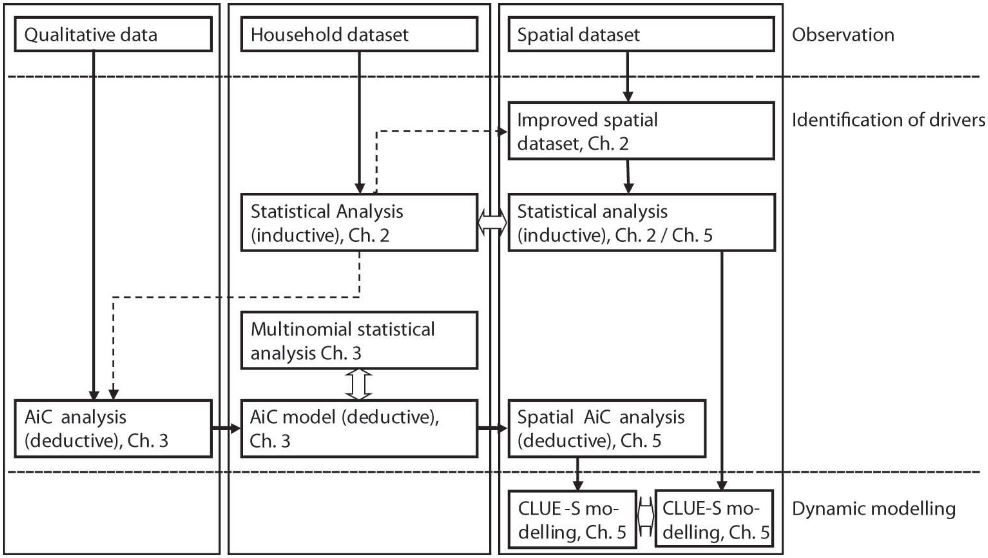


Figure 7.1: Sequence of methodological approaches throughout the present study. Solid arrows indicate direct links, dashed arrows show informing links and the wide arrows indicate comparisons

3), which was mainly based on the qualitative information. Subsequently, the AiC analysis was used to construct an actor decision model (AiC model) that was validated with the household data. Finally, the AiC analysis was translated to a spatial approach and used to parameterise the CLUE-S land use model (Chapter 5). In this way the AiC analysis serves as a cross-level and interdisciplinary approach. (Chapter 4 (multilevel analysis) is not included in Figure 7.1).

Many land change scientist seek the development of theories of land use change. As discussed in Chapter 3, moving up and down between empirical, inductive methods and theoretical, deductive methods as described above is a useful approach to stimulate theory development. Relations derived in inductive approaches can be used to structure the process of theory building. Subsequently, theoretical causal structures can be developed and tested. From there, one can move a step back in the direction of induction by calibrating the theoretical model with empirical data to get a better model fit. This process could go through several steps moving up and down between inductive and deductive approaches to improve the theories and the models to describe the land use processes. In this respect the study that forms this dissertation is an example of the approach described in Chapter 3.

With the presented combination of approaches it was possible to combine various disciplinary aspects in the AiC model. Applying different methods forces the research into an integrative approach. In this way we were able to construct an interdisciplinary AiC model and to incorporate process information in the spatially explicit land use model. Without this mix of approaches the analyses would have been dominated by one disciplinary paradigm and a truly integrative approach of the various aspects of the land use system would have been difficult.

7.4 Lessons learned from performing interdisciplinary research

As argued before, land use is a field of research that involves many disciplines. To come to integrated land use studies it is inevitable to do multidisciplinary or interdisciplinary research. Multidisciplinary research refers to a group effort in which a number of disciplines are represented which exchange information, for example between disciplinary models. The important difference with interdisciplinary research is that in an interdisciplinary study new paradigms and methods are developed apart from the existing disciplinary approaches and which have a position in between the contributing disciplines. This paragraph reports on the experiences of working on an interdisciplinary study. In retrospect these experiences are largely the same as formulated by Pickett *et al*. (1999) and Schoenberger (2001), which demonstrates these experiences have a kind of universal value.
In itself, the present study aimed at integrating disciplines but the project was also carried out in a larger interdisciplinary program. Cooperating within a group with people from various backgrounds turned out to be crucial in carrying out truly interdisciplinary research. Generally speaking, people are educated in a disciplinary way, which will lead to a disciplinary bias in their thinking and their approach to handle a research question, whether consciously or unconsciously. Involvement of people from different backgrounds will automatically lead to different perspectives. Thus, interdisciplinary research should be a group process. A drawback of a group process is that it will cost more time than working alone or in a disciplinary group. It takes time to understand each other to reach consensus about the way to proceed.
A number of issues can be identified that can contribute to the interdisciplinary research process. First of all, it is important to have a common problem or research question. Loosely formulated, in this project the common research question was: Why do people manage their land the way they do and why at that location? Secondly, it is important to have a number of methods in common. In this project an actor decision-making framework was adopted that was used by two of the researchers. Pickett *et al*. (1999) argue that both deductive and inductive approaches should be part of interdisciplinary research. Mature specialities have often well developed theories but less developed disciplines and interdisciplinary research can often benefit from inductive approaches. As was concluded in Chapter 3, inductive research can help to identify the factors that are important to land use and this information can be used as a guide in constructing mechanistic and causal hypotheses. Thirdly, having a research question and methods in common will help to develop a common vocabulary, which is very important for clear communication and cooperation. Disciplinary researchers all have their own research culture, where the meaning of words and concepts are known, but which cannot always easily be understood by others. A fourth help in interdisciplinary research is to have a common research site and, if possible, to share data.
Interdisciplinary research involves incorporating contradictory and ill-fitting elements. Unless an all-encompassing theory that includes all disciplines in a balanced manner is available, disciplines and disciplinary methods will compete for their position in a study. Disciplinary 'truths' and certainties are questioned and may even be violated. The most general theories in each of the contributing disciplines are often too abstract to link with other disciplines (Pickett *et al*., 1999). So, in interdisciplinary research it is necessary to scarify some of the disciplinary detail in order to establish links with other disciplines.

7.5 Perspectives in land change science

The LUCC project (Turner *et al*., 1995; Lambin *et al*., 1999), which played an important part in the development of land use science, started in 1995 and ended in October 2005. The LUCC project is succeeded by a new initiative called the Global Land Project (GLP, 2005). This provides the opportunity for renewed agenda setting for land change science. Below some of the thematic and methodological issues that are currently identified for the research agenda are discussed.

Two examples of thematic topics that need the attention of the land science community are globalisation and vulnerability. The locations of production and locations of consumption are disconnected more and more (*i.e.* globalisation) due to, amongst others, migration and urbanisation (Turner *et al*., 2004; LUCC scientific steering committee, 2005). Land change science has to include the links between local and global developments. A theoretical example that rewrites Von Thünen's theory in this respect by disconnecting the locations of production and consumption is in Walker and Solecki (2004). In this thesis the case of corn production shows that even for a distant frontier system the influence of goods for a distant market is large. Growing urban populations in the Philippines increase the demands for feed-corn, which is produced in the uplands, through increased meat consumption (Coxhead and Buenavista, 2001). Currently, the corn market is protected in the Philippines (Coxhead, 2000). Liberalisation of this market, which is currently considered, can lead to large shifts in the crops produced because feed corn may be imported instead of produced in the Philippines.

Vulnerability of society and ecosystems is another issue that needs more attention (Turner *et al*., 2004). Land use research tended to focus on slow variables and underlying factors. However, land use systems are also to a large extent determined by extreme events (both human and biophysical). Extreme events determine the resilience and collapse of systems and thereby the system's vulnerability (LUCC scientific steering committee, 2005). In the study area droughts and typhoons have important direct effects for the population in the area, but also largely determine the functioning land use system as such (Huigen and Jens, n.d.). Many of the land use decisions in the area cannot be explained without taking the extreme events into account. For example, for the case of banana Verburg *et al*. (2004a) found that based on accessibility banana would be more profitable than corn. However, in most instances corn is preferred above banana. A part of the explanation for this paradox turned out to be that the area is regularly hit by typhoons, which destroy the bananas and therefore impede production for more than a year.

The LUCC science community has developed a wide range of tools and methods to use in land use studies. Below we discuss some methodological issues that are currently advocated as being the way to proceed in land change science.

It is widely acknowledged that land use research is to be carried out in a comprehensive and integrated manner. Especially, integrated (computer) modelling by combining social and biophysical drivers, modelling of decision-making by agents, modelling of lag times, modelling thresholds, and multi-source data integration are promising methodologies (LUCC scientific steering committee, 2005). Under the umbrella of integrated approaches a multitude of approaches can be identified.

Many scholars describe the land use system as a so-called coupled human-environment system (e.g. Turner *et al*., 2004) and a number suggest treating this system as a complex system and adopting complexity theory in studying the system. Although complexity theory

itself is an ill-defined term (Manson, 2001) it includes a number of phenomena that are characteristic for land use change processes and which are studied to some extent already, for example self-organisation, emergence, path dependence and feedbacks (O'Sullivan, 2004). These phenomena are part of what Manson and O'Sullivan call aggregate complexity, which stems from a holistic and synergetic paradigm that deals with interactions of a variety of system components. These features mean that complex systems have often very different characteristics than systems that are in equilibrium, which is a basic assumption in many other economic and ecological theories.
Some scholars argue that 'land change science' is currently emerging as a new science (Turner *et al*., 2004; Rindfuss *et al*., 2004; Lambin *et al*., 2005). Lambin *et al*. (2005) argue that the time is ripe for an overarching theory: "Emerging sciences need their own theories". This call for an overarching theory is meant in a methodological sense rather than a call for new substantive theories as the theories of Von Thünen, Malthus and Boserup (Von Thünen, 1966; Malthus, 1967; Boserup, 1965). This overarching theory should incorporate issues such as behaviour of people, feedbacks, multiple levels, time and links with the broader world (Lambin *et al*., 2005). Although the authors also mention some of the difficulties and say that land use science is not yet able to produce such a theory there are some additional reservations to this call for an overarching theory. Apart from the question of whether it is possible to create an overarching theory, such a description has the risk of becoming incredibly complex, because all land use scientist have their own list of topics, processes and mechanisms that they would like to include in such a theory and these lists will never be the same for all scientists. Additionally, when such a large all-encompassing theory has to be implemented using computer models and practical tools one may be confronted with many difficulties regarding verification and validation of the system, data needs, computing capacity, etc. Furthermore, one all-encompassing theory may decrease the attention to search for alternative solutions. This may reduce the diversity of approaches in land use science, which is essential for integrative research as is concluded in this thesis.
In this thesis a framework for the analysis of environmental problems was adopted for the analysis of land use decisions. This so-called Action-in-Context framework (De Groot, 1992) is a promising tool in solving a large part of the land use jigsaw puzzle. Especially by combining the actors field (analysis of relation between primary, secondary, etc. actors) and the deeper analysis (in depth analysis of decisions of one actor or actor group), which was used in this thesis, the framework can incorporate a wide diversity of land use relevant issues. The deeper analysis was used throughout this study as a methodological framework for land use decisions. Furthermore, the multilevel approach (Chapter 4) is a very promising statistical method to bridge scales and levels and therefore to integrate disciplines. Although these methods link various parts of the land use system they do not provide the overarching methodological theory that includes all elements of the land use system.

With the above considerations in mind we would like to argue that theories of land use methodologies should focus on the combination of parts of the complex system at a level between disciplinary elements and an overarching theory. The combination of disciplinary elements of the system can lead to understanding of more and more sub-systems, which eventually proceeds in the direction of a larger theory. An all-encompassing theory of land use change is still far off. Theorising on these subsystems may provide smaller steps that contribute to the overarching theory and prevents a fixation on a final solution that would

draw attention away from alternatives that can provide important contributions. These theories of part of the system should aim at the true integration of some of the disciplines and methodological themes and should describe actual mechanisms and processes. Land use modelling has to focus on the development of mechanisms that enable integrative research rather that adding more elements to existing models. For example, integrating feedbacks, path dependency and emerging properties are not fully understood theoretically and neither is the potential of these issues sufficiently explored with modelling approaches. For example, integrated systems that include projections of the amount of land use change, allocation of land use, their effects and their feedbacks into the land claim and land use allocation are hardly (or not) available.
On the part of the substantive theories land use research would benefit from testing these theories in real world cases. In this thesis broad rational choice was tested by making the AiC framework operational in the study area, which proved to work well for the prediction of the occurrence of land use.

In this dissertation various methods were applied from different disciplinary perspectives. Both deductive, theoretical as well as inductive, statistical approaches were used. The joint understanding from these analyses enabled the integration of all important aspects in one modelling framework. The modelling was an important tool to organise the knowledge available about the complex system and added to the insight in the system as a whole. The combination of methods has been the key to improved understanding of the land use system.

References

Alonso, W. 1964. Location and Land Use. Toward a General Theory of Land Rent. Harvard University press, Cambridge, MA. 204 p.

Anselin, L. 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht. 284 p.

Atauri, J.A., de Lucio, J.V. 2001. The role of landscape structure in species richness distribution of birds, amphibians, reptiles and lepidopterans in Mediterranean landscapes. Landscape Ecology 16 (2), 147-159.

Bailey, S.A., Haines-Young, R.H., Watkins, C. 2002. Species presence in fragmented landscapes: modeling of species requirements at the national level. Biological Conservation 108, 307-316.

Balagtas-Mangabat, C. 2002. Philippines' forest-relevant land use policies: General overview and background in implementation in San Mariano, Philippines. Project working paper, CVPED, Cabagan, Isabela, Philippines.

Bebbington, A. 1999. Capitals and capabilities: A framework for analyzing peasant viability, rural livelihoods and poverty. World Development 27 (12), 2021-2044.

Bibby, C.J., Burgess, N.D., Hill, D.A. 1992. Bird Census Techniques. Academic Press, London, UK. 257 p.

Blaikie, P. 1985. The Political Economy of Soil Erosion in Developing Countries. Longman, Harlow, UK. 188 p.

Boserup, 1965. The Conditions of Agricultural Growth. Aldine, Chicago. 124 p.

Briassoulis, H. 2000. Analysis of land use change: Theoretical and modeling approaches. In: Loveridge, S. (ed.), The Web Book of Regional Science. West Virginia University, Morgantown.

Brooks, T.M., Mittermeier, R.A., Mittermeier C.G., Da Fonseca, G.A.B., Rylands, A.B., Konstant, W.R., Flick, P., Pilgrim, J., Oldfield, S., Magin, G., Hilton-Taylor, C. 2002. Habitat loss and extinction in the hotspots of biodiversity. Conservation Biology 16 (4), 909-923.

Brooks, T.M., Pimm, S.L., Collar, N.J. 1997. Deforestation predicts the number of threatened birds in insular Southeast Asia. Conservation Biology 11 (2), 382-394.

Brown, D.G., Walker, R., Manson, S., Seto, K. 2004. Modeling land use and land cover change. In: Gutman, G., Janetos, A.C., Justice, C.O., Moran, E.F., Mustard, J.F., Rindfuss, R.R., Skole, D.L., Turner, B.L., Cochrane, M.A. (eds.), Land Change Science: Observing, Monitoring and Understanding Trajectories of Change on the Earth's Surface. Kluwer Academic Publishers, Dordrecht/Boston/London. pp. 395-409.

Brown, D.G., Page, S., Riolo, R., Zellner, M., Rand, W. 2005. Path dependence and the validation of agent-based spatial models of land use. International Journal of Geographical Information Science 19 (2), 153-174.

Browne, W.J., Subramanian, S.V., Jones, K., Goldstein, H. 2005. Variance partitioning in multilevel logistic models that exhibit over-dispersion. Journal of Royal Statistical Society A 168, 599-614.

Brox, O. 1990. The common property theory – epistemological status and analytical utility. Human Organization 49 (3), 227-235.

Bruijnzeel, L.A. 2004. Hydrological functions of tropical forests: not seeing the soil for the trees? Agriculture, Ecosystems and Environment 104, 185-228.

Buckland, S.T., Elston, D.A. 1993. Empirical models for the spatial distribution of wildlife. Journal of Applied Ecology 30, 478-495.
Bunnell, F.L., Huggard, D.J. 1999. Biodiversity across spatial and temporal scales: problems and opportunities. Forest Ecology and Management 115, 113-126.
Centre for multilevel modelling, 2005. Software reviews of multilevel analysis packages http://www.mlwin.com/softrev/index.html (October 29, 2005).
Chomitz K.M., Gray D.A. 1996. Roads, land use, and deforestation: A spatial model applied to Belize. World Bank Economic Review 10 (3), 487-512.
Chomitz, K.M. Kumari, K. 1998. The domestic benefits of tropical forests. The World Bank Observer 13 (1), 13-35.
Chomitz K.M., Thomas T.S. 2003. Determinants of land use in Amazonia: A fine-scale spatial analysis. American Journal of Agricultural Economics 85 (4), 1016-1028.
Cleuren, H.M.R.H. 2001. Paving the Road for Forest Destruction. Key Actors and Driving Forces of Tropical Deforestation in Brazil, Ecuador and Cameroon. Leiden Development Studies, New Series, No. 1, Leiden University, Leiden, The Netherlands.
Coleman, J.S. 1990. Foundations of Social Theory. Harvard University Press, Boston. 993 p.
Costanza, R. 1989. Model goodness of fit: A multiple resolution procedure. Ecological Modelling 47, 199-215.
Couclelis, H. 2001. Modeling frameworks, paradigms, and approaches. In: Clark, K.C., Parks, B.E., Crane, M.P. (eds.), Geographic Information Systems and Environmental Modeling. Longman & Co, New York. pp. 34-48.
Couto, P. 2003. Assessing the accuracy of spatial simulation models. Ecological Modeling 167, 181-198.
Coxhead, I. 2000. Consequences of a food security strategy for economic welfare, income distribution and land degradation: The Philippine case. World Development 28 (1), 111-128.
Coxhead, I., Buenavista, G. 2001. Seeking sustainability: A synthesis of research in phase I of SANREM CRSP-Southeast Asia, 1993-98. In: Coxhead, I., Buenavista, G. (eds.), Seeking Sustainability: Challenges of Natural Resource Management in a Philippine Watershed. Philippines: Phil. Council on Agricultural Research, Natural Resources and Development, Los Baños. pp. 1-31.
Crawley, M.J. 1993. GLIM for Ecologists. Blackwell Science, Oxford. 379 p.
Dale, V.H. 1997. The relationship between land-use change and climate change. Ecological Applications 7 (3), 753-769.
Dauber, J., Hirsch, M., Simmering, D., Waldhardt, R., Otte, A., Wolters, V. 2003. Landscape structure as an indicator of biodiversity: matrix effects on species richness. Agriculture, Ecosystems and Environment 98 (1-3), 321-329.
De Groot, W.T. 1992. Environmental Science Theory; Concepts and Methods in a One-World, Problem Oriented Paradigm. Elsevier Science Publishers, Amsterdam. 583 p.
De Groot, W.T. 1999. Van vriend naar vijand naar verslagene en verder. Een evolutionair perspectief op de verhouding tussen ment en natuur. Nijmegen University Press, Nijmegen, The Netherlands. 50 p.
De Groot, W.T. 2003. Thunian land use theory, visions of nature, and the future of the Sierra Madre forest. In: Ploeg, J. van der, Bernardo, E.C., Masipiquena, A.B. (eds.), The Sierra Madre Mountain Range: Global Relevance, Local Realities. Golden Press, Tuguegarao City, Philippines.

De Groot, W.T., De Jong, F.M.W., Van den Berg, M.M.H.E. 1987. Population dynamics of duckweed in polder ditches. Hydrobiological Archives 109 (4), 601-618.

De Groot, W.T., Kamminga, E.M. 1995. Forest, People, Government: a policy-oriented analysis of the social dynamics of tropical deforestation. Dutch National Programme on Global Air Pollution and Climate Change, Bilthoven, the Netherlands. 183 p.

DENR-CENRO (Department of Environment and Natural Resources, Community Environment and Natural Resources Office) 1998. Operations map 1:50000, CENRO, Naguilan, Isabela, Philippines.

Doorman, F. 1991. A framework for the rapid appraisal of factors that influence the adoption and impact of new agricultural technology. Human Organization 50 (3), 235-244.

Easterling, W.E. 1997. Why regional studies are needed in the development of full-scale integrated assessment modelling of global change processes. Global Environmental Change Part A 7 (4), 337-356.

Elster, J. 1989. Nuts and Bolts for the Social Sciences. Cambridge University Press, Cambridge. 184 p.

Eppink, F.V., van den Bergh, J.C.J.M., Rietveld, P. 2004. Modelling biodiversity and land use: urban growth, agriculture and nature in a wetland area. Ecological Economics 51 (3-4), 201-216.

ESSC (Environmental Science for Social Change, Inc.) 1999. Decline of the Philippine Forest. The Bookmark, Inc., Makati City, Philippines.

Fairbanks, D.H.K. 2004. Regional land-use impacts affecting avian richness patterns in Southern Africa-insights from historical avian atlas data. Agriculture, Ecosystems and Environment 101 (2-3), 269-288.

Fazey. I., Fischer, J., Lindenmayer, D.B. 2005. What do conservation biologists publish? Biological Conservation 124 (1), 63-73.

Fox, J., Rindfuss, R.R., Walsh, S.J. Mishra, V. (eds.) 2003. People and the Environment: Approaches for Linking Household and Community Surveys to Remote Sensing and GIS. Kluwer Academic Publishers, Boston/Dordrecht/London. 344 p.

Garrity D.P., Kummer, D.M., Guiang, E.S. 1993. The Philippines. In: National Research Council Committee on sustainable agriculture and the environment in the humid tropics (ed.), Sustainable Agriculture and the Environment in the Humid Tropics. National Academy Press, Washington, DC. pp. 549–624.

Geist, H.J., Lambin, E.F. 2002. Proximate causes and underlying driving forces of tropical deforestation. BioScience 52 (2), 143-150.

General, A.A. 1999. Co-management of the environment through the Socialized Integrated Industrial Forest Management Agreement (SIFMA): The San Mariano experience. In: Bernardo, E.C., Snelder, D.J. (eds.), Co-managing the Environment; The Natural Resources of the Sierra Madre Mountain Range. Proceedings International Work Conference, 21-24 September 1998, CVPED/Plan International, Cabagan, Philippines.

Geoghegan, J., Pritchard Jr, L., Ogneva-Himmelberger, Y., Chowdhury, R.R., Sanderson, S., Turner II, B.L. 1998. "Socializing the pixel" and "pixelizing the social" in land-use and land-cover change. In: Liverman, D., Moran, E., Rindfuss, R.R., Stern, P. (eds.), People and Pixels: Linking Remote Sensing and Social Science. Committee on the Human Dimensions of Global Change, National Academy Press, Washington, D.C. 244 p.

Geoghegan, J., Cortina Villar, S., Klepeis, P., Macario Mendoza, P., Ogneva-Himmelberger, Y., Roy Chowdhury, R., Turner II, B.L., Vance, C. 2001. Modeling tropical deforestation in the Southern Yucatán Peninsular Region: Comparing survey and satellite data. Agriculture, Ecosystems and Environment 85 (1-3), 25-46.

Ghazoul, J., Hellier, A. 2000. Setting critical limits to ecological indicators of sustainable tropical forestry. International Forestry Review 2 (4), 243-253.

Gibson, C.C., Ostrom, E., Ahn, T.K. 2000. The concept of scale and the human dimensions of global change: a survey. Ecological Economics 32 (2), 217-239.

Gibson, L.A., Wilson, B.A., Aberton, J.G. 2004. Landscape characteristics associated with species richness and occurrence of small native mammals inhabiting a coastal heathland: a spatial modelling approach. Biological Conservation 120 (1), 75-89.

Gilligan, C. 1982. In a Different Voice: Psychological Theory and Women's Development. Harvard University Press, Cambridge.

Glaser, B., Strauss, A. 1967. The Discovery of Grounded Theory. Aldine, Chicago. 271 p.

GLP (Global Land Project). 2005. Science Plan and Implementation Strategy. IGBP Report No. 53/IHDP Report No. 19. IGBP Secretariat, Stockholm. 64 p.

Goldstein, H. 1995. Multi-Level Statistical Models. Kendall's library of statistics, Vol 3. Edward Arnold, London.

Guisan, A., Zimmerman, N.E. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135, 147-186.

Hargis, C.D., Bissonette, J.A., David, J.L. 1998. The behavior of landscape metrics commonly used in the study of habitat fragmentation. Landscape Ecology 13, 167-186.

Haynes, K.E., Fotheringham, A.S. 1984. Gravity and Spatial Interaction Models. Sage publications, Beverly Hills, Ca. 88 p.

Henle, K., Lindenmayer, D.B., Margules, C.R., Saunders, D.A., Wissel, C. 2004. Species survival in fragmented landscapes. Biodiversity and Conservation 13 (1), 1-8.

Hobbes, M., Stalpers, S.I.P., Kooijman, J., Le Thi Thu Thanh, Trinh Khanh Chi, Phan Thi Anh Dao. (Not dated). MFA in social context: A Vietnamese case study of combining the Material Flow Analysis and Action-in-Context frameworks. Journal of Industrial Ecology (accepted).

Hoshino, S. 2001. Multilevel modeling on farmland distribution in Japan. Land Use Policy 18 (1), 75-90.

Http://www.landsat.org.: Landsat image path 116, row 48, date: June 3, 2001. Source: Landsat.org, Basic Science and Remote Sensing Initiative, Michigan State University (http://www.landsat.org).

Huigen, M.G.A. 2004. First principles of the MameLuke multi-actor modelling framework for land use change, illustrated with a Philippine case study. Journal of Environmental Management 72 (1-2), 5-21.

Huigen, M.G.A., Jens, I. (Not dated). Socio-economic impact of super typhoon Harurot in San Mariano, Isabela, the Philippines. World Development (submitted).

Irwin, E.G. Geoghegan, J. 2001. Theory, data, methods: developing spatially explicit economic models of land use change. Agriculture, Ecosystems and Environment 85 (1-3), 7-24.

IUCN. 2005. 2005 IUCN Red List of Threatened Species. www.iucn.redlist.org. Downloaded on 25 August 2005.

IHDP. 2005. IHDP Update. Newsletter of the international human dimensions programme on global environmental change. Nr. 3/2005 "Land-Use and Land-Cover Change". IHDP, Bonn, Germany.

Jepsen, J.U., Topping, C.J. Odderskær, P., Andersen, P.N. 2005. Evaluating consequences of land-use strategies on wildlife populations using multiple-species predictive scenarios. Agriculture, Ecosystems and Environment 105 (4). 581-594.

Johns, A.D. 1996. Bird population persistence in Sabahan logging concessions. Biological Conservation 75, 3-10.

Jones, K., Duncan, C. 1995. Individuals and their ecologies: analysing the geography of chronic illness within a multilevel modelling framework. Health and Place 1 (1), 27-40.

Jongman, M. 1997. The Discours of Logging Bans. A Case Study of the Logging Moratotium in San Mariano, Isabela, The Philippines. MA-thesis, CML student report No 76. Leiden, Netherlands, and Cabagan, Isabela, Philippines.

Kennedy, R.S., Gonzales, P.C., Dickinson, E.C., Miranda, Jr., H.C. Fisher, T.H. 2000. A Guide to the Birds of the Philippines. Oxford University Press, Oxford.

Klein, M. 2003. Introducing Fruit-Trees at Smallholder's Farmers : Potential and Constraints. (Environment and development student report, nr. 160). Leiden; Cabagan: Leiden University, Centre of Environmental Science; Isabela State University, Cabagan Valley Programme on Environment and Development (CVPED).

Kok, K., Farrow, A., Veldkamp A., Verburg, P.H. 2001. A method and application of multi-scale validation in spatial land use models Agriculture, Ecosystems and Environment 85 (1-3), 223-238.

Kok, K., Veldkamp A. 2000. Using the CLUE framework to model changes in land use on multiple scales. In: Bouman, B.A.M., Jansen, H.G.P., Schipper, R.A., Hengsdijk, H. Nieuwenhuyse, A. (eds.), Tools for Land Use Analysis on Different Scales with Case Studies for Costa Rica. Kluwer, Dordrecht. pp. 35-63.

Kok, K., Winograd, M. 2002. Modelling land use change for Central-America, with special reference to the impact of Hurricane Mitch. Ecological Modelling 149, 53-69.

Kummer, D.M. 1992. Deforestation in the Post-War Philippines. Ateneo de Manila University Press, Manila; University of Chicago Press, Chicago. 177 p.

Lambin, E.F., Baulies, X., Bockstael, N., Fischer, G., Krug, T., Leemans, R., Moran, E.F., Rindfuss, R.R., Skole, D., Turner II, B.L., Vogel, C. 1999. Land-Use and Land-Cover Change: Implementation Strategy, IGBP Report No. 48/IHDP Report No. 10, IGBP, Stockholm. 125 p.

Lambin E.F., Turner II, B.L., Geista, H.J., Agbola, S.B., Angelsen, A., Brucee, J.W., Coomes, OT., Dirzo, R., Fischer, G., Folke, C., George, P.S., Homewood, K., Imbernon, J., Leemans, R., Xiubin Li, Moran, E.F., Mortimore, M., Ramakrisnan, P.S., Richards, J.F., Skanes, H., Steffen, W., Stone, G.D. Svedin, U., Veldkamp, A. Vogel, C., Jianchu Xu. 2001. The causes of land-use and land-cover change: moving beyond the myths. Global Environmental Change 11 (4), 261-269.

Lambin E.F., Ehrlich D. 1997. Land-cover Changes in Sub-Saharan Africa (1982-1991): Application of a change index based on remotely sensed surface temperature and vegetation indices at a continental scale. Remote Sensing of Environment 61 (2), 181-200.

Lambin, E.F., Geist, H.J., Lepers, E. 2003. Dynamics of land-use and land-cover change in tropical regions. Annual Review of Environment and Resources 28, 205-241.

Lambin, E.F., Geist H., Rindfuss, R.R. 2005. Land-use and land-cover change: developing and implementing an agenda for local processes with global impacts. In: IHDP, IHDP Update. Newsletter of the international human dimensions programme on global environmental change. Nr. 3/2005 "Land-Use and Land-Cover Change". IHDP, Bonn, Germany.

Laney, R.M. 2004. A process-led approach to modeling land change in agricultural landscapes: a case study from Madagascar. Agriculture, Ecosystems and Environment 101 (2-3), 135-153.

Liao, S. 2003. Knowledge management technologies and applications — literature review from 1995 to 2002. Expert Systems with Applications 25, 155-164.
Liverman, D., Moran, E., Rindfuss, R.R., Stern, P. (eds.) 1998. People and pixels: Linking Remote Sensing and Social Science, Committee on the Human Dimensions of Global Change National Academy Press, Washington D.C. 244 p.
Lobo, J.M., Martín-Piera, F. 2002. Searching for a predictive model for species richness of Iberian dung beetle based on spatial and environmental variables. Conservation Biology 16 (1), 158-173.
LUCC scientific steering committee. 2005. Frontiers in Land-use/cover change research. In: IHDP, IHDP Update. Newsletter of the international human dimensions programme on global environmental change. Nr. 3/2005 "Land-Use and Land-Cover Change". IHDP, Bonn, Germany.
Luoto, M., Virkkala, R., Heikkinen, R.K., Rainio, K. 2004. Predicting bird species richness using remote sensing in boreal agricultural-forest mosaics. Ecological Applications 14 (6), 1946-1962.
Mac Nally, R., Fleishman, E., Fay, J.P., Murphy, D.D. 2003. Modelling butterfly species richness using mesoscale environmental variables: model construction and validation for mountain ranges in the Great Basin of western North America. Biological Conservation 110 (1), 21-31.
Malthus, R., 1967 [1789]. Essay on the Principles of Population. 7th edition, Dent, London.
Manson, S.M. 2001. Simplifying complexity: a review of complexity theory. Geoforum 32 (3), 405-414.
McConnell, W. 2002. Preface. In: Parker, D.C., Berger, T., Manson, S.M. (eds.), Agent-Based Models of Land-Use and Land-Cover Change; Report and Review of an International Workshop, October 4-7, 2001. Indiana University, LUCC Report Series No. 6.
McConnell, W., Moran, E.F. 2001. Meeting in the Middle: The Challenge of Meso-Level Integration. An International Workshop on the Harmonization of Land Use and Land Cover Classification. LUCC Report Series No. 5. Indiana University, Bloomington, USA.
Menard, S. 2001. Applied Logistic Regression Analysis. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-106, Sage, Thousand Oaks, CA.
Menon, S., Pontius, R.G., Rose, J., Khan, M.L., Bawa, K.S. 2001. Identifying conservation-priority areas in the tropics: a land-use change modeling approach. Conservation Biology 15 (2), 501-512.
Mertens, B., Sunderlin, W.D., Ndoye, O., Lambin, E.F. 2000. Impact of macroeconomic change on deforestation in South Cameroon: Integration of household survey and remotely-sensed data. World Development 28 (6), 983-99.
Moonen, E. 2002. Improving Your Land…: From 'Swidden' to 'Rice-Paddy' in the Forest-Frontier of the Sierra Madre Mountains of Luzon, the Philippines. Centre of Environmental Science (CML), Leiden, The Netherlands.
Moran, E.F., Skole, D.L., Turner II, B.L. 2004. The development of the international land use and land cover change (LUCC) research program and its links to NASA's land cover and land use change (LCLUC) initiative. In: Gutman, G.; Janetos, A.C.; Justice, C.O.; Moran, E.F.; Mustard, J.F.; Rindfuss, R.R.; Skole, D.; Turner II, B.L.; Cochrane, M.A. (eds.), Land Change Science: Observing, Monitoring and Understanding Trajectories of Change on the Earth's Surface. Series: Remote Sensing and Digital Image Processing, Vol. 6. Kluwer, Dordrecht. 461 p.

Müller, D., Zeller, M. 2002. Land use dynamics in the central highlands of Vietnam: a spatial model combining village survey data with satellite imagery interpretation. Agricultural Economics 27 (3), 333-354.

Municipality of San Mariano and Housing and Land Use Regulatory Board. 2000. Comprehensive Development Plan San Mariano, Isabela (planning period 2000-2004).

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J. 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858.

NAMRIA. Not dated. Sheet 7476 IV San Mariano 1:50000, NAMRIA: National Mapping and Resource Information Authority, Department of Environment and Natural Resources, Fort Bonifacio, Makati, Manila, Philippines.

Nelson, G. C. 2002. Introduction to the special issue on spatial analysis for agricultural economists. Agricultural Economics 27 (3), 197-200.

Nelson, G.C., De Pinto, A., Harris, V., Stone, S. 2004. Land use and road improvements: A spatial perspective. International Regional Science Review 27 (3), 297-325.

Nelson, G.C., Harris, V., Stone, S.W. 2001. Deforestation, land use and property rights: empirical evidence from Darien, Panama. Land Economics 77 (2), 187-205.

Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W. 1996. Applied Linear Statistical Models, 4th edition. Irwin, Burr Ridge, Illinois. 1408 p.

Osborne, P.E., Alonso, J.C., Bryant, R.G. 2001. Modelling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. Journal of Applied Ecology 38, 458–471.

Ostrom, E. 1990. Governing the Commons: the Evolution of Institutions for Collective Action. Cambridge University Press, Cambridge.

O'Sullivan, D. 2004. Complexity science and human geography. Transactions of the Institute of British Geographers, 29 (1) 282-295.

Overmars, K.P., De Groot, W.T., Huigen, M.G.A. 2006. Comparing Inductive and Deductive Modeling of Land Use Decisions: Principles, a Model and an Illustration from the Philippines. Human Ecology (accepted).

Overmars, K.P., De Koning, G. H. J., Veldkamp, A. 2003. Spatial autocorrelation in multiscale land use models. Ecological Modelling 164 (2-3), 257-270.

Overmars, K.P., Verburg, P.H. 2005. Analysis of land use drivers at the watershed and household level: Linking two paradigms at the Philippine forest fringe. International Journal of Geographical Information Science 19 (2), 125-152.

Overmars, K.P., Verburg, P.H. 2006. Comparison of an actor-based and an empirical approach to specify a spatially explicit land use model. Land Use Policy (accepted).

Pan, W., Bilsborrow, R.E. 2005. The use of a multilevel statistical model to analyze factors influencing land use: A study of the Ecuadorian Amazon. Global and Planetary Change 47 (2-4), 232-252.

Parker, D.C., Berger, T., Manson, S.M. (eds.) 2002. Agent-Based Models of Land-Use and Land-Cover Change. LUCC Report Series No. 6. LUCC Focus I Office, Indiana University. 124 p.

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

Persoon, G.A., Van der Ploeg, J. 2003. Reviewing the projected future of San Mariano, a boomtown at the Sierra Madre forest fringe. Philippines Studies 51 (3), 451-473.

Perz, S.G., Walker, R.T. 2002. Household life cycles and secondary forest cover among small farm colonists in the amazon. World Development 30 (6), 1009-1027.

Perz S.G., Skole D.L. 2003. Social determinants of secondary forests in the Brazilian Amazon. Social Science Research 32 (1), 25-60.

Pickett, S.T.A., Burch jr., W.R., Morgan, J.M. 1999. Interdisciplinary research: Maintaining the constructive impulse in a culture of criticism. Ecosystems 2, 302-307.

Pijanowski, B.C., Brown, D. G., Manik, G., Shellito, B. 2002. Using artificial neural networks and GIS to forecast land use changes: A land transformation model. Computers, Environment and Urban Systems 26 (6), 553-575.

Polsky, C., 2004. Putting space and time in ricardian climate change impact studies: Agriculture in the U.S. Great Plains, 1969 1992. Annals of the Association of American Geographers 94 (3), 549-564.

Polsky, C., Easterling III, W.E. 2001. Adaptation to climate variability and change in the U.S. Great Plains: A multi-scale analysis of Ricardian climate sensitivities. Agriculture, Ecosystems and Environment 85 (1-3), 133-144.

Pontius Jr., R.G., Boersma, W., Castella, J-C., Clarke, K., De Nijs, T., Dietzel, C., Zengqiang Duan, Fotsing, E., Goldstein, N., Kok, K., Koomen, E., Lippitt, C.D., McConnell, W., Pijanowski, B., Pithadia, S., Mohd Sood, A., Sweeney, S., Tran Ngoc Trung, Verburg, P.H. 2005. Cross case comparison for several prominent land change models. Presentation at the 6th Open Meeting of the Human Dimensions of Global Environmental Change Research Community, Bonn 9-13 October 2005. Downloaded from: http://www.clarku.edu/~rpontius/ on 30-1-2006.

Pontius Jr., R.G., Cornell, J.D., Hall, C.A.S. 2001. Modelling the spatial pattern of land use change with GEOMOD2: Application and validation for Costa Rica. Agriculture, Ecosystems and Environment 85 (1-3), 191-203.

Pontius, Jr., R.G., Schneider, L.C. 2001. Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems and Environment 85 (1-3), 239-248.

Popper, K.R. 1963. Conjectures and refutations: the growth of scientific knowledge. Routledge and Kegan Paul, London.

Rasbash, J., Browne, W., Goldstein., H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., Lewis, T. 2000. A User's Guide to MLwiN (Second Edition), Institute of Education, London. 278 p.

Raudenbush, S., Bryk, A., Congdon K. 2004. HLM 6, Hierarchical Linear and Non-Linear Modelling. SSI, Scientific Software international.

Rindfuss, R.R., Walsh, S., Mishra, V., Fox, J., Dolcemascolo, G.P. 2003. Linking household and remotely sensed data. In: Fox, J., Rindfuss, R.R., Walsh, S.J., Mishra, V. (eds.), People and the Environment: Approaches for Linking Household and Community Surveys to Remote Sensing and GIS. Kluwer Academic Publishers, Boston. pp. 1-29.

Rindfuss, R.R., Walsh, S.J., Turner II, B.L., Fox, J., Mishra, V. 2004. Developing a science of land change: Challenges and methodological issues. PNAS 101 (39), 13976–13981.

Robinson, W. 1950. Ecological correlations and the behaviour of individuals'. American Sociological Review (15), 351 – 357.

Rotmans, J., van Asselt, M.B.A., Anastasi, C., Greeuw, S.C.H., Mellors, J., Peters, S., Rothman, D., Rijkens, N. 2000. Visions for a sustainable Europe. Futures 32 (9-10), 809 - 831.

Rudel, T., Roper, J. 1997. The paths to rain forest destruction: Crossnational patterns of tropical deforestation, 1975-90. World Development 25 (1), 53-65.

Schneider, L.C., Pontius Jr, R.G. 2001. Modeling land use change in the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems and Environment 85 (1-3), 83-94.

Schoenberger, E. 2001. Interdisciplinarity and social power. Progress in Human Geography 25 (3), 365-382.

Schwanen T., Dieleman F.M., Dijst M. 2004. The impact of metropolitan structure on commute behavior in the Netherlands: A multilevel approach. Growth and Change 35 (3), 304-333.

Serneels, S., Lambin, E.F. 2001. Proximate causes of land-use change in Narok District, Kenya: a spatial statistical model. Agriculture, Ecosystems and Environment 85 (1-3), 65-81.

Snelder, D.J. 2001. Forest patches in Imperata grassland and prospects for their preservation under agricultural. Agroforestry Systems 52, 207-217.

Snijders, T.A.B., Bosker, R.J. 1999. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling. Sage, New York. 266 p.

Sodhi, N.S., Liow, L.H., Bazzaz, F.A. 2004. Avian extinctions from tropical and subtropical forests. Annual Review of Ecology Evolution and Systematics 35, 323-345.

Solecki, W.D., Oliveri, C. 2004. Downscaling climate change scenarios in an urban land use change model. Journal of Environmental Management 72 (1-2), 105-115.

Staal, S.J., Baltenweck, I., Waithaka, M.M., deWolff T., Njoroge, L. 2002. Location and uptake: integrated household and GIS analysis of technology adoption and land use, with application to smallholder dairy farms in Kenya. Agricultural Economics 27 (3), 295-315.

Sterken, J. 2004. Banana-based farming systems in relation to soil quality in San Mariano municipality, Isabela Province, the Philippines. Environment and development student report, nr. 172. Leiden; Cabagan : Leiden University. Institute of Environmental Sciences; Isabela State University. Cabagan Valley Programme on Environment and Development (CVPED).

Stolle, F., Chomitz, K.M., Lambin, E.F., Tomich, T.P. 2003. Land use and forest fires in Jambi province, Sumatra, Indonesia. Forest, Ecology and Management 179 (1-3), 277-292.

Steiner, N.C., Köhler, W. 2003. Effects of landscape patterns on species richness—a modelling approach. Agriculture, Ecosystems and Environment 98 (1-3), 353-361.

Store, R., Jokimäki, J. 2003. A GIS-based multi-scale approach to habitat suitability modelling. Ecological Modelling 169, 1-15.

Swets, J.A. 1988. Measuring the accuracy of diagnostic systems. Science 240, 1285-1293.

Tadepally, H. 1999. Neglect and restoration of minor irrigation tanks in Krishna district, India : a local-level case study of motivation, capacity and partnership interventions. Maruthibala Publishers, Nuzvid.

Thijs, W. 2005. The Distribution of Bird and Bat Species in Remnant Forest Patches in Northeast Luzon, Philippines, in Relation to Landscape Characteristics. Student report 186, CML Leiden University and Dept. of Environmental Sciences, Wageningen University.

Tilman, D., May, R.M., Lehman, C.L., Novak, M.A. 1994. Habitat destruction and the extinction debt. Nature 371, 65 -66.

Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography 46, 234–40.

Tobler, W.R. 1979. Cellular geography. In: Gale, S., Olssen, G. (eds.), Philosophy in Geography. Reidel, Dordrecht. pp. 379-386.

Turner II, B.L., Moran, E.F., Rindfuss, R.R. 2004. Integrated Land-Change Science and Its Relevance to the Human Sciences In: Gutman, G., Janetos, A.C., Justice, C.O., Moran, E.F., Mustard, J.F., Rindfuss, R.R., Skole, D., Turner II, B.L., Cochrane, M.A. (eds.), Land Change Science: Observing, Monitoring and Understanding Trajectories of Change on the Earth's Surface. Series: Remote Sensing and Digital Image Processing, Vol. 6. Kluwer, Dordrecht. pp 431-447.

Turner II, B.L., Skole, D, Sanderson, S., Fischer, G., Fresco L., Leemans, R. 1995. Land Use and Land-Cover Change Science/Research Plan. IGPB Report No. 35 and HDP Report No. 7. 132 p.

Turner, I.M. 1996. Species loss in fragments of tropical rain forest: a review of the evidence. Journal of Applied Ecology 33, 200-209.

Turner, M.D. 2003. Methodological reflections on the use of remote sensing and geographic information science in human ecological research. Human Ecology 31 (2), 255-279.

Valmayor, R.V., Jones, D.R., Subijanto, Polprasid, P., Jamaluddin, S.H. 1990. Bananas and Plantains in Southeast Asia, INIBAP, Montpellier.

Vance, C., Geoghegan, J. 2002. Temporal and spatial modelling of tropical deforestation: a survival analysis satellite and household survey data. Agricultural Economics 27, 317-332.

Van den Top, G.M. 1998. The Social Dynamics of Deforestation in the Sierra Madre, Philippines. PhD. Thesis, Centre of Environmental Science, Leiden University. 437 p.

Van der Ploeg, J., Van Weerd, M. 2004. Devolution of natural resource management and Philippine crocodile conservation in San Mariano, Isabela. Philippine Studies 52 (3), 345-382.

Van Egmond, F. 2003. Land Use, Geomorphology and Soil properties in the San Mariano Watershed, Philippines. MSc. Thesis, Wageningen University, The Netherlands, student report, Leiden University, The Netherlands.

Van Weerd, M. 2002. Summary Report of the Fauna Study Results in the Northern Sierra Madre Natural Park by the Plan – Philippines / Northern Sierra Madre Natural Park – Conservation Project. NSMNP-CP, Cabagan, Philippines.

Van Weerd, M., Van Boven, G., Van der Ploeg, J. 2004. Using communication to involve local communities and local governments in protected area management: two related cases from Northeastern Luzon, Philippines. In: Hamú, D., Auchincloss, E., Goldstein, W. (eds.), Communicating Protected Areas. Commission on Education and Communication, IUCN, Gland, Switzerland and Cambridge, UK. pp. 139-154.

Vayda, A.P. 1983. Progressive contextualization: methods for research in human ecology. Human Ecology 11 (3), 265-281.

Vayda, A.P. Walters, B.B. 1999. Against political ecology. Human Ecology 27 (1), 167-179.

Veldkamp, A., Fresco, L.O. 1997. Reconstructing land use drivers and their spatial dependence for Costa Rica (1973 and 1984). Agricultural Systems 55, 19-43.

Veldkamp, A., Lambin, E.F. 2001. Predicting land-use change. Agriculture, Ecosystems and Environment 85 (1-3), 1-6.

Veldkamp, A., Kok, K., De Koning, G.H.J., Schoorl, J.M., Sonneveld, M.P.W., Verburg, P.H. 2001. Multi-scale system approaches in agronomic research at the landscape level. Soil and Tillage Research 58 (3-4), 129-140.

Veldkamp A., Verburg, P.H. 2004. Modelling land use change and environmental impact. Journal of Environmental Management 72 (1-2), 1-3.

Verbeek, M. 2000. A Guide to Modern Econometrics, John Wiley and Sons. 386 p.

Verburg, P.H., Chen, Y.Q. 2000. Multiscale characterization of land-use patterns in China. Ecosystems 3, 369-385.

Verburg, P.H., De Groot, W.T., Veldkamp, A. 2003. Methodology for multi-scale land-use modelling: concepts and challenges. In: Dolman, A.J., Verhagen, A., Rovers, C.A. (eds.), Global Environmental Change and Land Use. Kluwer Academic Publishers, Dordrecht/Boston/London.

Verburg, P.H., Overmars, K.P., Witte, N. 2004a. Accessibility and land use patterns at the forest fringe in the Northeastern part of the Philippines. The Geographical Journal 170 (3), 238-255.

Verburg, P.H., Ritsema van Eck, J., de Nijs, T.C.M., Dijst, M.J., Schot, P. 2004b. Determinants of land use change patterns in the Netherlands. Environment and Planning B 31, 125-150.

Verburg, P.H., Ritsema van Eck, J., de Nijs, T.C.M., Visser, H., de Jong, K. 2004c. A method to analyse neighbourhood characteristics of land use patterns. Computers, Environment and Urban Systems 28 (6), 667-690.

Verburg, P.H., Schot, P., Dijst, M., Veldkamp, A. 2004d. Land use change modelling: current practice and research priorities. GeoJournal 61 (4), 309-324.

Verburg, P.H., Soepboer, W., Veldkamp, A., Limpiada, R., Espaldon, V., Sharifah Mastura S.A. 2002. Modeling the spatial dynamics of regional land use: the CLUE-S Model. Environmental Management 30 (3), 391-405.

Verburg, P.H., Veldkamp, A. 2004. Projecting land use transitions at forest fringes in the Philippines at two spatial scales. Landscape Ecology 19 (1), 77-98.

Vitousek, P.M., Mooney, H.A., Lubchenco, J., Melillo, J.M. 1997. Human domination of earth's ecosystems. Science 277, 494-499.

Von Thünen, J.H. 1966. Der isolierte staat in beziehung der landwirtschaft und nationalökonomie. In: Hall, P. (ed.), Von Thünen's Isolated State. Pergamon press, Oxford.

Walker, R. 2004. Theorizing land-cover and land-use change: The case of tropical deforestation. International Regional Science Review 27 (3), 247-270.

Walker, R., Perz, S., Caldas, M., Silva, L.G.T. 2002. Land use and land cover change in forest frontiers: The role of household life cycles. International Regional Science Review 25 (2), 169-199.

Walker, R., Moran, E., Anselin, L. 2000 Deforestation and cattle ranching in the Brazilian Amazon: External capital and household processes. World Development 28 (4), 683-699.

Walker, R., Solecki, W. 2004. Theorizing land-cover and land-use change: The case of the Florida Everglades and its degradation. Annals of the Association of American Geographers 94 (2), 311-328.

Walsh, S.J., Bilsborrow, R.E., McGregor, S.J., Frizzelle, B.G., Messina, J.P., Pan, W.K.T., Crews-Meyer, K.A., Taff, G.N., Baquero, F.D. 2003. Integration of longitudinal surveys, remote sensing time-series, and spatial analyses: approaches for linking people and place. In: Fox, J., Rindfuss, R.R., Walsh, S.J., Mishra, V., (eds.), People and the Environment: Approaches for Linking Household and Community Surveys to Remote Sensing and GIS. Kluwer Academic Publishers, Boston/Dordrecht/London. pp. 91-130.

Walsh, S.J., Crawford, T.W., Welsh, W.F., Crews-Meyer, K.A. 2001. A multiscale analysis of LULC and NDVI variation in Nang Rong district, northeast Thailand. Agriculture, Ecosystems and Environment 85 (1-3), 47-64.

Walsh, S.J., Crews-Meyer, K.A. (eds.) 2002. Linking People, Place, and Policy. A GIS Science Approach. Kluwer Academic Publishers, Dordrecht. 348 p.

Walsh S.J., Evans T.P., Welsh W.F., Entwisle B., Rindfuss R.R. 1999. Scale dependent relationships between population and environment in Northeastern Thailand. Photogrammetric Engineering and Remote Sensing 65, 97–105.

WBCP. 2004. Checklist of Birds of the Philippines. Jensen, A., Pryor, S., Allen, D. (eds.), Wild Bird Club of the Philippines (WBCP), Manila.

White, R., Engelen, G., Uljee, I. 1997. The use of constrained cellular automata for high-resolution modelling of urban land-use dynamics. Environment and Planning B 24, 323–343.

Witte N. 2003. Accessibility as a Driver for Land Use Change. MSc. Thesis, Wageningen University, The Netherlands, student report, Leiden University, The Netherlands.

Zuiderwijk, A.B. 1998. Farming Gently, Farming Fast. Migration, Incorporation and Agricultural Change in the Mandara Mountains of Northern Cameroon. PhD dissertation, Leiden University, Leiden, The Netherlands.

Summary

Overmars, K.P. 2006. Linking process and pattern of land use change, illustrated with a case study in San Mariano, Isabela, Philippines. Thesis Leiden University, the Netherlands.

The conversion of the earth's land surface by human actions has been extensive in the past and is still on-going at a fast rate. Land use change does not affect all regions in the world in a similar way. One of the countries that is highly affected by land use changes is the Philippines. In the past century, a large part of the country was deforested as a result of intensive commercial logging activities and expansion of the agricultural area. Land use changes in the Philippines have major consequences for the landscape and the functions it can provide. Land use changes have caused biodiversity to be under threat and slopes have become unstable, which may cause landslides. Unsustainable land use practices restrict the opportunities for people to make a living in the future. These land use changes and their effects also apply for the study area of this research. The study area is a part of the municipality of San Mariano in the northeastern part of the Philippines. This area, which comprises 48,000 ha, transformed from a forested area with few inhabitants in the 1900s to an area that is currently largely cleared and which is home to approximately 4,000 families, which are predominantly dependent on agriculture. The area is situated in the transition zone between the lowlands of the Cagayan valley and the uplands of the Sierra Madre mountain range. At present, the study area has a land use gradient from intensive agriculture near San Mariano, with mainly rice and yellow corn, via a scattered pattern of rice, yellow corn, banana, grasses and trees to residual and primary forest in the eastern part. Large-scale commercial logging stopped in the area. Currently, the main land use changes are agricultural expansion and small-scale (illegal) logging activities.

Land use change forms the interface where the human and the natural system interact. Land change science is therefore a field that involves many disciplines. To study land use these various disciplines have developed their own paradigms and methods. However, disciplinary approaches can only cover part of the complex system responsible for land use changes. To understand the dynamics of land use change in a comprehensive way, new, interdisciplinary methodologies that integrate the many aspects of the land use system are necessary.
To position the research approaches of this dissertation in the wide array of methodologies in land use science two broad methodological approaches are identified: 'from pattern to process' and 'from process to pattern'. The pattern-based method can be described as spatially oriented, GIS (Geographical Information System) based approach, which starts with analysing land use patterns by identifying correlations between land use and its explanatory factors. The process-based approach originates from the social sciences and starts with analysing actors and processes and aims at modelling the land use pattern from these relations. Broadly speaking, the distinction between pattern-based and process-based research coincides with the distinction between inductive and deductive methodologies. The former is strong in describing patterns empirically, but has a weak connection with causal processes. The latter is strong in describing causal structure, but is often less easy to parameterise, calibrate and validate for real world cases.
The main objective of this dissertation is to develop methodologies to identify and integrate factors that are important in the land use system in order to describe and model the com-

plex land use system in a comprehensive manner. To facilitate the integration of human and natural sciences both 'pattern to process' and 'process to pattern' research is carried out. The methodological challenges that are addressed in this study include bridging differences in spatial scales, organisational levels and temporary scales; identification of appropriate units of analysis; combining different disciplinary paradigms and developing new paradigms that unify the disciplines into one concept.

As an exploratory study two datasets were analysed to identify the explanatory factors of land use in the area. First, a statistical analysis was performed on household survey data from interviews. This analysis included field characteristics as well as household variables and aims at explaining the occurrence of corn, banana or wet rice on a particular field. The results from this study were used to inform a second, spatial analysis. The factors that turned out to be important for the allocation of the land use types in the study area can be categorised in three groups: accessibility, origin of the land managers and biophysical constraints. Despite the efforts to integrate the approaches the factors that were selected by the stepwise procedure varied between the household analysis and the spatial analysis as well as the relative importance of the variables. These differences stem from differences in unit of analysis, differences in sample design, differences in the themes included and differences in the methods that were used to collect the data.

The statistical, inductive approach from the exploratory study reveals correlations rather than causal relations. To better understand, structure and describe the processes of land use in the area a theoretical, deductive framework was adopted, which consists of a qualitative model (the Action-in-Context framework) describing causal relations in actor decision-making. This framework was used to construct a quantified deductive model explaining crop choice on fields. This model was tested using the household data and compared with a statistical (inductive) analysis of the same data. The performance of both approaches is similar. A major difference between the two approaches is that the deductive approach tests the full causal structure, which leads to a better grip on causal relations and supports theory building whereas the statistical model is constructed to fit the data as best as possible.

An important way to integrate different disciplines is to integrate the levels of analysis of these disciplinary approaches. A statistical approach to combine different organisation levels and spatial scales is multilevel analysis. This method explicitly addresses the hierarchical levels in the data and shows what proportion of the variance occurs at which level. Aggregating or disaggregating variables to the unit of analysis, which may violate the statistical assumption of the model use, is not necessary with this method. The multilevel model for the case was informed by the results from the analyses above and incorporates the field, household and village level. The case study revealed the importance of the household level in explaining land use at the detailed level of the study area. In some of the constructed multilevel models the village variability could partly be explained by field variables. Generalising this observation, it can be concluded that in land use studies all organisational levels between the resolution and the extent should be examined on their potential importance in explaining land use. The strength of multilevel analysis is that it allows to make a multitude of propositions between higher and lower levels and scales and to test these relations.

Subsequently, the information from all preceding analyses was integrated in a dynamic spatial model, which is used to make projections of land use under different scenario conditions. The relations of the deductive household model were translated to the spatial level to create suitability maps that are used as input in a modelling exercise using the CLUE-S model (Conversion of Land Use and its Effects at Small regional extent). This approach was compared with a CLUE-S model that incorporates suitability maps derived with the statistical spatial analysis. For a land use projection for 2015 these two modelling approaches are different in 15 % of the cells, which can be contributed to the different specifications of the suitability maps. However, considering only the cells that actually changed the two approaches have only 50 % in common. The two different approaches to specify the land use model each have consequences for the use of the model in policy making. Inductive, statistical approaches are weaker in the description of causality and processes. This restricts models that are based on an inductive analysis to model large changes in processes, for example the introduction of a new land use type. If instead a theoretical, deductive approach is used to derive and describe relations between land use and its explanatory factors the models can be made more flexible and the introduction of new land use types and changes in processes during the modelling period can be facilitated. The CLUE-S model with the deductive approach to specify the land use suitability is most valuable in small study areas where detailed actor research can be carried out. Large-scale studies can best be carried out with an inductive approach and can be used for the rough identification of hotspots of land use change.

In order to use land use models in policy-making effectively the projections of future land use patterns should be translated into normative indicators. The Philippines are a global hotspot of biodiversity and the study area borders the largest contiguous forest area of the country. Therefore, an assessment of the effects of land use change on biodiversity was made. For three land use scenarios land use maps are projected for the year 2015 using the CLUE-S model with the deductive specification mentioned above. The scenarios are different in the level of agricultural expansion and forest conservation management. Furthermore, the relation between landscape characteristics and endemic forest bird species richness was determined. This relation was used to create maps with an indicator for the value of a location for endemic bird conservation for the present situation and for the projected land use maps. The results showed that the pattern of the effects of land use changes can be different from the pattern of land use changes themselves because land use changes have off-site effects and land use changes have different effects at different locations. The scenarios clearly show the areas that are under threat. The combination of a state of the art land use model and biodiversity mapping can provide quantitative indicators to project changes in biodiversity due to land use change. The land use model is capable of incorporating the human dimension of land use change and the competition between land use types. This is important to project the effects of policy measures on the land use system. The biodiversity assessments of the projected landscapes can be used to evaluate policy options for conservation management.

The main land use developments in the area are agricultural expansion and small-scale (*i.e.* non-commercial) logging. Especially under a high growth scenarios agricultural expansion poses a threat to the forest. So far, the area that is currently under forest was spared by its natural defence of steep slopes and inaccessibility. In a negative scenario all forested areas will eventually be used for agriculture. If the agricultural system practiced is unsustainable this development would only provide a solution until the area has degraded. As an alternative a large part of the foreseen agricultural expansion could be realized in areas that are currently under grassland. Furthermore, productivity of the current land use systems may be improved. The key to a sustainable future in San Mariano is to control agricultural expansion due to population growth (natural and migration), to direct agriculture expansion to appropriate areas, to invest in viable agricultural systems and to conserve natural resources. However, this approach demands strong governance and sufficient investments.

In this thesis it is especially the combination of approaches that have led to a greater understanding of the land use system in the study area. Qualitative information was used to describe land use processes in the area. Quantitative data were used to analyse the land use system at the household level and in a spatially explicit way for the complete study area. In the analyses both deductive and inductive research methods were used. All methods were aimed at integrating different levels and thematic information that originated from different disciplines. Moving between empirical, inductive methods and theoretical, deductive methods is a useful approach to stimulate theory building. Methods that can deal with multiple levels proved to be valuable for integration of disciplinary approaches, which often greatly differ in their unit of analysis.

Some scholars argue that the time is ripe for an overarching theory of land use change. It is doubtful if it is possible to find a theory that would be acceptable for all disciplines involved in land use science and which can cover all the important phenomena. An all-encompassing theory of land use change is still far off. I would argue that it is currently more fruitful to develop methodological theories of parts of the system that describe interactions and feedbacks between components of the system. This dissertation includes some examples of such theories and methods. The joint understanding from these analyses was combined in a modelling framework that added to the insights in the overall land use system.

Samenvatting

Overmars, K.P. 2006. Het relateren van patroon en proces in landgebruikstudies, geïllustreerd met een voorbeeld in San Mariano, Isabela, Filippijnen.

De bedekking van het aardoppervlak is in het verleden op uitgebreide schaal veranderd door menselijk handelen en dit gaat momenteel nog steeds in een hoog tempo door. Deze landgebruiksveranderingen hebben grote gevolgen voor het mondiale milieu doordat ze bijvoorbeeld het klimaat, ecosysteemfuncties en duurzaamheid beïnvloeden. Niet alle regio's in de wereld zijn in dezelfde mate beïnvloed door landgebruiksverandering. In de Filippijnen echter is in de afgelopen eeuw een groot gedeelte van het land ontbost door intensieve, commerciële houtkap en uitbreiding van de landbouw. De landgebruiksveranderingen hebben grote consequenties voor het landschap en de functies die het landschap heeft. Ze bedreigen de biodiversiteit en de stabiliteit van hellingen, wat aardverschuivingen kan veroorzaken. Niet-duurzame landgebruikspraktijken kunnen de kansen van mensen om in de toekomst in hun levensonderhoud te voorzien verkleinen. Deze veranderingen in landgebruik en de effecten daarvan gelden ook voor het studiegebied van dit onderzoek.
Het studiegebied is een deel van de gemeente San Mariano in het noordoosten van de Filippijnen. Dit gebied, dat 48.000 ha groot is, is veranderd van een bebost gebied met weinig inwoners in de jaren 1900 tot een gebied dat momenteel grotendeels is ontbost en waar ongeveer 4000 gezinnen wonen die voornamelijk afhankelijk zijn van landbouw. Het gebied vormt de overgang tussen het laagland van de Cagayan vallei en bergen van de Sierra Madre. Tegenwoordig heeft het studiegebied een landgebruiksgradiënt van intensieve landbouw met voornamelijk teelt van natte rijst en maïs nabij San Mariano in het westen via een fijnmazig patroon van rijst, maïs, bananen, gras en bomen naar overgebleven bos en primair bos in het oosten. Grootschalige commerciële boskap is gestopt in het gebied. Tegenwoordig zijn de uitbreiding van het landbouwgebied en kleinschalige (illegale) boskap de belangrijkste veranderingen van het landgebruik.

Landgebruiksverandering komt tot stand door interactie van sociale en natuurlijke systemen. De wetenschap die verandering van landgebruik onderzoekt bestaat daarom uit vele disciplines. In het bestuderen van landgebruik hebben de verschillende disciplines hun eigen paradigma's en methodes ontwikkeld. Echter, deze disciplinaire benaderingen kunnen alleen een gedeelte van het complexe dynamiek van landgebruiksveranderingen beschrijven. Voor een meer omvattend begrip zijn nieuwe methoden nodig die meerdere delen van het systeem integreren.
Om de onderzoeksaanpak te positioneren in de reeks methoden die worden gebruikt in landgebruiksonderzoek onderscheiden we twee globale methoden: "van patronen naar processen" en "van processen naar patronen". De patroongerichte methode kan worden omschreven als een ruimtelijk georiënteerde, op GIS (Geografisch Informatie Systeem) gebaseerde benadering, die begint met de analyse van ruimtelijke patronen en vervolgens correlaties probeert te vinden tussen landgebruik en de verklarende factoren daarvan. De procesgerichte aanpak is afkomstig uit de sociale wetenschappen; de methode begint met de analyse van actoren en processen en probeert vervolgens de patronen van landgebruik te modelleren uit deze relaties. Globaal gezien valt het onderscheid tussen de patroongerichte en de procesgerichte aanpak samen met het onderscheid tussen inductieve en deductieve methodologie. De kracht van de eerstgenoemde is om patronen te beschrijven op een

empirische manier. De kracht van de laatstgenoemde is het beschrijven van de causale structuur. Deze aanpak is vaak moeilijker te parameteriseren, te kalibreren en te valideren voor praktijkgerichte studies.
De belangrijkste doelstelling van dit proefschrift is om methoden te ontwikkelen om de factoren die belangrijk zijn in het landgebruiksysteem te identificeren en te integreren om daarmee het complexe landgebruiksysteem te beschrijven en te modelleren in een veelomvattende manier. Om de integratie van sociale en natuurlijke wetenschappen te vergemakkelijken is zowel onderzoek "van patronen naar processen" als "van processen naar patronen" gedaan. De methodologische uitdagingen die in deze studie worden behandeld zijn onder meer het overbruggen van verschillen in ruimtelijke schalen, organisatorische lagen en temporele schalen, het identificeren van de juiste eenheid voor analyse, het combineren van verschillende disciplinaire paradigma's en het ontwikkelen van nieuwe paradigma's die disciplines verenigen in één concept.

In een verkennende studie zijn twee datasets geanalyseerd om de factoren te identificeren die het landgebruik in het studiegebied verklaren. Als eerste is een statistische analyse uitgevoerd op data van een enquête onder huishoudens. Deze analyse bevatte zowel perceelskarakteristieken als huishoudensvariabelen en probeerde de aanwezigheid van maïs, bananen en natte rijstbouw op een bepaald veld te verklaren. De resultaten van deze studie zijn vervolgens gebruikt als informatie voor een tweede, ruimtelijke analyse. De factoren die belangrijk bleken te zijn voor het voorkomen van de verschillend landgebruikstypen in het gebied kunnen in drie categorieën worden verdeeld: bereikbaarheid, oorsprong van de landgebruikers en biofysische beperkingen. Ondanks de pogingen om alle factoren op te nemen in beide methoden bleken de factoren die werden geselecteerd door de stapsgewijze regressieprocedures en hun relatieve belang verschillend in de analyse van de huishoudens en de ruimtelijke methode. Deze verschillen komen voort uit verschillen in eenheid van analyse, verschillen in het ontwerp van de steekproef, verschillen in de thema's die zijn gebruikt en verschillen in de methodes om de gegevens te verzamelen.
De statistische, inductieve benadering van de verkennende studie toont veeleer correlaties dan causale processen. Om de processen die het landgebruik in het gebied bepalen beter te begrijpen en te structureren is vervolgens een theoretisch, deductief raamwerk toegepast. Dit raamwerk bestaat uit een causaal beslissingsmodel van actoren (het "Action-in-Context" raamwerk). Dit raamwerk is gebruikt om een gekwantificeerd, deductief model te construeren dat de gewaskeuze op een veld verklaart. Dit model is vervolgens getoetst met behulp van de huishoudensgegevens en vergeleken met een statistische (inductieve) analyse van dezelfde gegevens. De empirische prestaties ('fit') van de twee aanpakken zijn vergelijkbaar. Een groot verschil tussen de twee benaderingen is dat de deductieve aanpak de volledige causale structuur test, wat tot een beter begrip van de causale relaties leidt en bevorderend is voor het construeren van theorie, terwijl de statistische analyse is gemaakt om zo dicht mogelijk bij de empirische gegevens te blijven.
Een belangrijke manier om verschillende disciplines te integreren is om hun verschillende analyseniveaus te integreren. Een statistische methode om verschillende organisatieniveaus en ruimtelijke schalen te combineren is een 'multilevel-analyse'. Deze methode houdt expliciet rekening met de hiërarchische niveaus in de data en laat zien welk percentage van de variantie aanwezig is op welk niveau. Aggregeren of desaggregeren van variabelen naar één eenheid van analyse, wat de statistische aannames kan ondermijnen, is met deze methode niet nodig. De resultaten van de bovenstaande analyses zijn gebruikt in een voorbeeldstudie van een multilevel-aanpak die gebruik maakt van een veld-, huishoudens- en

dorpsniveau. Deze studie bracht naar voren dat het huishoudensniveau belangrijk is in het verklaren van landgebruik op het gedetailleerde niveau van de studie. In een aantal van de multilevelmodellen kon de variabiliteit tussen dorpen deels verklaard worden door veldvariabelen. Dit generaliserend kan worden geconcludeerd dat in studies naar landgebruik alle organisatieniveaus die liggen tussen de resolutie van het GIS en de grootte van het gebied bekeken moeten worden om te zien of ze potentieel belangrijk zijn in het verklaren van landgebruik. De kracht van de multilevel-analyse is dat deze methode het mogelijk maakt om hypotheses met betrekking tot de hogere en lagere organisatieniveaus en ruimtelijke schalen te maken en deze daadwerkelijk te toetsen.

Vervolgens is de informatie van alle voorgaande analyses geïntegreerd in een ruimtelijk, dynamisch model, dat is gebruikt om projecties van landgebruik te maken onder verschillende scenariocondities. De relaties uit het deductieve actorenmodel zijn vertaald naar het ruimtelijke niveau om daarmee geschiktheidskaarten te maken die gebruikt konden worden in een toepassing van het CLUE-S model ('Conversion of Land Use and its Effects at Small regional extent'). Deze aanpak is vergeleken met een CLUE-S model dat gebruik maakte van geschiktheidkaarten die afgeleid zijn van een ruimtelijk-statistische analyse. Voor een landgebruiksprojectie voor 2015 zijn deze twee modellen verschillend in 15 % van de gridcellen van de GIS-kaart, wat kan worden toegeschreven aan de verschillende specificaties van de geschiktheidskaarten. Echter, als alleen de cellen in aanmerking worden genomen waarin verandering wordt voorspeld hebben ze maar 50 % gemeen. Voor het maken van beleid is daarom een keuze tussen de twee verschillende manieren om het landgebruiksmodel te specificeren van belang. Inductieve, statistische methoden zijn zwakker in het beschrijven van causaliteit en processen. Dit beperkt het nut van deze modellen om grote veranderingen in processen te modelleren, bijvoorbeeld de introductie van een nieuw landgebruikstype. Als in plaats daarvan een theoretische, deductieve methode wordt gebruikt om relaties tussen landgebruik en zijn verklarende factoren af te leiden en te beschrijven kunnen de modellen flexibeler worden gemaakt, en kan het effect van bijvoorbeeld de introductie van nieuwe landgebruikstypen en veranderingen in processen tijdens de modelleerperiode zichtbaar worden gemaakt. Het CLUE-S model met deductieve specificatie van de landgebruiksgeschiktheid is het meest waardevol in kleine studiegebieden waar gedetailleerd onderzoek naar de actoren gedaan kan worden. Studies in grote gebieden kunnen het best worden uitgevoerd met de inductieve benadering en kunnen dan bijvoorbeeld gebruikt worden voor de identificatie van 'hotspots' van landgebruiksverandering, onder aanname van voortgang van bestaande processen.

Om landgebruiksmodellen effectief te gebruiken in het maken van beleid moeten de projecties van landgebruikspatronen zoveel mogelijk vertaald worden in normatieve indicatoren zoals bijvoorbeeld welvaart of biodiversiteit. De Filippijnen is een mondiale 'hotspot' van biodiversiteit en het studiegebied grenst aan het grootste aaneengesloten stuk bos in het land. Daarom is een inventarisatie gemaakt van de effecten van landgebruik op biodiversiteit. Voor drie landgebruikscenario's zijn projecties gemaakt naar de landgebruikskaart van 2015 met behulp van het CLUE-S model met de deductieve modelspecificatie zoals genoemd in de vorige alinea. De scenario's verschillen in de hoeveelheid landbouwuitbreiding en het bosbeschermingsbeleid. De relatie tussen landschapskarakteristieken en de soortenrijkdom van endemische bosvogelsoorten is bepaald. Deze relatie is gebruikt om kaarten te maken met een indicator voor de waarde van een locatie voor het beschermen van endemische vogels voor de huidige situatie en voor de voorspelde landgebruikskaarten. De resultaten laten zien dat het patroon van de effecten van landgebruiksveranderingen kan verschillen van het patroon van de landgebruiksveranderingen zelf, omdat landgebruiksveranderingen effecten buiten de locatie zelf kunnen hebben en omdat landgebruiksveranderingen verschillende effecten hebben op verschillende locaties. De scenario's laten duidelijk de gebieden zien die bedreigd worden. De combinatie van dit landgebruiksmodel volgens de huidige stand van de wetenschap en het karteren van biodiversiteit kan daarom kwantitatieve indicatoren verschaffen om veranderingen in biodiversiteit als gevolg van landgebruiksveranderingen te projecteren. Het landgebruiksmodel is in staat om de menselijke dimensie van landgebruiksverandering mee te nemen, evenals de competitie tussen landgebruikstypen. De inventarisatie van de biodiversiteit van de geprojecteerde landschappen kan worden gebruikt om beleidsmaatregelen ten behoeve van natuurbehoud te evalueren.

De belangrijkste landgebruiksontwikkelingen in het gebied zijn uitbreiding van de landbouw en kleinschalige houtkap. Speciaal in de scenario's met een hoog groeicijfer vormt de uitbreiding van de landbouw een bedreiging voor het bos. Tot nu toe is het gebied dat momenteel bebost is gespaard gebleven vanwege zijn natuurlijke verdediging van steile hellingen en ontoegankelijkheid. In een negatief scenario zullen alle beboste gebieden uiteindelijk gebruikt worden voor landbouw. Als alternatief zou een groot deel van de landbouwuitbreiding gerealiseerd kunnen worden in de gebieden die momenteel grasland zijn. Daarbij zou de productiviteit van het huidige landbouwsysteem kunnen worden verbeterd. De sleutel naar een duurzame toekomst in San Mariano is het beheersen van de uitbreiding van de landbouw als gevolg van populatiegroei (natuurlijk en migratie), het sturen van landbouwuitbreiding naar geschikte plaatsen, te investeren in levensvatbare landbouwsystemen en het beschermen van natuurlijke hulpbronnen. Echter, deze aanpak vraagt sterk bestuur en voldoende investeringen.

In dit proefschrift is het in het bijzonder de combinatie van methoden dat tot een groter inzicht in het landgebruiksysteem in het gebied heeft geleid. Kwalitatieve informatie is gebruikt om landgebruiksprocessen in het gebied te beschrijven. Kwantitatieve gegevens zijn gebruikt om het landgebruiksysteem te analyseren op het huishoudniveau en op een ruimtelijke manier voor het hele studiegebied. In de analyses zijn zowel deductieve en inductieve onderzoeksmethoden gebruikt. Alle methoden streven naar het integreren van verschillen niveaus en thematische informatie die is voortgekomen uit verschillende disciplines. Het op en neer bewegen tussen empirisch georiënteerde inductieve methoden en meer theoretisch georiënteerde deductieve methoden is een bruikbare benadering om ontwikkeling van methodologische theorie te stimuleren. Methoden die kunnen omgaan met verschillende niveaus bleken waardevol te zijn voor het integreren van disciplinaire benaderingen die vaak erg verschillen in hun eenheid van analyse.
Sommige wetenschappers betogen dat de tijd rijp is voor een alles-overbruggende theorie van landgebruiksverandering. Het is twijfelachtig of het mogelijk is om een theorie te vinden die acceptabel is voor alle disciplines die betrokken zijn in landgebruikswetenschappen en die alle belangrijke fenomenen kan omvatten. Een allesomvattende landgebruikstheorie is nog ver weg. Ik zou willen betogen dat het momenteel productiever is om theorieën te ontwikkelen voor delen van het systeem, met name die welke interacties en terugkoppelingen tussen componenten van het systeem kunnen beschrijven. Dit proefschrift bevat enkele voorbeelden van zulke theorieën en methoden. Het inzicht van deze analyses samen is gecombineerd in een modelleerraamwerk dat bijdraagt aan het inzicht in het totale landgebruiksysteem.


ACE 597

Curriculum Vitae

Koen Pieter Overmars was born on the 7th of August 1976 in the Noordoostpolder, the Netherlands. From September 1988 to June 1994 he followed secondary education at the Zuyderzee College in Emmeloord. In the following year Koen studied Industrial Design Engineering at the Delft University of Technology. During this year Koen found out that his interest was more in the natural environment than in the industrial world. After obtaining the propaedeutic exam (cum laude) in Delft he decided to change his field of study. In September 1995 he started the study Soil, Water and Atmosphere at Wageningen University. He finished his study in September 2000 (cum laude) in the specialisation Soil Inventory and Land Evaluation. Before starting his Ph.D. research Koen was working as a GIS and river modelling consultant for Meander Consultancy and Research. The research that resulted in this dissertation was conducted at the CML (Institute of Environmental Sciences) at Leiden University and at the Laboratory of Soil Science and Geology at Wageningen University. This research started in May 2001 and was finished in October 2005.