

Land-use Change in the Central Highlands of Vietnam

A spatial econometric model combining
satellite imagery and village survey data

Dissertation

zur Erlangung des Doktorgrades
der Fakultät für Agrarwissenschaften
der Georg-August-Universität Göttingen

vorgelegt von

Daniel Müller

geboren in Friedrichshafen

Göttingen, March 2003

-
1. Referent: Prof. Dr. Manfred Zeller
2. Koreferent: Prof. Dr. Stefan von Cramon-Taubadel

Tag der mündlichen Prüfung: 22. Mai 2003

Author:

Daniel Müller
Agricultural Economist, M.Sc.

Contact:

Institute of Rural Development
Georg-August-University of Göttingen
Waldweg 26
37073 Göttingen
Phone: +49 (0) 551 - 395668
Fax: +49 (0) 551 - 393976
Email: danielix@gmx.net

Abstract

This dissertation investigates the role of geophysical, agroecological, and socioeconomic determinants of land-use change during the last 25 years in two districts of Dak Lak province in the Central Highlands of Vietnam. The analysis of these determinants allows to assess the influences of various rural development policies on land-cover changes. Landsat satellite images from the same cropping period of the years 1975, 1992 and 2000 are interpreted to detect land-cover change between the two time periods. A survey in randomly selected villages provides primary recall data on socioeconomic and policy variables hypothesized to influence land-use change. Secondary data on rainfall, soil suitability, and topography were obtained from meteorological stations, from a digital soil map and a digital elevation model. All data were spatially referenced using GIS software. Survey data is merged with spatially explicit raster data using accessibility catchments, which are designed to approximate village areas based on the estimated travel costs from each cell to the village location. A reduced-form, multinomial logit model is used to estimate the influence of hypothesized determinants on land use and the probabilities that a certain pixel has one of five land-cover classes during either of the two periods under consideration.

Results suggest that the first period from 1975 to 1992 was characterized by land-intensive agricultural expansion and the conversion of forest into grass and agricultural land. During the second period, since 1992, the rapid, more labor- and capital-intensive growth in the agricultural sector was enabled by the introducti-

on of fertilizer, improved access to rural roads and markets, and the expansion of irrigated areas. These policies, combined with the introduction of protected forest areas and policies discouraging shifting cultivation during the second period reduced the pressure on forests while at the same time increasing agricultural productivity and incomes for a growing population. Forest cover during the second period mainly increased due to the regeneration of areas formerly used for shifting cultivation.

Policy simulations mimic the influences from potential policy interventions on land use. Land-cover categories for the policy simulations are aggregated to three classes and a spatial sample is drawn to concentrate the analysis on changes influenced by anthropogenic interventions. In that way, the focus of the simulations is on land-use changes within already cultivated areas and at the agricultural frontier where most land-use changes relevant for policy makers take place. The simulations are limited to the second period from 1992 to 2000 as this is the relevant period for potential policy measures.

The four rural policy scenarios are carried out for low and high levels of investments in irrigation facilities, for an enlarged area under forest protection and for a combination of high investments in irrigation combined with increased forest protection, an often mentioned first-best policy option. Spatially explicit simulations of these policy interventions facilitate spatially explicit results, which can enable decision makers to formulate and implement geographically targeted rural development interventions. Simulation results suggest modest land-use changes in the expected directions under the four scenarios. A visual assessment of the location of the expected changes enables the identification of probable hotspots of land-cover conversions and allows for an ex-ante evaluation of the impacts of proposed policy interventions on the magnitude of land-use change.

Zusammenfassung

Diese Dissertation untersucht den Einfluss von geophysikalischen, agrarökologischen und sozioökonomischen Bestimmungsfaktoren auf Veränderungen der Landnutzung während der letzten 25 Jahre in zwei Distrikten der Provinz Dak Lak im Zentralen Hochland Vietnams. Aufbauend auf der Analyse dieser Bestimmungsfaktoren werden die Wirkungen verschiedener politischer Maßnahmen ländlicher Entwicklung auf Änderungen der Landbedeckung bewertet. Landsat Satellitenbilder aus den gleichen Anbauperioden der Jahre 1975, 1992 und 2000 werden interpretiert, um Veränderungen der Landbedeckung während dieser Zeitspanne herauszuarbeiten. Primärdaten wurden anhand einer Umfrage in zufällig ausgewählten Dörfern erhoben und erlauben die Überprüfung der Einflüsse sozioökonomischer Variablen und Politikindikatoren auf Landnutzungsveränderungen anhand aufgestellter Hypothesen. Verwendete Sekundärdaten über Niederschlag, Bodeneignung und Topographie stammen von meteorologischen Messstationen, einer digitalen Bodenkarte und einem digitalen Höhenmodell. Alle Daten sind in georeferenzierter Form in einem Geographischen Informationssystem gespeichert. Die Daten aus der Dorfumfrage werden mit räumlich expliziten Rasterdaten auf Basis von Dorfeinzugsgebieten verbunden. Diese Dorfeinzugsgebiete basieren auf der Höhe der geschätzten Transportkosten von jeder Rasterzelle in der Forschungsregion zur geographischen Position des nächstgelegenen Dorfes. Ein multinomiales Logitmodell in reduzierter Form wird geschätzt, um den Einfluss der bestimmenden Faktoren auf die Flächennutzung herauszuarbeiten und die Wahrscheinlichkeiten zu berechnen, die eine bestimmte Zelle

hat, in eine der fünf Landabdeckungskategorien während einer der zwei Perioden zu fallen.

Ergebnisse zeigen, dass die erste Periode von 1975 bis 1992 durch landintensive landwirtschaftliche Expansion und die Umwandlung von Waldflächen in Gras- und Ackerland gekennzeichnet ist. Während der zweiten Periode seit 1992 wurden das hohe, mehr arbeits- und kapitalintensive Wachstum im landwirtschaftlichen Sektor durch die Einführung von Düngemitteln, den verbesserten Zugang zu Straßen und Märkten sowie durch die Expansion bewässerter Flächen ermöglicht. Diese Politikstrategien, zusammen mit der Ausweisung weiterer geschützter Waldflächen und Anreizen zur Eindämmung des Wanderfeldbaus während der zweiten Periode, verringerten den Druck auf die Waldflächen bei gleichzeitiger Erhöhung der landwirtschaftlichen Produktivität und der Einkommen für eine wachsende Bevölkerung. Der Waldbestand erhöhte sich in der zweiten Periode hauptsächlich auf Flächen, die früher für Wanderfeldbau verwendet wurden.

Politiksimulationen werden modelliert, um Einblicke in die Auswirkungen möglicher Politikinterventionen auf die Flächennutzung zu erhalten. Die Landbedeckung wird für die Simulationen in drei Kategorien aggregiert. Eine Stichprobe von Zellen wird dergestalt ausgewählt, dass die Analyse auf Änderungen konzentriert werden kann, die durch anthropogene Eingriffe verursacht wurden. So wird erreicht, dass die simulierten Flächennutzungsänderungen innerhalb bereits genutzter Gebiete und angrenzender Zonen analysiert werden, in welchen der überwiegende Teil der Flächennutzungsänderungen liegt, die für politische Entscheidungsträger relevant sind. Ausserdem werden die Simulationen auf die zweite Periode von 1992 bis 2000 begrenzt, da dies die relevante Periode für mögliche Politikmaßnahmen ist.

Die vier Politikscenarien sollen mögliche Auswirkungen politischer Interventionen auf landwirtschaftliche Nutzflächen in der Forschungsregion simulieren. Simulationen werden durchgeführt für reduzierte und höhere Investitionen in die Bewässerungsinfrastruktur, für mehr Flächen unter Waldschutz und für eine Kombination hoher Investitionen in Bewässerungsinfrastruktur zusammen mit erhöhtem Waldschutz, eine häufig erwähnte win-win Politik. Räumlich explizite Simulationen von Politikeingriffen ermöglichen räumlich explizite Ergebnisse, welche die politischen Entscheidungsträger befähigen können, politische Richtlinien für ländliche Regionalentwicklung mit einer besseren räumlichen Zielorientierung zu formulieren und umzusetzen. Die Ergebnisse der Simulationen zeigen begrenz-

te Flächennutzungsänderungen in allen vier Szenarien. Eine visuelle Interpretation der erwarteten Änderungen ermöglicht die Identifizierung vermeintlicher hot-spots, an denen Flächenänderungen geschehen können. Weiterhin läßt eine ex-ante Bewertung der Auswirkungen des jeweiligen Politikeingriffes Rückschlüsse auf die Höhe möglicher Flächennutzungsänderungen zu.

Acknowledgments

The dissertation at hand is the result of a three-year research project. In the course of these three years a large number of people supported the development of this dissertation making it impossible name them all.

I am grateful to Professor Dr. Manfred Zeller, director of the Institute of Rural Development (IRD) of the Georg-August-University Göttingen, for his supervision, guidance and support. I am also deeply grateful to Dr. Regina Birner of the IRD for numerous intensive discussions and her assistance throughout the three years as well as to Prof. Dr. Stefan von Cramon-Taubadel from the Chair of Agricultural Policy of the Georg-August-University Göttingen for his thorough and helpful evaluation of this dissertation.

Dr. Hans Helmrich, Chief Technical Advisor of the Sustainable Resources in the Lower Mekong Basin Project (SMRP), initiated the research project and always supported and encouraged us. Michael Glück, Technical Advisor of SMRP, was in charge of the administrative part of our Vietnam component. I am grateful to them for their support and much of the improvement of my management skills are due to their guidance. At the Ministry of Agriculture and Rural Development I acknowledge the constant backing of the national project director and vice-director of the Department of Forest Development, Dr. Nguyen Hong Quan. I also want to thank Dr. Nguyen Hai Nam, at the time national coordinator of SMRP, and the staff of SMRP in the Hanoi office.

In Dak Lak province uncountable people were involved in the research project at village, commune, district and provincial level. They treated me with hospitality,

curiosity and respect. The friendly and encouraging atmosphere in the Department of Agriculture and Rural Development (DARD) contributed essentially to the final outcome of my studies. In this context, I would like to thank particularly Mr. Phan Muu Binh, director of the DARD in Dak Lak province. Without his support our research would not have been possible. I am indebted to Tran Huu Nghi, Tran Ngoc Thanh and the other members of SMRP project in Dak Lak. Their helpful assistance and openness was a basis for the fruitful work in Dak Lak. I would like to express my sincere gratitude to Nguyen Ai Lien, who did an outstanding job as research assistance and mastered in administration, translation, and data management throughout the time we worked together. The dissertation also profited substantially from the three MSc students of the research project in Dak Lak, namely Tran Huu Nghi, Pham Tan Ha, and Nguyen Thi Thanh Huong. They impressed me with their professional skills, commitment and reliability. Special credits go to my friend Tran Huu Phu, who made my 18 months stay in Dak Lak a very pleasant one.

Related to our activities in Yen Chau district of Son La province, I am indebted to the Chief Technical Advisor of the Social Forestry Development Project (SFDP) of GFA/GTZ, Dr. Elke Förster, and to the Technical Advisor of SFDP, Dr. Björn Wode, who assisted and promoted us. I want to thank the two MSc students who carried out their MSc research in Yen Chau under my supervision, Ms. Ta Thi Tham and Ms. Le Thi Giang. I enjoyed working together with them.

I appreciated the collaboration with my colleagues at the IRD, particularly with Miet Maertens, María A. Mániz Costa, Pham Manh Cuong, and Stefan Schwarze. Their stimulating suggestions and constructive criticism inspired and encouraged me. I want to thank Matthias Nott for introducing me to the typesetting system \LaTeX , which I used for the preparation of this thesis, Michael Epprecht from IFPRI-Vietnam for numerous fruitful discussions on GIS and personal computing, Darla Munroe for the intensive and interesting debates via email and Christoph Feldkötter for his invaluable contributions in GIS and Remote Sensing aspects.

Last, but of course not least, I want to thank my parents. This thesis is dedicated to them!

Göttingen, May 2003

Daniel Müller

Affiliations

The research project entitled “Development of a Watershed Information System for the Assessment of Land Use Systems and Conflicts in the Mountain Regions of the Lower Mekong Basin – Case Studies in Vietnam and Thailand” was initiated by and is affiliated with the [Sustainable Management of Resources in the Lower Mekong Basin Project](#) (SMRP) and its Vietnamese partner institutions, the [Ministry of Agriculture and Rural Development](#) (MARD) together with its line department on provincial level. SMRP is a technical cooperation project between the [Mekong River Commission](#) (MRC) and the [German Agency for Technical Cooperation](#) (GTZ), commissioned by the [German Ministry of Economic Cooperation and Development](#) (BMZ).

The [Tropical Ecology Support Program](#) (TÖB) of the GTZ supported the research project on behalf of the BMZ’s environmental division, which provided the funding for a period of three years.

Scientific guidance and supervision came from the [Institute of Rural Development](#) (IRD) of the [Georg-August University of Göttingen](#).

The research project is endorsed as a Ph.D. dissertation of the [Land Use and Land Cover Change Project](#) (LUCC) by the [International Geosphere-Biosphere Programme](#) (IGBP) and the [International Human Dimensions Programme](#) (IHDP).

Contents

Abstract	III
Zusammenfassung	V
Acknowledgments	IX
Affiliations	XI
Table of Contents	XIII
List of Tables	XVII
List of Figures	XIX
Abbreviations	XXI
1 Introduction	1
1.1 Background	1
1.2 Problem analysis	3
1.3 Objectives	5
1.4 Outline	5
2 Theoretical background	9
2.1 Conceptual framework	9

2.2	Hypotheses of land-use change	14
2.3	The land rent model	16
2.4	Spatially explicit models of land-use change	20
2.4.1	Applications to forestry	21
2.4.2	Applications to agriculture	22
2.5	Summary	23
3	Methodology	25
3.1	Econometric approaches	25
3.1.1	Multinomial logit	27
3.1.1.1	Goodness-of-fit	29
3.1.1.2	Independence of irrelevant alternatives	30
3.1.2	Spatial econometric aspects	31
3.1.2.1	Spatial scale	33
3.1.2.2	Spatial interpolation	34
3.1.3	Possibilities to correct for spatial effects	35
3.1.4	Assessing predictive accuracy	37
3.2	Data collection	38
3.2.1	Land cover data	38
3.2.2	Geophysical data	42
3.2.3	Village survey	44
3.2.3.1	Socioeconomic village characteristics	46
3.2.3.2	Policy variables	47
3.3	Integration of socioeconomic and spatially explicit data	48
3.3.1	Village border approximation	49
3.3.2	Data integration and scale of analysis	51
3.4	Corrections for spatial effects	52
3.5	Expected effects of regressors	53
3.5.1	Geophysical data	54
3.5.2	Socioeconomic village characteristics	56
3.5.3	Policy variables	57
3.6	Methodological constraints and data limitations	59
3.6.1	Endogeneity issues	59
3.6.2	Data issues	60

3.7	Summary	62
4	Descriptive Analysis	63
4.1	Dak Lak province	64
4.2	Research area	66
4.2.1	Geophysical indicators	67
4.2.2	Socioeconomic history	70
4.2.3	Policy review	73
4.3	Outcome Variables	80
4.3.1	Land-cover changes	80
4.3.2	Agricultural development indicators	85
4.3.3	Economic growth and welfare	88
4.3.4	Environmental indicators	95
4.4	Summary	97
5	Econometric Analysis	99
5.1	Multinomial logit	100
5.2	Empirical results	101
5.3	Prediction accuracy	104
5.4	Independence of irrelevant alternatives	109
5.5	Model shortcomings	110
5.6	Summary	112
6	Policy simulations	113
6.1	Model specifications	114
6.2	Empirical results from base estimation	115
6.3	Description of scenarios	118
6.4	Scenario results	119
6.4.1	Low investments in irrigation	120
6.4.2	High levels of government investments	120
6.4.3	Effect of increased forest protection	122
6.4.4	Effect of high levels of government investments combined with forest protection	122
6.5	Summary	123

7 Conclusions and Policy Implications	125
Bibliography	131
Appendix	145
Glossary	161

List of Tables

3.1	Prediction matrix	37
3.2	Specifications of Landsat satellite images	39
3.3	Characteristics of Landsat ETM+	40
3.4	Measurement, level and source of exogenous variables	45
3.5	Kernel to derive spatially lagged slope	54
3.6	Predicted signs of exogenous variables	55
4.1	Agricultural land use and elevation	68
4.2	Population and area of Lak and Krong Bong districts, 2001	72
4.3	Commune population and population density, 1999	73
4.4	Average Euclidean distance to road categories	78
4.5	Land cover in km^2 and percent of total land	82
4.6	Landscape metrics	84
4.7	Correlation coefficients of selected variables	95
5.1	Correctly predicted cells at different cut-off points	105
5.2	Prediction matrices	106
5.3	Percentage of correctly predicted cells by category	107
5.4	Hausman test of IIA assumption	110
6.1	Prediction matrix for base estimation	116
6.2	Hausman test of IIA assumption, base prediction	118

6.3	Scenario 1 – Low levels of government investments	120
6.4	Scenario 2 – High investments in irrigation	121
6.5	Scenario 3 – Increased forest protection	122
6.6	Scenario 4 – High investments in irrigation combined with forest protection	123
A.1	Descriptive statistics for land use	148
A.2	Descriptive statistics for independent variables	149
A.3	Multinomial logit estimation of land use, model 2000 (closed forest as comparison group)	150
A.4	Multinomial logit estimation of land use, model 1990 (closed forest as comparison group)	151
A.5	Measures of fit for multinomial logit models	152
A.6	Likelihood ratio tests for combining outcome categories . . .	152
A.7	Likelihood ratio tests for independent variables	153
A.8	Descriptive statistics for variables in simulation	154
A.9	Multinomial logit estimation for base prediction (non-agricultural land as comparison group)	155

List of Figures

2.1	Conceptual framework of land-use change	11
2.2	Conceptual framework for econometric models	13
2.3	Land rent gradients and land use zones	19
3.1	Distance decay functions	32
3.2	Village boundaries based on accessibility, 2000	50
4.1	Location of research area	65
4.2	Population growth in Dak Lak province	66
4.3	Research area with administrative boundaries	67
4.4	Slope categories by land-cover class, 2000	69
4.5	Slope categories by land-cover class, 1992	69
4.6	Slope categories by land-cover class, 1975	70
4.7	Population growth in research area	71
4.8	Ethnic composition	74
4.9	Number of villages and major events	75
4.10	Introduction of technologies and primary schools	77
4.11	Travel time in hours to various locations	79
4.12	Research area on top of Landsat ETM+	81
4.13	Land-cover change trajectories from 1975 to 2000	83
4.14	Permanent upland and paddy area of survey households	86

4.15	Paddy rice yield by ethnic group and in total	87
4.16	Upland rice yield	88
4.17	Crop introduction and cultivation extent	89
4.18	Wealth indicators	90
4.19	Wealth indicators by ethnic group	91
4.20	Wealth indicators by distance from all-year road	92
4.21	Household food shortages	93
4.22	School enrollment by ethnic group	94
4.23	Change in state of natural resources, 1980 to 1990	96
4.24	Change in state of natural resources, 1990 to 2000	96
5.1	Odds ratios of selected explanatory variables, model 2000	102
5.2	Odds ratios of selected explanatory variables, model 1990	103
5.3	Maximal predicted probabilities, model 2000	108
5.4	Probability differences between observed and predicted categories, model 2000	109
6.1	Spatial sample for policy simulations	114
6.2	Prediction map for base estimation	117
A.1	Land cover maps	145
A.2	Rainfall and temperature in research area	146
A.3	Scatterplot matrix of selected socioeconomic indicators	147
A.4	Scenario 1 – Low investments in irrigation	156
A.5	Scenario 2 – High investments in irrigation	157
A.6	Scenario 3 – Increased forest protection	158
A.7	Scenario 4 – High investments in irrigation combined with forest protection	159

Abbreviations

BMZ	German Ministry for Economic Cooperation and Development
CGE	Computable General Equilibrium model
DANIDA	Danish Agency for Development Assistance
DARD	Department of Agriculture and Rural Development
DEM	Digital Elevation Model
df	Degrees of Freedom
DSM	Digital Soil Map
e.g.	Example Given
EM	Electromagnetic Energy
ETM+	Enhanced Thematic Mapper Plus
EROS	Earth Resources Observation Systems
FCRP	Fixed Cultivation and Resettlement Program
FIPI	Forest Inventory and Planning Institute
FPD	Forest Protection Department
GCP	Ground Control Points
GDLA	General Department for Land Administration
GDP	Gross Domestic Product
GIS	Geographic Information System
GPS	Global Positioning System

GTZ	Deutsche Gesellschaft für Technische Zusammenarbeit GmbH (German Agency for Technical Cooperation)
GSO	General Statistics Office
ha	Hectare
HEPA	Hunger Eradication and Poverty Alleviation Program
ICO	International Coffee Organization
i.e.	Id Est (that is)
IGBP	International Geosphere-Biosphere Programme
IHDP	International Human Dimensions Programme
IIA	Independence of Irrelevant Alternatives
IRD	Institute of Rural Development
km	Kilometer
LCC	Land-Cover Change
LR	Likelihood Ratio
LUCC	Land-Use and Land-Cover Change
LURC	Land-Use Right Certificate
LHS	Left-hand Side
LR	Likelihood Ratio
LUC	Land-Use Change
MARD	Ministry of Agriculture and Rural Development
m	Meter
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimator
MOLISA	Ministry of Labor, Invalids and Social Affairs
mm	Millimeter
MNL	Multinomial Logit model
MNP	Multinomial Probit
MRC	Mekong River Commission
MSS	Multispectral Scanner
NEZ	New Economic Zone
NIAPP	National Institute for Agricultural Planning and Projection
NL	Nested Logit
NPK	Nitrogen, Phosphorous, and Potash (Potassium)
NTFP	Non-Timber Forest Products

OLS	Ordinary Least Squares
P	Probability
PC	People's Committee
PDF	Probability Density Function
RHS	Right-hand Side
RS	Remote Sensing
SD	Standard Deviation
SMRP	Sustainable Management of Resources in the Lower Mekong Basin Project
SRV	Socialist Republic of Vietnam
t/ha	Tons per Hectare
TM	Thematic Mapper
TÖB	Tropenökologisches Begleitprogramm (Tropical Ecology Support Program)
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
USD	US Dollar
VICOFA	Vietnam Coffee and Cocoa Association
VLSS	Vietnam Living Standards Measurement Survey
WB	World Bank

CHAPTER 1

Introduction

1.1 Background

Since the end of the war in 1975 against the United States Vietnam has undergone dramatic changes in its economic environment. The communist government in the reunified Vietnam banned individualized ownership of land after the war and collectivized agricultural production activities under the rigid control of the state administration. Hundreds of thousands of people were resettled from Northern to Southern Vietnam to work in the newly established agricultural collective system in the decade following the war. After severe periods of food shortages and hunger in the beginning of the 80s the Vietnamese government initiated its economic renovation policy (*doi moi*), which shifted control over productive assets from collective to individual ownership. The introduction of the household responsibility system allowed rural households to exercise limited control over labor and land, thereby stipulating a reorganization of agricultural production. The shift of control in production processes accompanied by massive migration programs and rapid natural population growth induced agents to adapt their land-use decisions to the changing political and economic circumstances. Consequences can be found on various levels with multiple effects on socioeconomic conditions and environmental indicators. It is these political, demographic and economic transitions in Vietnam over the past

two and a half decades that are crucial for understanding the process of development in general, and land-use change in particular.

Between 1975 and 1990 forest cover in Vietnam decreased from 34% to 28% (Nguyen Manh Cuong, 1999). Almost five million hectares of forest were lost in Vietnam from 1943 to 1997. Forests covered 41% of the country's total land area in 1945 compared to 28% in 1993 (Do Dinh Sam, 1994). Population increased from 27 to 79 million between 1945 and 2001. At present, about three quarters of the population are engaged in agricultural production, thereby generating a massive demand for land. This demand will continue to rise as population is projected to increase to 123 million in 2030.¹ The fast growing rural population with population densities among the highest in South-East Asia puts high pressure on land use stemming from a need to expand and intensify agricultural production. At the same time, the Vietnamese government is attempting to restore forest cover to the level of 1943 with huge reforestation programs and changing tenure regimes like the allocation of forest land to individuals, groups and communities.

Dak Lak is the province with the most abundant forest resources in Vietnam. Population increased more than five-fold since liberation day in 1975 and doubled in the last decade. Increasing population creates a growing demand for agricultural land and forest products, which leads to a degradation of forest quality and a decrease in forest cover. A growing population along with suitable natural and economic conditions and an evolving market economy stimulated the commercialization of agriculture. The emergence of the large-scale cultivation of coffee is the most prominent example in this respect and led to a transformation of the rural economy in the recent history of Dak Lak. Coffee production increased from negligible amounts to 450,000 tons making Vietnam the second biggest world producer and inducing massive land-use change and widespread deforestation. Between 1975 and 2000 forest cover in Dak Lak declined by 25%.

The degradation of land and forest resources can have direct adverse impacts on the entire ecological infrastructure, the hydrological system, soil properties and on the local and regional climate. Against this background a research project was initiated from which this thesis emanated. The thesis attempts to assess the impacts of policy, technology, socioeconomic and geophysical conditions on land use and land-use change and its related consequences on economic growth, welfare, and the

¹Source: www.library.uu.nl/wesp/populstat/Asia/vietnamc.htm, accessed February, 2003.

protection of forests. The research project takes place in two districts of Dak Lak province in the Central Highlands of Vietnam. Dak Lak exhibits an interesting case in the study of land-use dynamics with its abundant forest resources, ethnic diversity, high immigration rates, and dynamic agricultural and socioeconomic development, particularly during the last decade.

1.2 Problem analysis

Land-use change, the physical change in land cover caused by human action on this land, is a common phenomenon associated with population growth, market development, technical and institutional innovation and policy action. Change in land use can have various consequences on natural resources like an alteration of the vegetation cover, a change in physical soil characteristics, in plant and animal populations, and effects on hydrological externalities (Turner II et al., 1995; Lambin et al., 1999; Aylward, 2000). Land cover and land-use change influences, and is affected by, climate change, loss of biodiversity and the sustainability of human-environment interactions and is seen as an essential component in all sustainability considerations (Lambin et al., 1999). Land-use change leads to a change in, and is affected by, socioeconomic indicators such as rural income, agricultural productivity, wealth and education. A better understanding of the complex interactions of these changes over time should enable decision makers at regional and national level to formulate regionally adapted policy interventions, which stimulate benefits and counteract negative consequences by simultaneously considering the trade-offs among economic, environmental, and social objectives in the process of sustainable rural development.

Patterns and processes of land-cover and land-use change are essentially spatial processes. Spatial data analysis gives valuable insights into processes of land-cover and land-use change and their underlying causes. It can facilitate spatial planning and the analysis of influences on land use stemming from socioeconomic, agroecological, geophysical and policy variables.

Household and village level statistical analysis of farm practices and land use in developing countries have long suffered from having to use poor proxy variables to represent spatially related factors. Distance or time to market outlets or service providers are often based on recall data and crude judgments from survey questionnaires. Geophysical variables like slope are often averaged per farm or village.

Agroecological zones are generally only roughly differentiated. All other spatially related effects are usually captured by simple location dummy variables. The coarse nature of these proxies has limited the ability of researchers to quantify the directions and magnitude of effects resulting from spatially explicit factors and largely restricted conclusions to non-spatial interpretations, assessments and outcomes.

Conversely, geographic information system (GIS)-based analysis typically ignores the socioeconomic and other household-specific factors that may vary markedly even within a relatively small zone of analysis. The practical reason is that spatial data for these variables cannot be easily obtained. Wealth indicators and education levels, access to credit and government services as well as household size are a few examples of a number of variables, which can be of critical importance in determining the choices and constraints of farmers. Analysis that rely heavily on remotely sensed data may poorly predict outcomes in which an individual household's socioeconomic characteristics and available resources play an important role. A combination of spatially explicit geophysical variables, measured at a pixel-level with a high spatial resolution, with socioeconomic characteristics of individuals, households or communities, measured usually for certain administrative boundaries or interpolated from georeferenced locations, is expected to yield further insights into dynamics and patterns of land-use change.

In this dissertation, novel methods for combining data from a village-level survey with remote sensing data derived from Landsat satellite images and GIS data are developed. By applying GIS techniques the location and spatial characteristics of objects are considered and analyzed explicitly, in that way providing additional insights compared to conventional survey research and data analysis on land-use aspects. Econometric analysis is carried out using an ex-post research design and a reduced-form, spatially referenced multinomial logit model. The multinomial logit models estimate the direction and strength of various exogenous and predetermined driving forces on changes in land cover. Changing right-hand side variables enables the simulation of the effects of potential policy interventions on land-cover change including an assessment of the spatial consequences and a possible identification of hotspots.

1.3 Objectives

To support the sustainable development of upland watersheds the research project aims to assess the driving forces of land-use change and the effect these changes have on the conditions of natural resources, human welfare and economic output in upland watersheds of Dak Lak Province. A better understanding of the complex interactions of these changes over time should assist decision makers at regional and national level in the formulation of regionally adapted policy interventions, which stimulate benefits and counteract negative consequences by simultaneously considering the trade-offs among economic, environmental and social objectives in the process of sustainable rural development.

The objectives of this dissertation are to assess the exogenous driving forces of land-use change in a spatially explicit way. Land-cover data derived from satellite image interpretation is linked with data from a survey in randomly selected villages to test several hypotheses of land-use change and to identify driving forces and characteristic processes of land-use change. Specifically, the aim is to assess the land-use responses of local agents to increasing population densities, technological change and improved market access. As land-use responses the two options of agricultural expansion and intensification will be examined in detail. Improved information of processes and patterns of land-use changes combined with spatially explicit data on relevant indicators and outcomes should provide decision makers with a better basis for land use planning, land allocation and the implementation of various other rural development and resource management policies.

1.4 Outline

Chapter 2 outlines the theoretical background of induced technical and institutional innovation and elucidates the conceptual framework with respect to land-use changes induced by local agents. Several main hypotheses related to the conceptual framework are stated and adapted to the prevailing conditions in the research area. Combining the models of Ricardo and von Thünen introduces the concept of space and an explanation for the spatial variation in land use and, in that way, adds a spatial perspective to the conceptual framework. An overview of previous applications of spatially explicit models of land-use change follows with special attention to spatial regression models, the approach employed for data analysis in this thesis.

The methodology employed for econometric estimation is presented in chapter 3. As a consequence of the different data sources employed, innovative techniques were developed to integrate data from the spatially explicit sources and from the village survey. To estimate the spatially explicit data, techniques of spatial econometrics are necessary, which account for the peculiarities of considering two dimensions in X and Y direction. Based on this, the chosen methodology to correct for spatial effects in the limited dependent variable framework is outlined and conventional and spatially explicit assessments of predictive power are described. The illustration of model details is followed by the data collection procedures with respect to the derivation of the land-cover data from satellite images, the collection and manipulation of GIS-data, and the socioeconomic village survey.

Chapter 4 outlines the descriptive statistics of primary and secondary data for Dak Lak province and, in more detail, for the research area. This chapter aims to provide a characterization of the natural and economic situation in the research area over the last two decades. The focus is on exogenous data employed in the econometric estimation, but also includes various endogenous indicators. Major trends in agricultural and socioeconomic development and in the change in environmental indicators should enable the reader to obtain a historical picture of the changes since 1975 and an idea about the present situation.

The empirical results of the multinomial logit estimations are presented in chapter 5. One model is estimated for the period between 1975 and 1992 and one for 1992 to 2000. The results from the econometric models are discussed and relevant changes between the first and second period highlighted. The chapter further investigates model statistics, measures for goodness-of-fit and predictive accuracy for both models and concludes with allusions to model shortcomings.

In chapter 6 four potential policy scenarios of rural development interventions are simulated. The approach taken in this study provides the possibility to mimic the consequences of policy options for land use in a spatially explicit way. Apart from model statistics, the prediction maps of the simulated changes allow for a spatial assessment of expected land-cover changes and indicate potential hotspots of land-cover conversions as a result of changing policies given the data at hand. Conclusions and policy recommendations in chapter 7 summarize the policy implications derived from the descriptive analysis and the spatially explicit econometric model and contain a discussion of results and their policy implications. Possibilities

are mentioned for additional analyses using a spatially explicit data set as the one employed in this study and references are made concerning future research needs.

CHAPTER 2

Theoretical background

This chapter outlines the theoretical background including the theory of induced innovation to explain processes of land-cover and land-use change. The theoretical implications and the possible effect of choice variables, decision parameters and exogenous macro-level influences on land-use change are discussed. To connect the driving forces of land-use change to their related outcomes, several hypotheses are formulated based on the conceptual framework. As a theoretical foundation for the analysis of land-use change and the introduction of the concept of space into the economic theory of land use, the land rent model of von Thünen including its assumptions and implications will be explored in detail. The combination of the von Thünen model with theories from David Ricardo provides a consistent theoretical basis to explicitly consider the location of objects in the econometric estimation. Several spatially explicit empirical studies are discussed with a focus on spatial regression models, the approach taken in this thesis.

2.1 Conceptual framework

[Land cover](#) and [land use](#) are two linked components, which for a long time have been studied separately. Land cover as the physical state of the land, was primarily a concern of natural scientists while land use as the human employment of land, was predominantly the focus of social scientists ([Turner II & Meyer, 1994](#)). *Land-*

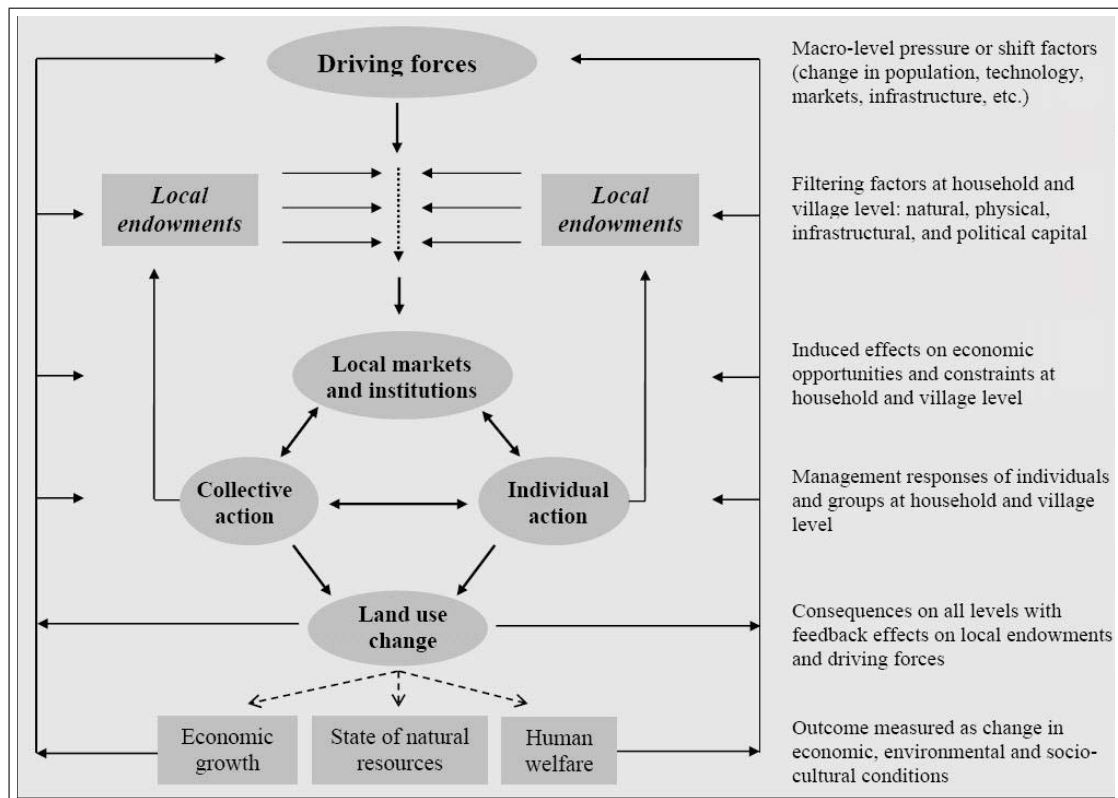
cover change is related to conversions and modifications of existing land cover, where the former is a change from one class to another while the latter a change of the conditions within one class (Turner II & Meyer, 1994). *Land-use change* includes either a change of an existing land-use category or a change in the intensity of an existing land use (Turner II & Meyer, 1994).

The conceptual framework employed to study land-cover and land-use change draws heavily from the theory of induced innovation (Boserup, 1965; Ruthenberg, 1980; Ruttan & Hayami, 1984). Two key processes of land-use change can be distinguished as a source for increasing agricultural production: *First*, agricultural expansion into previously uncultivated areas, which usually takes place at an extensive and constant technological level; and, *second*, agricultural intensification on already cultivated land. Intensification involves the substitution of labor-intensive and capital-intensive inputs for land-intensive production techniques by means of technological changes such as the introduction of better irrigation facilities, of improved crop varieties and fertilizer.

Agricultural expansion and intensification might coexist and occur simultaneously. If agricultural production gets more profitable, e.g. through increases in agricultural output prices, farmers have an incentive to expand the area under cultivation given the availability of suitable land and labor. This may result in forest clearing. Changes in the relation between agricultural output prices might have a similar effect as different crops exert different pressures on farmers to expand their area under cultivation (Kaimowitz & Angelsen, 1998). Profitability of agriculture is also influenced by changes in input prices. An increase in fertilizer prices makes farming less profitable and can reduce land devoted to agricultural production. On the other hand, it might invoke farmers to substitute land for fertilizer, in that way leading to forest clearing. Improvements in labor-saving technologies can equally induce farmers to expand agricultural area as it relaxes labor constraints. Boserup (1965) and Ruthenberg (1980) emphasize the responses of households, communities, and societies to pressures and opportunities induced by *exogenous* shift factors such as population growth, improvement of rural road networks, and investment in markets and the agribusiness sector, thereby enabling the introduction of higher-value perennial crops. Other studies suggest the impact of population pressure to be relatively small (Pender et al., 1999, 2001a). In the research area, additional pressures on land-use change stem from the introduction of protected areas during the

1990s and from a more effective enforcement of state policies discouraging [shifting cultivation](#).

Figure 2.1: Conceptual framework of land-use change



Source: Adopted from [Scherr et al. \(1996\)](#) and [Kaimowitz & Angelsen \(1998\)](#)

Driving forces at the macro-level induce a pressure for change on local agents (see the conceptual framework of land-use change in figure 2.1 with associated variables at the same height on the right side). In the research area the agents of land-use change are primarily individual farmers. State companies involved in agricultural or forest activities like state forest and state coffee enterprises had little influence. The conceptual framework in figure 2.1 does not consider dynamics explicitly, but can be extended for a dynamic analysis of several periods by subscribing the elements with the time periods t_1, t_2, \dots, t_n (see e.g. [Scherr et al., 1996](#)). Driving forces include population growth, altering input and output prices, technological change and infrastructural improvements such as road upgradings. As underlying factors, these exogenous or predetermined driving forces stimulate a transformation in the

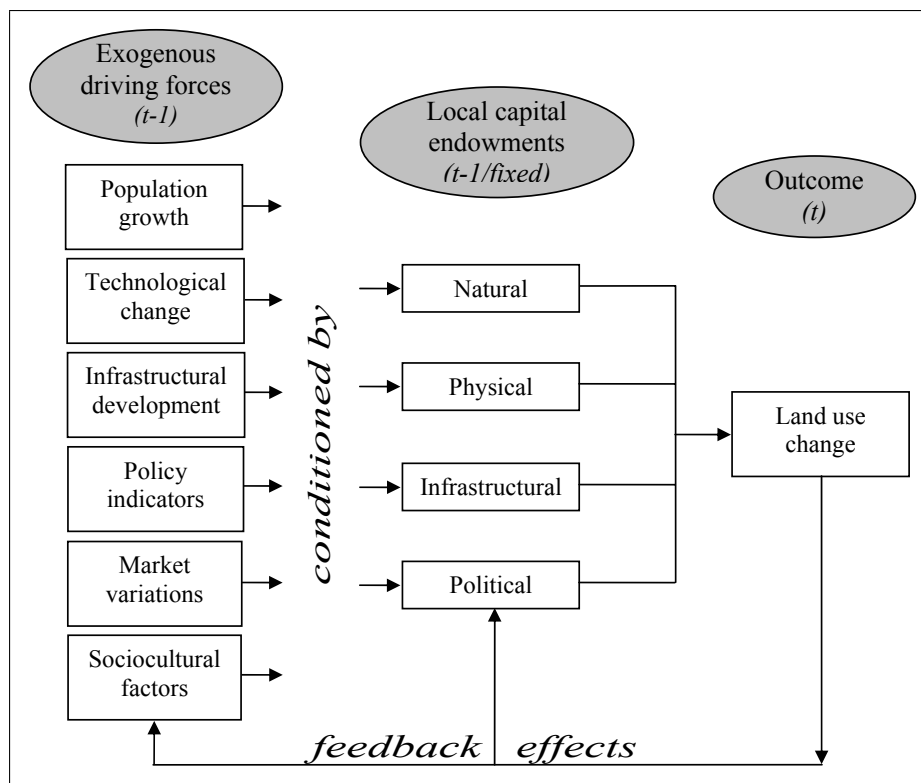
characteristics of agents and their decision parameters. Village institutions and organizations might evolve in a characteristic way as a response to the changing environment (Ruttan & Hayami, 1984). This will in turn provoke a change in the variables on which individual or collective management decisions are grounded, such as the choices of farmers concerning the allocation of production factors and consumption patterns. A consequence might be land-use change in its various forms, measured as the physical change from one major land-cover class to another on a given plot and during a specified period. It is this consequence, which is the focus of this thesis. Land-use change may result in changing levels of human welfare, can alter natural resource conditions and affect indicators of economic growth. Consequently, agents will react to a pressure for change and adjust their decisions according to the new circumstances (Scherr et al., 1996; Templeton & Scherr, 1997; Kaimowitz & Angelsen, 1998; Pender et al., 1999). Therefore, the processes described in figure 2.1 are dynamic and various feedback effects have to be considered, represented by the arrows in the figure.

This thesis attempts to model the outcome in figure 2.1 (land-use change) as a function of various exogenous driving forces and local endowments with spatially explicit variables on natural, physical, infrastructural and political capital. The center of figure 2.1 with local markets and institutions and with collective and individual action is not considered in the reduced-form models of chapter 5 and 6, which only consider exogenous and predetermined indicators as right-hand side (RHS) variables. On empirical grounds, only land-use change is considered as the single outcome in the econometric model. Endogenous outcomes of land-use change, represented in figure 2.1 by the dashed lines, will be investigated in the descriptive analysis of chapter 4, where changes in human welfare, the state of natural resources and indicators for economic growth are accounted for. The change in forest cover is one exception as an outcome for the state of the natural resources and is included in the econometric model via the dependent variable land cover.

Changes in land use of local agents are induced by the exogenous driving forces at the macro and regional level (Scherr et al., 1996; Templeton & Scherr, 1997; Pender et al., 1999). Along with the exogenous switch from a planned rural economy to a household responsibility system in the early 1990s in Vietnam, changes in prices, transaction costs, and the availability of technologies induced private farmers and entrepreneurs to adjust their land-use patterns rapidly, as can be witnessed in

Dak Lak Province. Moreover, the Vietnamese policy transformation towards more market orientation was supported by significant public investments in the rural infrastructure and in agricultural extension services, which promoted the adoption of new technologies, crops and animal husbandry systems. As an indication for the scale of improvement in the rural infrastructure, the average distance from a village to an all-year road in the research area decreased from 18 to six kilometers in the last two decades under consideration. Apart from policy and socioeconomic factors, changes in the amount of agricultural and forested area are determined by a range of mainly time-invariant natural conditions (Ruthenberg, 1980) such as rainfall, topography, soil quality, and other geophysical and agroecological variables.

Figure 2.2: Conceptual framework for econometric models



Source: Adopted from Scherr et al. (1996)

Figure 2.2 adapts the conceptual framework from figure 2.1 to the model employed for data analysis and to the peculiarities caused by the spatially explicit framework. Following this tailored conceptual framework in figure 2.2 and taking into account recent advances in the literature on spatial analysis, exogenous driv-

ing forces, conditioned by village level endowments induce an economic reaction of local agents, manifested as changes in land use. Accordingly, land-use change is explained as a function of endogenous (and therefore lagged) population growth, the introduction of new agricultural technologies and crops, operational variables for measuring the spatial placement of policy-induced investments in roads and market infrastructures, exogenous policy indicators, and variations in market prices proxied by transportation costs. These exogenous driving forces are filtered by local agents based on their endowments with natural capital (level and variance of rainfall, soil suitability, altitude and slope of land), socioeconomic village characteristics, education, road access as well as variables representing the investment into irrigation. The causal relationships between these variables will be examined in a spatially explicit framework and their respective direction and magnitude quantified (see chapter 5).

2.2 Hypotheses of land-use change

Population growth induces agricultural intensification by increasing the scarcity of land relative to labor through a shortening of fallow cycles and an increase in labor input per unit of land. Growing pressure on land induces a shift to higher-value crops on existing fields and an expansion of agriculture into more fragile, marginal areas (Boserup, 1965). On the other hand, as the scarcity of land increases, it gets more valuable, which in turn may lead to enhanced resource management practices and higher land investment on lands where secure long-term property rights exist and, therefore, an ownership effect for its users is created (Tiffen et al., 1994; Scherr & Hazell, 1994). In the particular case of Dak Lak there are two mechanisms of population growth: immigration, both controlled and uncontrolled, and natural growth. The two processes might be very distinctive in its effects on land-use change and its outcomes. Gradual natural growth may lead to a gradual expansion of agricultural land or to a gradual intensification of agricultural production on existing plots. The impact of immigration might be very different, as migrating people typically have to bring new land under agricultural production and, in many cases, establish new settlements. As a consequence, immigration can lead to a much higher degree of agricultural expansion and to a more rapid degradation of resources. Immigration schemes controlled by the Vietnamese government attempted to resettle migrants to more favorable areas with suitable natural conditions. Uncontrolled migration, on the other hand, generated a variety of settlements in more remote and inaccessible

areas, as can also be observed in the research area. If agricultural expansion proceeds into more fragile areas, which are more susceptible to resource degradation, the adverse effect on the state of natural resources might exacerbate.

Besides a high variation between villages, the complex processes of population growth in Dak Lak also lead to a high variability of agricultural activities within the villages. Increasing internal differentiation between and within villages might be followed by social inequity and conflicts. Increasing income differentiation might also hamper collective action to establish local level institutions for natural resource management or agricultural production.

Roads and improved *market access*, e.g. through infrastructure development, often increase the intensity of agricultural input use and the productivity of agriculture and reduces the risk associated with investments in agricultural production (Lee et al., 2001). In that way, better access to market can reduce the need for land expansion (Reardon et al., 2001). If roads open up formerly inaccessible areas, colonization pressures can result in environmental degradation (Lee et al., 2001).

Technological development, e.g. new management practices or improved crop varieties, may promote a shift to cash crops and increase farm income, if prices are not falling at the same time (Pender et al., 1999). Yield-increasing technological progress provides opportunities to increase agricultural production on the same amount of land, thereby reducing the pressure on forests. More production of cash crops might in turn boost the demand for irrigation water, e.g. for coffee production, thereby affecting hydrological conditions. Effects on soil conditions may be ambiguous depending on the techniques applied (use of cover crops or shade trees) and the crops planted (coffee, cocoa, fruit trees, rubber). Moreover, technological innovations may increase the demand for agricultural land and create an incentive to further deplete natural forest resources (Baland & Platteau, 1998). Distinct results might emerge at different levels of labor intensity of technological progress. Labor-intensive progress tend to limit the amount of land under cultivation given household's labor constraints while labor-saving technologies release resources, which can be used to expand land under cultivation (Angelsen & Kaimowitz, 2001).

In Vietnam, the *State* is playing a continuing and important role in production and marketing of cash crops. State enterprises have quasi-monopolies, e.g. in coffee and rubber production, and provide farmers with subsidized inputs and purchase their products at a fixed price. Consequently, the mechanisms and outcomes might

be different for farmers enrolled in state-controlled cash crop or in food crop production. As the State has a direct influence on land-use change in certain areas of Dak Lak, one might refer to this phenomenon as directed or *imposed change* rather than induced change (Burmeister, 1987). The outcomes of these imposed changes are possibly different from the outcomes of induced changes. The promotion of plantation agriculture on better quality soils could force subsistence farmers out of these lands and lead to an expansion of land allocated to the production of food crops in more marginal areas, thereby leading to more rapid resource degradation. Additional income from cash crop production could, on the other hand, enable farmers to invest in sustainable, long-term resource use.

2.3 The land rent model

In 1817 David Ricardo (Ricardo, 2002) assumed a heterogeneous landscape where land quality influences land use and farmers cultivating on better land receive a rent, which Ricardo termed the *land rent*. Assuming a heterogeneous landscape, Ricardo did, however, not consider the relative location of production activities. The German farmer and economist Johann Heinrich von Thünen was the first one to explicitly consider the emergence of a certain spatial arrangement of land use. His pioneering book “Der isolierte Staat” (*The Isolated State*) was first published in 1826 (von Thünen, 1990). There, he emphasized the importance of transport costs for the development and the spatial arrangement of land-use structures around market locations. In that way, von Thünen introduced the notion of space to economic analysis in general and to the analysis of the spatial patterns of land use in particular. Von Thünen’s work is based on following assumptions (Briassoulis, 2000):

- ▷ The market is located centrally within an Isolated State, which is self-sufficient and exists without any external influences.
- ▷ The Isolated State is surrounded by an unoccupied wilderness.
- ▷ The land of the Isolated State is a uniform flat plain, equally traversable in every direction.
- ▷ The soil quality and climate do not vary throughout the Isolated State.
- ▷ Farmers in the Isolated State transport their own goods via oxcart directly to the market across the land, therefore, abstracting from roads and rivers.

- ▷ Farmers maximize their profits.

According to the above constraints von Thünen assumes the landscape as a flat, uniform plain, where movements in all directions are equally comfortable. **Absolute distances** (e.g. kilometers) therefore equal **relative distances** (e.g. travel time). Transport costs in his model are a linear function of distance without consideration of the amount of this distance and the means of transport. Von Thünen further abstracts from any variations in the productive potential of the land, thereby assuming that every hectare would produce the same output with the same amount of input. The market place in the center of the *Isolated State* has a monopoly and receives all goods of all surrounding farmers. The farmers are price takers of the stable market price, which can neither be influenced by individual nor by collective action. Von Thünen also abstracts from any outside influence and sees production merely as a reaction to economic conditions.

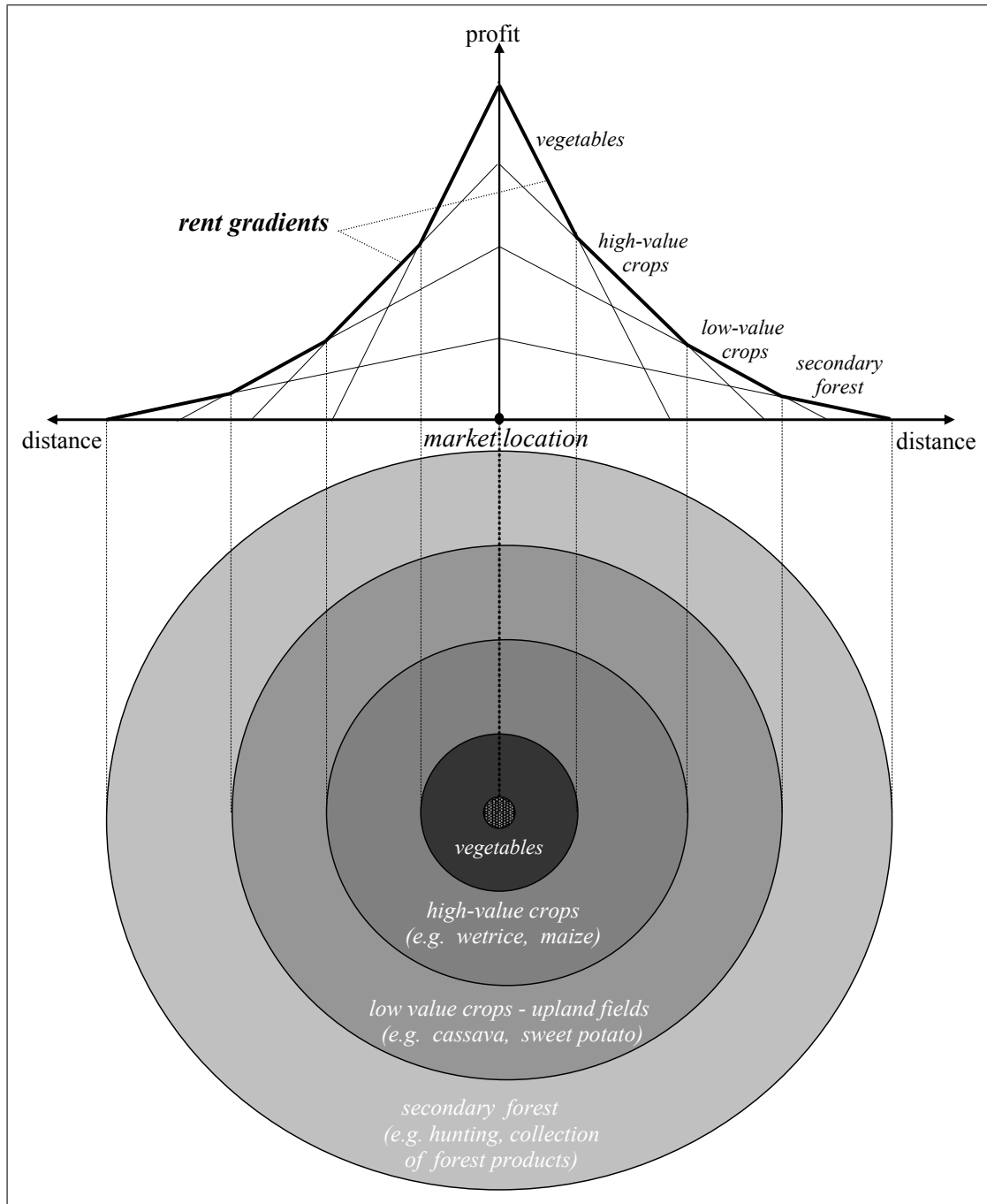
Farmers in a *von Thünen world* have perfect information and perfectly use this information to maximize profits. Summarizing the von Thünen world, each piece of land will be devoted to the use in which it would yield the highest rent. Under his imposed conditions, the various land uses are expected to occupy a series of concentric rings surrounding the market location (see figure 2.3). Von Thünen was the first scholar to introduce a theory for the spatial arrangement of agricultural patterns based on the distance of land to the central market, which determines the value of the land or the land rent. The land rent in turn leads to a certain spatial structure of land use and provides economists with a model to understand the spread of economic activities around a central market, or, as **Krugman** calls it, the “*centrifugal forces*” (1995). The establishment of this land rent theory earned von Thünen the title of the first *truly modern economist*. Economic geography and regional science claim von Thünen as one of the fathers of their discipline. He is also considered the founder of *marginal analysis* in neoclassical economics for the establishment of a theory of diminishing returns for the marginal productivity of labor using calculus (von Thünen, 1960, cited in Samuelson, 1983). His revolutionary equation for this marginal product of labor is carved into his tombstone in Belitz, Germany.

Land users are willing to pay or *bid* for the use of land. If each activity’s *bid rent* is displayed for each location over a specified area, a bid rent surface emerges with highest rents as peaks and the lowest as holes. Drawing a cross-section of the

rent surface along a line returns the so-called *rent gradient*, which is a measure for the relative advantage one plot has over another. Figure 2.3 shows an example for land use allocation in von Thünen's Isolated State adapted to the research area of this study. Simplifying actual land use, the rent gradients could be separated by the production of vegetables and other perishable products with high transport costs and yielding a high value per hectare in the inner ring next to the central market. Moving away from the market, the cultivation of intensive high-value annual crops like paddy and maize might prevail in the next concentric ring, followed by more extensive lower-value crops, usually planted on upland fields. The outermost ring could be reserved to land-intensive activities and the use of secondary forest for activities like hunting and the collection of timber and non-timber forest products. Further away, where transport costs exceed the expected profits, land would be covered by primary forest or other natural vegetation not influenced by human activities. The land uses in figure 2.3 are graphed as concentric rings around the corresponding market location, assuming transport costs, which incur all costs of the movements of goods over space, are a linear function of distance and therefore determine the net income derived from farming activities.

Ricardian models of land use explain the existence of different land rents with differences in land quality that arise from a heterogeneous landscape. Land of better quality or higher soil fertility generates higher rents. Land with higher quality generates surpluses for farmers compared to farmers with land of lower quality. Von Thünen found an explanation for the emergence of land use patterns and differing land prices over space as a function of distance from urban centers in his "featureless plain". These two theories are the basis for most economic models of land-use change. Combining the two theories by integrating the inherent features of plots (Ricardo) with distance measures (von Thünen) and relaxing some additional assumptions provides a consistent economic theory to explain land-use changes in a spatially explicit manner.

Figure 2.3: Land rent gradients and land use zones



Source: Author, following von Thünen (1990)

2.4 Spatially explicit models of land-use change

Most economic phenomena are essentially spatial processes ([van der Veen & Otter, 2001](#)). They not only vary over time, but also over space. Economists usually utilize economic subjects like households, villages or countries as their basic research unit, which leaves them working at coarse scales due to spatial aggregation. Natural scientists, on the other hand, use natural units – and here frequently continuous surfaces – to study processes of their interest. They do rarely base their scientific work on administrative or economic units.

The introduction of the concept of space presented in the previous section provides the theoretical foundation to apply new economic approaches to model determinants of land use. A key feature of these models is that all [data](#) elements contain locational attributes, require special techniques for the estimation and yield location-specific results. The majority of these economic models on land-use change link survey or census data to spatially explicit data, usually generated by means of [Geographical Information Systems](#) (GIS), in order to estimate the effects of exogenous factors on land-cover change. Tropical deforestation plays a dominant role in these models, which describe land use as a binary choice, i.e. deforested or not. [Kaimowitz & Angelsen \(1998\)](#) review some of these models and separated modeling approaches according to the level, i.e. household/firm, regional, national and macro-level models. Model types include analytical models, regressions, simulations and computable general equilibrium (CGE) models. Within these groups they consider spatial regression and spatial simulation models as examples for spatially explicit models. In this study the discussion focuses on spatially explicit regression models as this family of models reflects the estimation approach selected for data analysis in [chapter 5](#) and [6](#).

Founded on the work of von Thünen, [Chomitz & Gray \(1996\)](#) developed a widely cited empirical model based on survey and satellite data from Belize, which allows to study land-use change in a spatially explicit way. They assume that land will be devoted to the activity yielding the highest rent and assess the effect of roads on land use by employing the land rent model based on theories developed by Ricardo and von Thünen (see [section 2.3](#)). The model is estimated using land cover data derived from satellite images, spatially explicit geophysical variables, distance measures and dummy variables for the tenure status of the land. Road existence, road quality and market infrastructure are important factors affecting land rent as they

influence transport costs and therefore farm gate input and output prices. Other factors affecting land rent are soil quality and distance to local and regional markets (Chomitz & Gray, 1996). Land rent in their model is defined by future returns (vector of output prices P times output quantity Q) minus costs (vector of input prices C times input quantity W) for each location i and at each time T . δ denotes the discount rate at each location i . A similar model is employed by, among others, Nelson & Hellerstein (1997); Munroe et al. (2001); Deininger & Minten (2002). Formally, the net present value R of the land rent function is given by:

$$R_{ikT} = \int_0^1 (P_{ikT+t}Q_{ikT+t} - C_{ikT+t}W_{ikT+t})e^{-\delta t} \quad (2.1)$$

Household-level statistical analysis of farm practices in developing countries have long suffered from having to use poor proxy variables to represent spatially related factors. Geophysical variables like altitude and slope are often averaged per plot, farm or village. Agroecological zones are generally only roughly differentiated. All other spatially related effects are usually captured by simple dummy variables for the location of objects without a consideration of spatial effects. The crude nature of these proxies has limited the ability of researchers to quantify the directions and magnitude of the effects resulting from spatially explicit factors on farmer's choice of technology and on their allocation of resources. Examples of studies employing empirical estimations without explicit incorporation of the locations of objects include Southgate et al. (1991), Elnagheeb & Bromley (1994), Godoy et al. (1997), and Bergeron & Pender (1999).

2.4.1 Applications to forestry

Deininger & Minten (2002) employ satellite imagery data and municipio-level census information aggregated to grids of one square kilometer to estimate a probit model for deforestation in Mexico. A dummy for protected areas reveals that pixels closer to roads had a significantly higher chance to remain forested if they were part of a protected area. A similar econometric approach using a binary logit specification was applied to a smaller area with higher resolution data by Geoghegan et al. (2001). Pfaff (1997) merges remotely sensed data on land cover, which he aggregated to municipio-level, with socioeconomic data to address determinants of deforestation in the Brazilian Amazon. His multivariate analysis shows that increasing road

density has a strong positive effect on deforestation in one county and in neighboring counties. Larger distance from markets south of the Amazon leads to less and better soil quality increases deforestation. Pfaff notes that the positive effect of population density on deforestation was only evident when population was the sole variable. The population effect disappeared after the inclusion of other relevant indicators from his land-use model. Mertens et al. (2000) integrate a household survey and remotely sensed land cover data for Southern Cameroon.¹ They define boundaries of agricultural areas by estimating the spatial extent of villages into the forest and along the road network taking village population into account. These estimated village boundaries they use as the base unit to merge household and land cover data. Results of their statistical analysis indicate that higher deforestation rates are related to population growth, increased marketing of food crops, and agricultural expansion.

Vance & Geoghegan (2002) use survival analysis in a spatially explicit model to analyze the effect of explanatory variables at the household level on the probability of deforestation in southern Mexico. They employed sketch maps to link plot level data to Landsat Thematic Mapper (TM) images for two time periods in each of their four zones of analysis. Apart from several household level socioeconomic determinants they find the probability of deforestation to be non-linear over time for the duration of a household's occupancy of a plot. The probability of deforestation is first decreasing then increasing in their specified quadratic relationship as a result of the rotation periods, family life cycles and learning effects. Munroe et al. (2002) employ a random effects probit in panel formulation to estimate the probability of deforestation. Significant influences on forest cover changes they found for output price relations, infrastructural development and topographic variables.

2.4.2 Applications to agriculture

Fox et al. (1994) published one of the first models known to the author, which employs techniques of spatial econometrics applied to the study of land use. They analyze the relationship between farmer's characteristics and crop choice on plot level. They combine remote sensing (RS) data with a household census and used generalized least squares, multinomial logit and semivariogram analysis. In addition, they attempt to correct for spatial autocorrelation using dummy variables to indicate

¹Their article does not explicitly state which particular regression model they used.

the location of their study villages. Later, these “direction dummies” were removed, because of high collinearity with the distribution of ethnic groups. Household labor endowment, work force and economic status are major influencing variables in their plot-level spatial framework. They suggest that an improved transportation network encourages farmer to switch to more intensive farming techniques. [Chomitz & Gray \(1996\)](#) assess the effect of roads on land use and employ a land rent model based on the theories from Ricardo and von Thünen. They estimate a multinomial logit model with 11,000 observations of rectangular grid cells with a size of one square kilometer as base unit. Their findings suggest that market access, quality of land and tenure status affect the probability of agricultural land use with different effects on the likelihood of commercial versus semi-subsistence farming. Road construction in areas with low agricultural potential and low population densities may lead to habitat fragmentation and low economic returns. [Nelson & Hellerstein \(1997\)](#) estimate a similar model for Mexico including corrections for spatial dependence. They identify locations, which are more difficult to access, to be more prone to deforestation. Both models rely on satellite data to derive land cover and employ georeferenced geophysical data and limited secondary information for socioeconomic characteristics as independent variables.

[Serneels & Lambin \(2001\)](#) derive driving forces of land-use change within a cell-based modeling framework applied to a rural district in Kenya. The multiple logistic regressions reveal distance to markets and agricultural potential as most important explanatory variables. They note that the perception of land rents differs between smallholder and large-scale entrepreneurs. [Nelson et al. \(2001b\)](#) simulate the effects of a reduction in transport costs on land use. Road resurfacing did not have a strong effect on overall deforestation, however it decreased plywood forests and increased areas under pasture. Dummy variables for park and concession areas and a transportation cost surface serve as further socioeconomic explanatory variables. In a further article [Nelson et al. \(2001a\)](#) simulate the elimination of legal restrictions on land use and found that property rights indeed influence people’s land use.

2.5 Summary

This chapter introduced a conceptual framework to study land-use change. The framework draws from the theories of induced technical and institutional innovation and provides a concept to explain processes of the expansion of agricultural land

and agricultural intensification. The economic theories of Ricardo and von Thünen are outlined, which allow for the introduction of a heterogeneous landscape and the spatial variation in the value of land. In that way, the concept of space is introduced into the model. The combination of the conceptual framework with an explicit consideration of space provides the basis for the econometric model presented in the next chapter. The literature review at the end of this chapter exemplifies previous spatially explicit empirical applications for deforestation and agricultural land use.

Methodology for Data Collection and Econometric Modeling

This chapter outlines the modeling approach taken in this study. The mathematical background of limited dependent variable models and the econometric model employed for empirical analysis are presented. The econometric peculiarities of spatially explicit data analysis, which stem from the introduction of spatial effects, i.e. effects over space in X and Y direction like spatial dependence, are discussed and possibilities are mentioned to correct for these effects in a limited dependent variable framework. Empirical data collection by means of a satellite image interpretation, a village survey, interpolation techniques and distance measures is outlined.

This study utilizes a novel approach to combine these data sources using village accessibility catchments. The derivation of the accessibility catchments are outlined and several aspects of spatial econometrics relevant to the analysis are highlighted. Regressions are estimated for four types of regressors. The expected effects of the employed regressors are discussed for the geophysical data, the characteristics of the socioeconomic history of the villages and for policy-related variables.

3.1 Econometric approaches

In the case of land use models, researchers usually deal with dependent variables with several unordered discrete states representing the land-cover categories. These states are usually labeled with sequential numbers. The limited number of the ob-

served states of the dependent variables is related to an unobserved or *latent variable* (Anselin, 2002). This latent variable has an underlying propensity to be classified into one of the discrete states of land cover, either by the image interpreter or by the automatic classification procedure. Consequently, there is a certain distribution attached to each of the discrete states of land cover. For the binary case, the latent, continuous variable y^* can be linked to its binary representation by the equation (Long, 1997):

$$y_i = \begin{cases} 1 & \text{if } y_i^* > \tau \\ 0 & \text{if } y_i^* \leq \tau \end{cases} \quad (3.1)$$

The parameter τ in equation 3.1 represents a threshold and for observations of $y^* \leq \tau$ the observed binary variable y_i takes the value 0. If $y^* > \tau$, the dependent variable y_i is equal to 1. As the dependent variable y_i is unobserved, ordinary least squares estimation (OLS) is not appropriate, but instead maximum likelihood (ML) estimation has to be utilized (Long, 1997). ML estimation requires knowledge about the distribution of the error terms. If the error terms are assumed to be normally distributed, the *probit* model is used for a binary y_i . Under the assumption of a logistic distribution of the errors the *logit* model is applicable.

Models for categorical outcomes are nonlinear and nonlinearity has implications for the proper interpretation of these models. In a linear model, the marginal change (equal to the partial derivative or slope) is the same at all values of the explanatory variables. The distinguishing feature in model interpretation is that the effect of a given change in an independent variable is the same regardless of the value of that variable at the start of its change and regardless of the level of the other variables in the model (Long & Cheng, forthcoming).

In nonlinear models the effect of a change in a variable depends on the values of all variables in the model and is no longer simply equal to the parameters in the model. Nonlinear models like probit and logit models can be estimated with ML. In ML the parameter values are obtained based on a likelihood function, which calculates the greatest likelihood of obtaining the observed sample (Long & Freese, 2001). Under the assumptions of identically and independently distributed disturbance terms and a certain underlying probability density function (PDF), the maximum likelihood estimator (MLE) is consistent, efficient and asymptotically normal. These desirable properties hold as the sample size approaches infinity (Greene, 1997).

The assumption of a distribution for the variance of the residuals, $Var(\epsilon)$, in latent variable models changes the values of the coefficients, but does not affect the computed value of the probability that a certain outcome occurs (Long, 1997). Changing the assumed variance for the error terms affect the spread of the distribution, but not the proportion of the distribution above or below the threshold. If a different distribution $Var(\epsilon)$ is assumed, the values of the coefficients change in a uniform way and only the magnitude of the coefficients is affected. The significance tests are similar since they are not affected by $Var(\epsilon)$ (Long & Freese, 2001).

In general, the marginal change and discrete change will not be equal. The two measures of change differ since the model is nonlinear and the rate of change is constantly changing. The discrete change measures the actual amount of change over a finite range in an independent variable while the marginal change assesses the instantaneous rate of change. The two measures will be similar when the discrete change occurs over a region of the probability curve that is roughly linear (Long, 1997).

3.1.1 Multinomial logit

To explore relationships between exogenous and predetermined variables and the land-cover categories as left-hand side (LHS) variables, the *multinomial logit* specification (MNL) was employed. Predetermined variables are treated as exogenous and, therefore, produce asymptotically consistent estimators (Greene, 1997). The estimation of a reduced-form models is employed to represent potentially endogenous relationships among the variables (Greene, 1997). MNL models estimate the direction and the impact of a set of explanatory variables on the categorical dependent variable by predicting a probability outcome associated with each category of the dependent variable. The probability that $Y = i$ can be stated as (Greene, 1997):

$$Prob(Y_i = i) = \frac{\exp \beta'_i x_i}{\sum_{j=1}^J \exp \beta'_j x_i} \quad (3.2)$$

which applies, if the error terms are independently and identically distributed as log Weibull (McFadden, 1973). In equation 3.2, J denotes the number of land-cover classes used for analysis, β_i is a vector of estimation parameters and x_i are the

exogenous variables for all Y_i and at all locations i .¹ Normalizing on all probabilities yields a log-odds ratio (Greene, 1997):

$$\ln \left[\frac{p_{ij}}{p_{ik}} \right] = x'_i(\beta_j - \beta_k) \quad (3.3)$$

The probability of obtaining land use j at location i (p_{ij}) relative to the probability of obtaining land use k at i (p_{ik}) is termed the *odds ration*, which does not depend on the other choices (Greene, 1997). The dependent variable is expressed as the log of the odds of one alternative relative to a comparison group. The comparison group² is the outcome category, for which the coefficients are arbitrarily set to zero. The coefficients for different base categories differ as they have a different interpretation, but the predicted probabilities stay the same since they are not affected by the coefficients. If model assumptions hold, the maximum likelihood estimators are consistent, efficient and asymptotically normally distributed for large samples (Long & Freese, 2001). The likelihood function of a multinomial logit model is the product of the probability for each observation of each discrete dependent category. It can be written as (Maddala, 1983):

$$L(\beta_1, \beta_2, \dots, \beta_{J-1}) = \prod_{i=1}^n p_{i1}^{y_{i1}} p_{i2}^{y_{i2}} \dots p_{iJ}^{y_{iJ}} \quad (3.4)$$

In equation 3.4, p_{ij} denotes the probabilities for each observation i and the $j = 1, 2, \dots, J - 1$ outcome categories to be estimated. Taking logs, equation 3.4 can be stated as the log-likelihood:

$$\log L = \sum_{i=1}^n \sum_{j=1}^J y_{ij} \log(p_{ij}) \quad (3.5)$$

McFadden (1973) proved that the log-likelihood in equation 3.5 is globally concave, which renders the maximization problem straightforward and ensures unique estimates for the ML function.

¹As all observations in this study are georeferenced, observations i are equal to locations i .

²The comparison group is also called the base alternative or the reference group.

3.1.1.1 Goodness-of-fit

Likelihood-ratio (LR) tests compare the logged likelihood (equation 3.5) from the full model with the one from a reduced model omitting explanatory variables. Formally, the hypothesis that x_i does not significantly affect the dependent variable is tested with (Long & Freese, 2001):

$$H_0 : \beta_{i,1|b} = \dots = \beta_{i,J|b} = 0 \quad (3.6)$$

In equation 3.6, b denotes the base category necessary to identify the model and $\beta_{i,b|b}$ is necessarily 0. The MNL estimates binary logit models for all possible combinations of outcome categories. Since estimating binary logits for all categories includes redundant information, $J - 1$ binary logits are estimated (Long & Freese, 2001). To test the hypothesis with $J - 1$ parameters, likelihood-ratio (LR) and Wald tests can be used. The empirical applications in chapter 5 and 6 report only LR test statistics as the two tests showed no major differences.

Significance of individual estimators can be tested with z-statistics. If model assumptions hold, the estimates produced by the MLE are asymptotically normally distributed as:

$$\hat{\beta}_k \approx N\left(\beta_k, \sigma_{\beta_k}^2\right) \quad (3.7)$$

The hypothesis that the estimated parameter β_k is equal to the true value of β^* ($H_0 = \beta_k = \beta^*$) can be tested with (Long & Freese, 2001):

$$z = \frac{\hat{\beta}_k - \beta^*}{\hat{\sigma}_{\hat{\beta}_k}} \quad (3.8)$$

If H_0 is true, then the z-statistics z is distributed approximately normal with a mean of zero and a variance of one for large samples. The z-statistics will be reported for the parameter estimates of the MNL models in chapter 5 and 6.

For polychotomous dependent variables the upper bound of R^2 is much less than one due to the characteristics of the likelihood function in equation 3.4. This likelihood function can only reach an upper bound of one, because the probabilities p_{ij} are smaller or equal to one. A perfectly fitted model would therefore result in an R^2 smaller than one (Maddala, 1983). Therefore other test statistics need to be employed to measure the goodness-of-fit of the model. Maddala, among others,

suggests as a *Pseudo* – R^2 measure the Cragg & Uhler’s R^2 to assess the goodness-of-fit of the full model:

$$R_{CU}^2 = \frac{1 - (L_I/L_F)^{2/N}}{1 - L_I} \quad (3.9)$$

In equation 3.9 L_I denotes the maximum of the log likelihood when maximized with respect to the intercept or constant term only. L_F is the ML upon convergence of the full model including all coefficients. In the empirical application, the likelihood-ratios of both L_F and L_I as well as the R_{CU}^2 are reported (see table A.5 in the appendix).

3.1.1.2 Independence of irrelevant alternatives

The (logged) odds ratio depends on the odds ratios of other alternatives and all odds sum up to one. It should be noticed that the odds-ratio in equation 3.3 is independent from other choices, because the disturbances in equation 3.2 are – by assumption – independently distributed. As a result, the relative probabilities among remaining outcomes are unaffected, if any alternatives are added or removed.³ This phenomenon is called the assumption of *independence of irrelevant alternatives* (IIA) and implies that no correlation between the error terms exists and that they are homoscedastic, i.e. they have constant variance. A violation of the assumption of IIA will lead to inefficient, but not to inconsistent parameters (Greene, 1997). Hausman’s specification test is a possibility to test for the validity of the assumption of IIA (Hausman & McFadden, 1984). McFadden et al. (1976) developed a likelihood ratio test to assess the IIA, which Small & Hsiao (1985) further developed.

The Hausman test of IIA estimates all coefficients for the full model $\widehat{\beta}_F$ including all categories. Then, a reduced model estimates a subset of coefficients ($\widehat{\beta}_S$) by eliminating one or more outcome categories. Under the H_0 hypothesis that IIA holds, $\widehat{\beta}_F - \widehat{\beta}_S = 0$, and $\widehat{\beta}$ is a consistent estimator. The Hausman test has degrees of freedom equal to the rows of the vector $\widehat{\beta}_S$. Small & Hsiao’s test divides the sample into two random sub-samples of about equal size. An unrestricted model is estimated on both samples. The weighted average of the resulting coefficients from both samples is computed. A restricted sample is taken from the second sub-sample by eliminating all cases with a chosen value of the dependent variable. After that,

³The classical example is the red bus/blue bus paradox (see Maddala, 1983, for a discussion).

the MNL is estimated using the restricted sample, yielding estimates and likelihood. Degrees of freedom are equal to $i + 1$ where i is the number of independent variables.

Both test are asymptotically distributed as χ^2 . The Hausman as well as the Small-Hsiao test often provide inconsistent estimates for a proper judgment of the IIA assumption (Long & Freese, 2001). According to McFadden (1973), the IIA implies that a MNL should only be used when outcome categories can plausibly be separated into independent categories by individual decision makers. In the empirical application only the results from the Hausman test will be presented in section 5.4 of chapter 5.

3.1.2 Spatial econometric aspects

Spatial econometrics incorporate the explicit consideration of spatial effects into the estimation of econometric models, because the presence of spatial effects inhibits the application of standard econometric techniques (Anselin, 1988). According to Anselin (1988), spatial econometrics is defined as:

“The collection of techniques that deal with the peculiarities caused by space in the statistical analysis of regional science models.”

Spatial econometrics therefore deals with various methodological differences that follow from the explicit consideration of spatial effects in economic models. Analytical and computational complexity arise from the explicit consideration of two dimensions, i.e. in X and Y (east-west and north-south) direction, in data and models. Anselin (1988) distinguishes between two main types of spatial effects, *spatial autocorrelation* and *spatial heterogeneity*.

Methodologically, *spatial autocorrelation* (or *spatial dependence*) is similar to the long known autocorrelation in time series models. Generally, the term *autocorrelation* expresses a lack of independence among observations (Anselin, 1988). Spatial autocorrelation is expressed in Tobler’s “First Law of Geography” (1979), as cited by Anselin (1988). It states that:

“Everything is related to everything else, but near things are more related than distant things.”

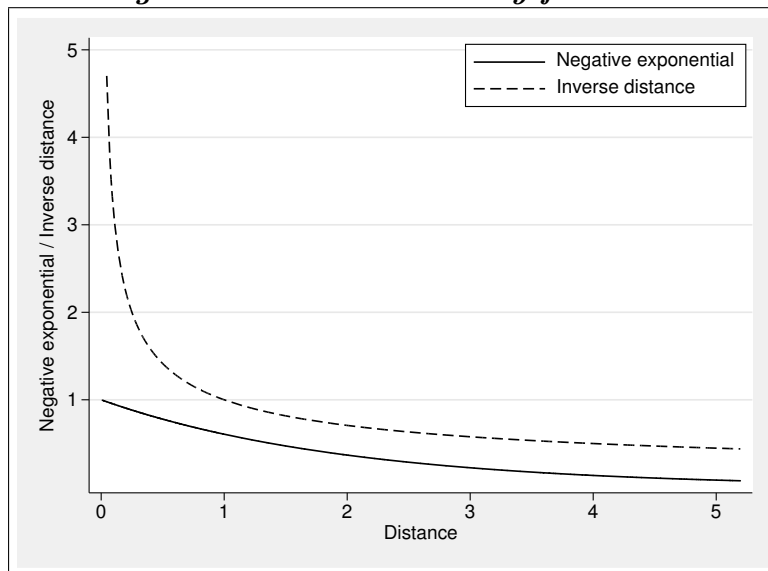
Spatial data is usually found to exhibit higher degrees of relationship, when observations are found close together than when points are located further away from each other. This implies that relative locations, i.e. the location of an observation

in relation to the location of its neighbors, are introduced including the variation of this relationship with increasing distance. Consequently, spatial autocorrelation can be expressed as a functional relationship between activities at one point in space and elsewhere (Anselin, 1988). Or formally:

$$y_i = f(y_j, y_k, \dots, y_n) \quad (3.10)$$

The value of the observed variable y at location i is (spatially) dependent on the value of y at the other locations j, k , until location n . A decreasing influence of neighboring values as the distance from y increases would be expected for most applications. That assumes that the further a neighboring location is away from the considered location the less influence it has. Commonly used functions to express this decay in influence at increasing distances between an origin i and all destinations j are for example the negative exponential ($f(d_{ij}) = e^{-\gamma d_{ij}}$) and the inverse distance function ($f(d_{ij}) = d_{ij}^{-\gamma}$) (Anselin, 2002). Figure 3.1 depicts these two distance decay functions for the example of $\gamma = 0.5$.

Figure 3.1: Distance decay functions



Source: Author, after Anselin (2002)

As for many applications in economics, geography and natural resource management the distance between objects and the location of objects relative to other

objects are of significant importance. Thus, spatial autocorrelation is a relevant phenomenon for a wide range of applications. Spatial autocorrelation can be formally expressed by the moment condition (Anselin, 2000):

$$\text{Cov}[y_i, y_j] = E[y_i, y_j] - E[y_i] \cdot E[y_j] \neq 0 \text{ for } i \neq j \quad (3.11)$$

In equation 3.11 i and j refer to individual observations (locations) and $y_{i(j)}$ is the value of a random variable of interest at that location, e.g. a spatial stochastic process yielding clusters of values in space. If covariances between the variables are not equal to zero they become meaningful, because it would imply a certain spatial structure or some kind of spatial interaction in the observations, which is of interest for the interpretation (Anselin, 2000). There are two distinct forms of spatial autocorrelation. First, *substantial spatial autocorrelation* in theory-driven models refers to a spatial relationship in the dependent variable, e.g. via spill-over effects, and potentially carries meaningful information. Second, in the presence of *nuisance spatial autocorrelation* the error terms are spatially correlated, which often arise in data-driven specifications and might stem from spatial aggregation or omitted variables (Anselin, 2002).

Spatial heterogeneity is present, if the units of observation are not homogeneous across space as is the case with unequal population densities or varying technological development (Anselin, 1988). Another source for spatial heterogeneity are effects varying over space like central place hierarchies and leading or lagging regions. The presence of spatial heterogeneity is usually considered by integrating varying parameters, random coefficients or switching regression approaches to account for the structural changes over space present in the data (Anselin, 1988).

3.1.2.1 Spatial scale

Often there is a mismatch between the scale of measurement of the observations under consideration. A scale mismatch and the need to integrate data from various sources tends to result in spatially dependent and spatially heterogeneous observations (Anselin, 2001b). Ignoring these aspects may lead to biased and inefficient estimates and wrong inferences. The application of spatial econometric techniques to correct for spatial effects therefore becomes necessary. In this study, the dependent variable land cover is derived from remotely sensed data and is computed as a surface for regular, square pixel cells of 50 meters on each side. The resulting grid does neither match the administrative units (commune and village territories)

nor the areal extent of underlying natural entities (e.g. watersheds). To deal with this fact cells can be aggregated to the usually larger administrative units, however, with a corresponding loss of variability inherent in the high-resolution spatial data. Smaller spatial scales (larger cells) lead to a less accurate estimate of the dependent variable. Furthermore, the variability in grid cells will not be constant across space, therefore, resulting in *heteroskedasticity* (Anselin, 2001b).

As Anselin (2001b) notes an interesting question is to which extent regressions at an aggregate level can identify the underlying model and to which extent they allow for proper inferences. Inferences about underlying micro-units rely on restrictive assumptions and inferences about them should be avoided (Anselin, 2001b).

3.1.2.2 Spatial interpolation

Little attention is paid to statistical implications of using predicted values as explanatory variables in techniques of *spatial interpolation* (Anselin, 2001b). Spatial interpolation is a technique to infer from point measurements on values for all locations, thereby “creating” data for previously “empty” regions.⁴ A standard spatial regression model has the form:

$$y_i = \beta x_i + \gamma z_i + \epsilon_i \quad (3.12)$$

In equation 3.12 y denotes the dependent variable, x_i and z_i the vectors of independent variables. β and γ represent the regression coefficients and ϵ the error term. Introducing a predicted value on the RHS in z_i will change equation 3.12 to (Anselin, 2001b):

$$y_i = \beta x_i + \gamma s(z)_i + \epsilon_i \quad (3.13)$$

where

$$s(z)_i = z_i + \psi_i \quad (3.14)$$

correspond to the predicted value of z at location i .

According to Anselin (2001b) the error structure ψ_i in equation 3.14 might be biased or its expected value might not be equal to zero. Unbiased predictions for

⁴One common example is the creation of continuous temperature surfaces from point measurements obtained from meteorological stations.

z_i are unlikely and only few spatial interpolation techniques generate unbiased predictors for the interpolated value of the explanatory variable z_i . In addition, the error structure will contain a spatial structure and errors will be higher the further away they are from the original data points (Anselin, 2001b). This might lead to inconsistent estimators for γ_i . In this study interpolation techniques are used for the amount and the variance of rainfall and for the population surface. To correct for the spatial effects stemming from the interpolation techniques, the regular spatial sample described in next section is employed. However, no statistical tests were applied to assess the amount of spatial dependence as the standard definition for residuals does not apply in a qualitative dependent variable framework (De Pinto & Nelson, 2002).⁵

3.1.3 Possibilities to correct for spatial effects

The existence of spatial relationships among observations can result in unreliable estimates and misguided statistical inference of the parameters (Anselin, 1988). Econometric problems with spatial data can be due to interactions among neighboring agents (see section 3.1.2). Spatial effects can also emerge when data from different sources, different sample designs or varying aggregation rules is used (Anselin, 1988). The integration of data from various sources will tend to result in spatially dependent as well as spatially heterogeneous observations.

At present, there are no models and test statistics available which account for substantive spatial interaction in a qualitative dependent variable framework (Anselin, 2001a, cited in Munroe et al., 2001).⁶ To correct for spatial dependence in qualitative dependent variable models, three types of ad-hoc corrections, which are frequently applied simultaneously, can be found in the literature (De Pinto & Nelson, 2002):

(a) *Regular sampling from a grid*: With regular spatial sampling a number of cells are selected in a systematic way. This is usually done by keeping only cells which are a specified distance away from the nearest selected neighbor, resulting in a non-contiguous subsample of the data. It involves a *coding scheme*

⁵ Recently, Kelejian & Prucha (2001) developed a technique to calculate the specification test *Moran's I* for the assessment of residual spatial autocorrelation in a limited dependent variable framework. This test is not considered in this thesis as it involves considerable matrix manipulations, which are outside the scope of this thesis.

⁶The Kelejian-Prucha test assesses residual (or nuisance) spatial autocorrelation (see footnote 5).

as defined by Besag (1974) and permits to apply standard estimation methods (Anselin, 2000). Regular spatial sampling was, among others, applied by Nelson & Hellerstein (1997); Nelson et al. (2001a, forthcoming); Munroe et al. (2002); Müller & Zeller (2002, forthcoming).

(b) *Spatial lag variables with index values for latitude and longitude:* One common techniques to correct for spatial effects is to include two additional explanatory variables in the estimation representing the latitude and longitude of each observation. This method is likely to be helpful when the spatial effect is caused by an unobserved variable that varies linearly over the area (De Pinto & Nelson, 2002). Fox et al. (1994); Nelson et al. (2001a); Munroe et al. (2002); Müller & Zeller (2002) included index values for latitude and longitude in their empirical models.

(c) *Spatially lagged geophysical variables:* The inclusion of spatially lagged geophysical variables was applied by Munroe et al. (2001); Nelson et al. (2001a) and Müller & Zeller (forthcoming). It involves the inclusion of an exogenous variable as a *spatial lag operator* of geophysical variables such as slope. A spatial lag is a weighted average of values in neighboring locations and, therefore, *smoothes* the surface. This approach approximates the use of a contiguity spatial weight matrix applied to selected RHS variables (De Pinto & Nelson, 2002).

De Pinto & Nelson (2002), Munroe et al. (2001) and Munroe et al. (2002) evaluate the effectiveness of these ad-hoc methods to correct for spatial effects on model accuracy in limited dependent variable models. Munroe et al. (2001) implement a coding scheme with different sizes of the regular spatial sample and use a *join count statistic* to test the hypothesis of spatial randomness in the land use choices. However, Munroe et al. (2001) point out that the disadvantage is the direct application of the test to the manifestation of a latent process. Spatial effects could be present in the true underlying process but absent in its realization and vice versa. Munroe et al. (2002) found a trade-off between spatial sampling and model accuracy using the Kappa statistic and the recently developed statistic to calculate the *Moran's I* in a limited dependent variable framework, which was at first implemented by Kelejian & Prucha (2001).

De Pinto & Nelson (2002) assess the success of removing spatial autocorrelation for qualitative dependent variable models using Kelejian & Prucha's *Moran's I* statistic. They conclude that these "ad-hoc" methods are not completely successful in eliminating the spatial effects. Still, all three of the above techniques are able to partly correct for spatial effects.

3.1.4 Assessing predictive accuracy

One measure of goodness-of-fit, which allows a spatial assessment of the estimation, is the prediction matrix where actual values are tabulated against the predicted values. This tabulation can also be examined graphically in prediction maps. It facilitates the assessment of model accuracy as it is a one-number summary of the model fit, because it reduces the model complexity to a single number. It is also known as the "percentage predicted correctly". The prediction matrix for a binary model can take the four forms described in table 3.1.

Table 3.1: Prediction matrix

$y = 1$	and predicted	$y = 1$	→	correct prediction
$y = 1$	and predicted	$y = 0$	→	incorrect prediction
$y = 0$	and predicted	$y = 1$	→	incorrect prediction
$y = 0$	and predicted	$y = 0$	→	correct prediction

Source: Hosmer & Lemeshow (2000)

Calculated predicted probabilities can also be employed to evaluate the predictive power of the MNL regressions. Once predicted probabilities are calculated for the entire choice set, one can assume that for all observations the highest probability is the predicted value for the dependent variable. As Nelson & Geoghegan (2002) point out, the probabilities for five categories can range from 20.1% to 99%. Therefore, this measure does not distinguish between "good" and "bad" predictions of the individual observations. Another approach would be to assign threshold values for correct predictions, e.g. to assume that only probability values bigger than 70% are correct predictions. Comparison of the number of correctly predicted pixels to the actual number of pixels can reveal the predictive power for each category (Nelson & Geoghegan, 2002). Once predicted probabilities are calculated, a tabulation of observed values against predicted values yields the prediction matrix. Actual

locations are typically in the rows of the matrix while columns show predicted values. Correct predictions are shown in the diagonal of the matrix. In section 5.3 the goodness-of-fit of the empirical application is assessed using prediction matrices including prediction maps that facilitate a visual analysis of the spatial accuracy.

3.2 Data collection

This section describes the techniques used to collect primary and secondary data. In short digressions the basics of remote sensing (RS), some specifications of the employed RS data, and description of the data collection technique for the left-hand side variable land cover will be introduced. The right-hand side variables are separated into geophysical variables mainly determining the natural agricultural potential, variables describing the socioeconomic village characteristics influencing land use, and policy-induced variables. Data aspects for the various categories of the right-hand-side variables will be presented as well as the hypothesized effects of the RHS variables on land use.

After discussion with government officials the research project focused on the period after liberation day in 1975. Survey recall data was collected for the years 1980 and 1990. Both years were linked to events, which most villagers could remember and therefore provided suitable points of reference for the recall questions. Furthermore, the two periods were conceptually convenient as one was before and the other after the introduction of the market economy.

3.2.1 Land cover data

Precisely defined, land cover is the (bio-)physical ground cover of the land surface and immediate subsurface. Land use, on the other hand, describes human employment of that land. In that sense, hunting would be an example of land use while scrub could be the corresponding cover. However, in literature the two are often used synonymously. In this study, land “cover” will be referred to denote the categories of the dependent variables and land “use” when interpreting the results of the analysis.

Changes in land cover were measured using time series of remotely sensed satellite data. Remote sensing (RS) relies on the measurement of *electromagnetic energy* (EM). Most passive remote sensing sensors measure the wavelength of the sunlight reflected from the earth’s surface. The resulting data can be used to derive *information* about characteristics of the earth’s surface.

Table 3.2: Specifications of Landsat satellite images

Name of scanner	Spatial resolution [m]	Number of bands
Enhanced Thematic Mapper (ETM+)	30*30 ^a	7
Thematic Mapper (TM)	30*30	6
Multispectral Scanner (MSS)	79*79	4

Source: Bakker et al. (2000); ERDAS (1999)

^aThe ETM+ has one panchromatic band with a spatial resolution of 15m (see table 3.3).

In this study, Landsat satellite images⁷ were used, because of its low costs, especially in relation to the area covered.⁸ Another advantage of Landsat satellite images, especially in developing countries where costs often restrict the use of remote sensing data, is the lack of a copyright, which permits the sharing of data among departments and research institutions.

Landsat images measure the sunlight mainly in the spectrum of the visible wavelengths. The Landsat satellites are equipped with a spaceborne remote sensing scanner. It is called a *multispectral scanner*, because the reflected energy is measured simultaneously for different wavelength bands. Each wavelength covers an interval of the electromagnetic band and within each wavelength the average reflected energy is measured. Each of the bands are related to specific characteristics on the earth's surface. For example, the infrared wavelengths give an indication about the type and health of the vegetation (Bakker et al., 2000).

The remotely sensed measures of EM are stored in a regular grid format with a specified number of rows and columns. In the case of Landsat, each wavelength band is stored in a separate layer of data. The bands of the image can be combined in different ways to visualize specific characteristics of land cover. For example, false color composites can be achieved by assigning a band combination of 5-4-3 to the Red, Green and Blue (RGB) channels of the computer monitor. Selected technical characteristics of Landsat images with their respective *spatial resolution* are listed in table 3.2.

⁷The Landsat Program of the U.S. Geological Survey (USGS) is managed from the Earth Resources Observation Systems (EROS) Data Center. For detailed information on the Landsat project see <http://landsat7.usgs.gov/index.php>.

⁸The swath wide of one Landsat ETM+ scene is 183 km by 170 km.

Table 3.3: Characteristics of Landsat ETM+

Band number	Spectral range [μm]	Spatial resolution [m]
1	0.45 to 0.52	30
2	0.53 to 0.61	30
3	0.63 to 0.69	30
4	0.75 to 0.90	30
5	1.55 to 1.75	30
6	10.40 to 12.5	30
7	2.09 to 2.35	30
PAN	0.52 to 0.90	15

Source: ERDAS (1999)

Visual on-screen interpretation is chosen as the technique to derive land cover information from the Landsat satellite images.⁹ Landsat images for the research area were interpreted for the years 1975 (Multi-Spectral Scanner, MSS), 1992 (Thematic Mapper, TM), and 2000 (Enhanced Thematic Mapper Plus, ETM+).¹⁰ Selected technical specifications of the Landsat ETM+ scanner are listed in table 3.3. Note that the Landsat ETM+ has an additional panchromatic band (i.e. an 8-bit greyscale image, labeled *PAN* in table 3.3) with a pixel size of 15 by 15 meters. This panchromatic band has considerable advantages like, for example, to delineate line features like roads and rivers more accurately.

Visual image interpretation extracts information by a visual analysis, which is done by an image interpreter. Colors, structures and patterns of an image are related to real world land cover. On-screen digitizing was used to derive land-cover maps from each image. Visual image interpretation has the advantage that the ability

⁹Another technique to derive land cover information is digital [image classification](#) where an operator instructs the computer to perform an image classification based on the spectral characteristics. The two most common approaches are *unsupervised classification*, where clustering algorithms partition the image into a number of discrete categories, and *supervised classification*, where the image is clustered based on the spectral characteristics of identified sample areas or training fields ([Bakker et al., 2000](#)).

¹⁰To better match remote sensing data with recall information, satellite images from 1980 and 1990 should have been used. However, no cloud-free satellite data for these years were available within the budget capacities of the research project.

of humans to derive conclusions from specific land cover features can be exploited by the interpreter. *Spontaneous recognition* refers to the ability of an interpreter to identify objects at first glance, e.g. irrigation channels or even the boundaries of bigger agricultural plots. In addition, *logical inference* allows the interpreter to draw conclusions based on ground control points, verifications, professional knowledge and field experience (Bakker et al., 2000). *Ground truthing* enable the researchers to collect reference data and to gain detailed knowledge of land cover conditions and their representation in the satellite image on the computer screen. Ground truthing was repeatedly carried out before and during image interpretation for familiarization with the appearance of actual land uses on the screen, for verification and for validation of the interpretation results.

Interpretation elements are hue, texture, shape, size, pattern, site and association. *Hue* refers to the colors of the image, which results from the amount of energy reflected from the earth' surface. It is defined in the intensity-hue-saturation system and allows approximately 10,000 variations in color (Bakker et al., 2000). *Texture* relates to the frequency of tonal change and can, for example, be described as fine, smooth, coarse. The *shape* of objects allows the interpreter to distinguish agricultural fields, roads, irrigation facilities, and other objects recognizable from their silhouettes. Differences in *size* can enable the identification of different agricultural land uses like small shifting cultivation fields and (usually) bigger, adjacent permanent fields. The *pattern* of pixel clusters give an indication of the spatial arrangement of objects. Paddy fields, for example, typically show different spatial patterns, i.e. rectangular, than mixed agricultural or shifting cultivation fields. *Site* refers to the absolute as well as relative geographic location of objects. Irrigation channels will generally be close to irrigated agricultural fields. *Association* permits to make inferences from a combination of objects. A limitation of visual image interpretation is that it is only what the name says it is, an *interpretation*. If different persons interpret the same image, results will consequently be different as they are based on the subjective opinion of each of the interpreters.

Land cover mapping is a process of aggregating the image data into several discrete categories, i.e. land-cover classes. This implies that an almost continuous surface is reduced to discrete classes with a corresponding loss of information. Based on the objectives of this research, available data sources and local conditions it was decided to derive nine land-cover classes from the satellite images.

All three Landsat images were taken in the spring season between March 4th and April 7th. Hence, they stem from the same cropping season and from comparable climatic conditions. This enhances the interpretation as the spectral reflection of land cover is easier to compare across the different images. In addition, recently updated 1:50,000 [Universal Transverse Mercator](#) (UTM) topographic maps were scanned, [georeferenced](#), and merged to obtain a consistent set of base information. These maps allow the verification of land cover delineation using additional point information and linear features such as contours, roads, and rivers. The interpretation resulted in land cover maps, which indicate physical land cover of the respective years for the major land-use classes. Overlays of land-cover maps from different points in time produce quantitative estimates of changes in land cover over time.

To reduce model complexity and to ease model interpretation, the land-cover data are aggregated into five classes, which were used for subsequent econometric analysis:

1. mixed agricultural land, including cash and food crops as well as plots under shifting cultivation;
2. paddy fields, both with one and two crops a year;
3. closed canopy, dense forest;
4. open canopy, mixed and secondary forest;
5. mixed grass land containing bamboo, bush, and shrub as well as re-growing shifting cultivation fields.¹¹

3.2.2 Geophysical data

To a considerable extent, data on geophysical variables determine the agricultural potential inherent in a specific plot of land. Geophysical variables include rainfall time series, soil types, elevation, and slope values. Measurement, level and source of the exogenous variables are depicted in table 3.4. Rainfall data stems from five rainfall stations in and several others around the research area. Daily time series were available starting from Liberation Day in 1976 until 1999. These time series

¹¹Field verification by means of ground-truthing revealed that most shifting cultivation plots, which were in fallow for one year, were still classified as agricultural land. Longer fallow periods result in the category “mixed grassland”.

allowed for the calculation of the yearly mean and variance of daily rainfall amounts for both time periods considered in the data analysis, i.e. from 1975 to 1992 and from 1992 to 1999.¹² Continuous surfaces for these layers were created by interpolating the point coverage with a regularized spline interpolator (ESRI, 1999; Mitas & Mitasova, 1999).

A digital soil map (DSM) on a scale of 1:250,000 has been reclassified into nine suitability classes for paddy and mixed agricultural production with “one” exhibiting the most suitable soils and “nine” the least suitable soils. The suitability classes are based on independent classifications from numerous soil experts from the Department of Agriculture and Rural Development of Dak Lak province and from the Tay Nguyen University of Buon Ma Thout. The average of all these classifications is used as the soil suitability class in the estimation. The Digital Elevation Model (DEM) from the Mekong River Commission (MRC) is based on contour lines from American topographic UTM maps, scaled at 1:50,000. The DEM has a spatial resolution of 50 by 50 meters. Slope degrees were calculated from this DEM.

A dummy variable for land suitable for permanent agricultural production is computed on a pixel basis. This variable takes on the value *one* for pixels suitable for agricultural production and zero for pixels, which are already cultivated, have a slope over 15 degrees and belong to a soil suitability class better than four. The lagged value for 1992 and 1975, respectively, of this spatially explicit dummy variable convertible land is hypothesized to have a significant influence in the regressions.¹³

¹²The rain-fed cropping season is almost all-year-round, and only four months of the year, i.e. December to March, are considered to be the dry season. During this dry season, however, there is considerable variation with respect to the commencement and end of rainfall due to regional diversity in altitude and micro-climate (proximity to forests). Rainfall during the dry season influences soil moisture for the planting season and has an effect on water availability in semi-technical irrigation systems that use collected rainfall water as a source. Therefore, the mean and the variance of rainfall data for the entire year was calculated and included in the empirical application.

¹³This variable was later dropped, because it led to estimation problems as it determined the category of paddy land of the dependent variable almost perfectly. Then, the likelihood functions can't be computed as the coefficient for paddy would be close to negative infinite. This leads to missing values in the variance-covariance matrix and therefore to missing standard errors, which prohibits the calculation of model statistics (StataCorp, 2003).

3.2.3 Village survey

A village survey was conducted in 101 randomly selected villages out of a total of 191 villages in the research area using a standardized questionnaire.¹⁴ To enable better generalization, the villages were selected according to a stratified random sampling with two strata, one for ethnic Vietnamese (*Kinh*) and one for all other ethnic groups. Proportional random sampling was used so that the villages were randomly selected in proportion to the actual number of sampling units in the total population (all villages) of each strata (Poate & Daplyn, 1993). The proportional stratified random sampling ensures that each of the strata is selected according to the share of the respective group. Therefore, no weighting factors were applied in the regressions.

The survey provides primary data on the current status and changes in land use practices since Liberation Day in 1975, and on the causes of these changes and the perception of the villagers concerning the effects on economic, socio-cultural, and ecological variables. Recall techniques were employed to obtain estimates from villagers for Liberation Day in 1975, for 1980, and 1990. Additional socioeconomic indicators were collected from interviews with officials of all communal People's Committees (PC) in the research area and from secondary sources such as official government statistics.

¹⁴The questionnaire can be obtained from the [author](#) upon request.

Table 3.4: Measurement, level and source of exogenous variables

Variable	Level of measurement	Pixel/village	Source
Geophysical indicators			
Slope	degrees	pixel	Digital Elevation Model (DEM)
Elevation	100 meters	pixel	DEM
Soil suitability ^a	9 classes	pixel	Digital soil map (DSM)
Mean rainfall amount and variance	100 millimeters	pixel	point interpolation
Land suitable for conversion	0/1	pixel	land use maps and DEM
Socioeconomic history of village			
Years since village establishment	2000 - year ^b	village	survey
Ethnic composition	0/1	village	survey
Lagged population	households	pixel	survey (point interpolation)
Hero village	0/1	village	survey
Policy Variables			
Distance to all-year road	km	pixel	road network/survey
Years since primary school	2000 - year ^b	village	survey
Years since introduction of NPK	2000 - year ^b	village	survey
Years since construction of reservoir	2000 - year ^b	village	survey
Increase in irrigated area	hectare	village	survey
Nature reserve areas	2000 - year ^b	pixel	official park boundaries
Spatial Effects			
X and Y coordinates	every 5 th	pixel	sampled from grid

Source: Author

^aSoil suitability classes are based on expert opinion and range from one for the highest to nine for the lowest soil suitability. They are included for both paddy and mixed agricultural land uses.

^bIf not observed in the village, this variable takes the value zero.

3.2.3.1 Socioeconomic village characteristics

A metric variable is created for the time since the foundation of the villages at their current location, defined as the *age of the villages* in years. Villages in the research area are usually inhabited by kinh or ethnic minority people.¹⁵ Mixed villages do not exist, apart from one or two Vietnamese families, who usually own a shop in an ethnic village. For that reason, a dummy variable coded one for an ethnic minority village, and zero otherwise, is included to capture the effect of *ethnic composition* on patterns of land use. Another dummy is coded one for villages awarded the status of a *hero village*. Villages, which assisted the Northern Vietnamese government in the fight against the South and the U.S.A were awarded this title from the communist government after the American War.¹⁶

For population data from the village survey an inverse distance weighted interpolation technique is applied to derive a continuous population surface. Inverse distance weighting calculates a linearly weighted combination from the a set of sample points, i.e. the village locations. The assumption behind inverse distance weighting is that closer points are more spatially autocorrelated than points further away (DeMers, 2002). The continuous surface is interpolated from point values at village locations where the influence of a particular point's value on the interpolated grid cell value depends on how far the point is from the cell being interpolated (ESRI, 1999). As barriers the village accessibility catchments (explained in section 3.3.1) were used and points outside village catchments have no influence. Points inside the barriers are calculated using inverse distance weighted interpolation in ArcInfo[®]. Thus, the fact that the population is highest at actual village locations and pressure decreases as one moves away from population centers is accounted for. The population surface aims to represent population distribution more realistically with a distribution of the data values as a property of location rather than an entire administrative area represented by a polygon (Bracken, 1994). As the villages are not scattered over space across several hamlets or small settlements, this assump-

¹⁵People of ethnic Vietnamese origin are called “*Kinh*” people, which will be used hereafter to distinguish them from people belonging to ethnic minority groups, referred to as “*ethnic people*”. Politically, both groups are Vietnamese citizens. Also note that the expression “*minority*” is used in a purely quantitative sense indicating that “*ethnic people*” are outnumbered by the “*Kinh*” in Vietnam.

¹⁶The *American War* in Vietnamese wording is equivalent to the *Vietnam War* as the rest of the world calls it.

tion is fairly realistic. The lagged value of the continuous population surface was included in the data analysis as a predetermined variable and, therefore, treated as an exogenous factor in the reduced-form equations.

3.2.3.2 Policy variables

Policy-induced variables reflect investments from external sources such as governments and donors in infrastructure and technologies as well as restrictions on land use imposed by the government.

Surface [vectors](#) for roads obtained on a scale of 1 : 250,000 are corrected for the research area to a scale of 1 : 50,000 using the digitized UTM base maps and the pan-chromatic band from the Landsat ETM+ with a ground resolution of 15 meters (table 3.3). Market access is measured by the Euclidean distance to the nearest all-year road for trucks and to the nearest district center. Based on recall information, a secondary road network is approximated for the year 1990. This road network is highly correlated with Euclidean distance to district capitals. The time since primary school existence in a village is proxied by number of years since the establishment of a primary school and included as an indicator of provincial and national government investments in human capital formation as a basic driving force for land-use change.

Technology variables are incorporated in the analysis with regard to the time since the introduction of a new technology for agricultural production. Included are (a) the number of years since the introduction of Nitrogen-Phosphorus-Potassium (NPK) fertilizers as a yield-increasing technology, (b) the number of years since the construction of a reservoir for irrigation of agricultural land to capture the time period since the establishment of improved irrigation facilities, and (c) the increase of irrigated area per village in the previous period resulting from additional or improved irrigation services.

The boundaries for National Parks and Nature Reserves are included as an additional policy variable ([Chomitz & Gray, 1996](#); [Deininger & Minten, 2002](#)). The time of protection is incorporated as the number of years since a pixel belongs to a protected area, zero otherwise.

There is possibly a range of other important variables, which can have a substantial effect on land use. For instance, land tenure and property rights shifted recently from a system based on customary rights to an official recognition of prop-

erty rights.¹⁷ In the research area, the first land use titles for agricultural land were issued in 1996. Until today, 65% of the interviewed households received these certificates – 45% of them in the past two years. Therefore, it is too early to estimate the effects of more secure property rights on land use given the temporal resolution of the land-cover data, which is only measured for three points in time, i.e. 1975, 1992 and 2000.¹⁸

3.3 Integration of socioeconomic and spatially explicit data

Farmers, and other land users, make their land management decisions based on objectives and constraints of their households and individual members. At the village level, an aggregate behavior of all land users and other stakeholders in land allocation and management is observed. This can also be considered as aggregated household behavior, which in turn is aggregated behavior of household members, i.e. individuals. To assign the socioeconomic entities to the pixels in the landscape requires a mechanism to link the two data sources is required (Rindfuss et al., 2002). Ideally, to integrate grid-based data with survey data, the scale of the analysis should match the unit of decision-making. Yet, in Vietnam as in most developing countries, plot maps and village boundaries are not available. This renders spatial modeling on the level of decision-making a time-consuming and costly task due to the necessary delineation of the spatial extent of plots or villages, e.g. using [Global Positioning Systems](#) (GPS). In this study, the spatial extent of plots and village boundaries could not be delineated with the time and budget available.¹⁹

¹⁷For an excellent discussion about the recent history of land tenure and its effects on village land use for an area in Northern Vietnam, see e.g. [Sikor \(1999\)](#).

¹⁸A change in tenure status might, in addition, influence land use mainly on existing agricultural plots by altering the crop composition. Changes in crop composition could, however, not be measured with the land-cover data at hand (see also subsection 3.2.1).

¹⁹Delineating plot boundaries as the base unit of analysis would match the unit of decision-making of an individual land user much more accurate, thereby being considerably closer to economic reality. However, this is obviously not feasible for an area of 2,327 km² with approximately 20,000 households involved in agriculture, each typically having several dispersed plots. Examples of analysis at the plot level linking geophysical and agroecological data on plots with socioeconomic data of farms or villages include [Fox et al. \(1994\)](#); [Mertens et al. \(2000\)](#) and [Vance & Geoghegan \(2002\)](#).

3.3.1 Village border approximation

Different approaches can be considered to delineate village areas. Possible approaches to define village boundaries based on Euclidean distance are (a) concentric rings around the village centroid with the radius as a proxy for travel time. With an uneven distribution of points and many points close together this will result in overlapping boundaries of the resulting polygons. Another approach are (b) [Thiessen polygons](#), which partition the area into non-overlapping regions, where each coordinate within one area is closer to this area's data point (i.e. the village centroid in our case) than to any other data point (i.e. other villages, [Martin, 2000](#)). For a graphical example of case (a) and (b) see [Entwisle et al. \(1998\)](#) and for an application to spatially explicit land use modeling applying Thiessen polygons see e.g. [Müller & Zeller \(2002\)](#). These two purely geometric approaches have similar shortcomings in that they assume equal travel time and transportation costs in all directions, abstracting from surface impedance as well as from road existence and quality ([Bigman & Fofack, 2000](#)). Equivalent to von Thünen's model, absolute distance corresponds to relative distance in the geometric measures (see section [2.3](#)).

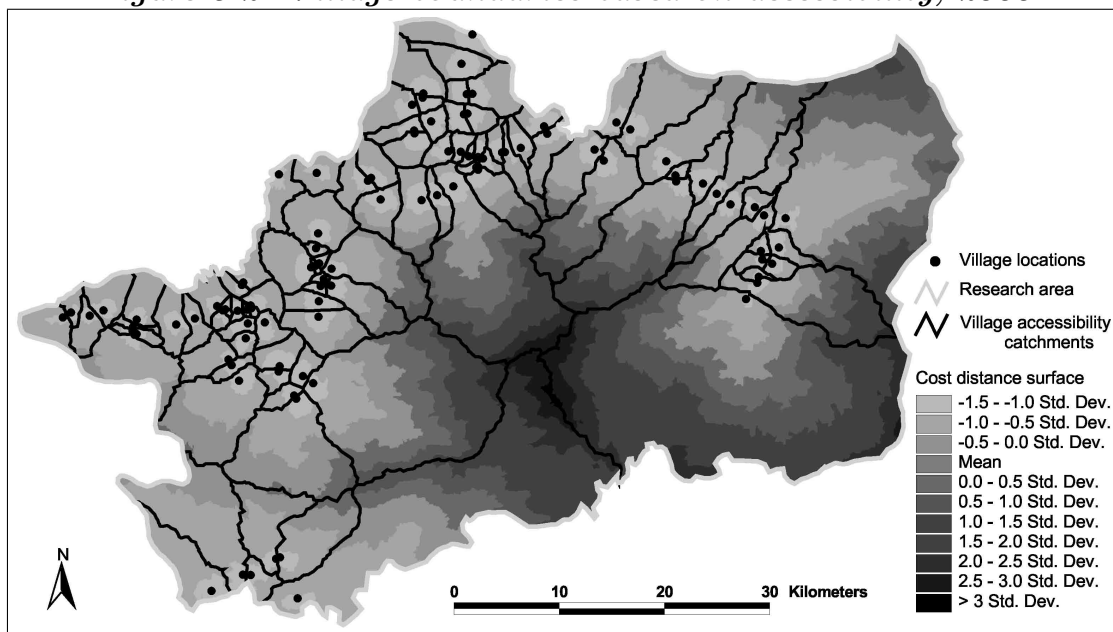
However, transportation costs are a critical indicator for an economy as it affects the movement of people as well as the purchase, distribution and sale of goods and services. Transportation costs also have fundamental effects on rural development. Especially in rural settings transportation costs are a major factor determining market access, and are often highly correlated with the degree of market integration. Consequently, the incorporation of locational attributes is essential in regional rural development analysis as it determines access to markets and political centers.

Another approach to approximate village boundaries is the use of *accessibility catchments* based on transport costs. Accessibility can be defined as the ability for interaction or contact with sites of economic or social opportunity ([Deichmann, 1997](#); [Bigman & Deichmann, 2000](#)). Spatial accessibility is similar to Euclidean distance functions, but instead of calculating the actual distance from one point to another, the shortest cost distance (or the accumulated travel cost) from each cell to the nearest source cell is determined. A second exception is that cost distance functions apply distance not in geographic units but in cost units ([ESRI, 2000](#)).

To achieve fairly accurate measures of the time needed to travel from each cell in the raster surface to a target location, i.e. the village location, the required travel time to the next village is estimated. To approximate this travel time an

accessibility surface is created in ArcView[®] GIS using the cost distance algorithm provided by a standard extension of ArcView.²⁰ Accessibility is calculated on a friction surface, which is a cost measure to move goods from any location in the study region (each cell) to the nearest target cell (the nearest village location). This results in irregularly shaped village boundaries (see figure 3.2), which are designed to reflect access conditions on the ground more accurately (Bigman & Fofack, 2000).

Figure 3.2: Village boundaries based on accessibility, 2000



Source: Author

The friction surface constructed in this study is a combination of the road network incorporating road quality, slope values and land cover. As no actual data on travel time are available, arbitrary values are assumed for the time to travel across the different road qualities, different land covers and different slope values. These values are designed to approximate the time needed to traverse a particular cell (Müller & Zeller, forthcoming). In the study region road quality is highly variable. Therefore, the road network is classified into four quality classes ranging from paths for walking, loosely surfaced seasonal roads, which are usually impassable during the entire rainy season, loosely surfaced roads impassable during rainfall peaks of

²⁰This extension is called *Cost Distance Grid Tools* and is available for free at <http://arcscrips.esri.com>.

the rainy season and tarred, hard-surfaced all-weather and all-year roads. These classes originate from a classification done by villagers in the survey interviews.

The costs to travel over certain land-cover categories are approximated with the highest impedance factor assigned to primary forest and the lowest to mixed agricultural land. Slope values are reclassified using the function $(1 + \frac{slope^2}{50})$, following Nelson et al. (2001b). This function has the desired property that costs increase more rapidly as slope increases. The shortcoming that no actual travel times are available results in travel estimates without units, which only allow to undertake ordinal comparisons of market access between any two points in the research area. However, the interest is on retrieving the influence of varying accessibility on land use rather than conducting a study on the effects of transport costs on land use.²¹ As approximated travel times serve only to define the village boundaries necessary for spatial analysis later on, real absolute travel time is not relevant for the purpose of the study at hand.

The cost-distance algorithm is then applied to derive a set of explicitly defined catchments, generated around each village location and similar in their construction to the delineation of watersheds (river catchments), which are usually delineated using digital elevation models. The generated maps differ from Thiessen polygons in that they are defined (a) by local geophysical (slope), agricultural (land cover) and economic (road quality) factors, and, (b) according to the purpose of this study (Nelson et al., 2001b).

The vast areas of virtually unused and remote forest land are implicitly integrated in the estimations of chapter 5 and 6 via geophysical indicators such as topography, soil quality, and slope as well as distance to major roads, which all render land use unprofitable on these marginal lands. The high variability in the size of the village accessibility catchments in figure 3.2 should therefore not distort regression results considerably.

3.3.2 Data integration and scale of analysis

The data is aggregated to one spatial resolution (i.e. to one grid size), which is used as the unit of analysis for all regressions in this study. Most of the data at hand is

²¹More precise data on transport costs enable a better approximation of the spatial variability of prices. Even more useful could be a spatially explicit database on input and output prices, which would implicitly capture the effects of transport costs as well as other indicators for price formation like market power.

available at this resolution, some variables had to be [resampled](#). The research area was spatially referenced to pixels with a size of 50 by 50 meters. Each pixel therefore represents an area of a quarter of a hectare. After resampling all variables, which are available as georeferenced surfaces in the GIS, to above grid size, the data of the entire research are stored in a [raster](#) GIS and contain 964,000 grids cells. The chosen cell size is arbitrary, which might result in measurement errors or even biased estimates (see subsection [3.1.2.1](#)). Combining data at different cell sizes is, however, mathematically challenging and usually complicates interpretation.²²

Survey data is combined with the raster data from the GIS based on the accessibility catchments defined in the previous section. Hence, socioeconomic survey data takes the same value for all cells located within the approximated village area of an interviewed village. As the number of villages differ in the considered periods, one cost distance surface is constructed for the year 2000 (see figure [3.2](#)) and another one for the year 1992. Both contain approximated village areas for already existing villages at the respective point in time. All the surfaces of the raster GIS are merged in the statistical software package based on the unique locational identity codes of each raster cell. Econometric analysis is based on this data set and every pixel in the GIS surfaces equals one observation in the statistical software package.

3.4 Corrections for spatial effects

To compensate for potential spatial dependence in the dependent variables the three ad-hoc techniques described in section [3.1.3](#) are applied to estimate the models presented in chapter [5](#).

(a) *Regular spatial sample*: Besag's coding scheme was used to draw a spatial sample of observations from the grids covering the research area ([Besag, 1974](#)). The regular spatial sample was drawn by selecting every fifth cell in the X and Y directions so that no selected cells are physical neighbors and all cells are separated by a distance of 200 meters. The sampling of non-neighbors increases the likelihood that observations are spatially independent, although the degree of spatial dependence is not assessed. This sampling procedure allows to reduce the sample size and, therefore, calculation time. However,

²²One potential econometric approach is multilevel modeling where explanatory variables of both individual- and area-level effects are estimated in a nested form. In that way, variables measured at varying scales can be considered.

a considerable arbitrariness is involved in the selection of the coding scheme [Anselin \(2000\)](#).

- (b) *Latitude and longitude as exogenous variables:* Exogenous variables indicating the X and Y location (the rows and columns in the grid) are included as additional regressors for X and Y coordinates.
- (c) *Spatially lagged slope:* A spatial lag operator for slope is incorporated as a RHS variable. A nine by nine *kernel*, shown in table 3.5, is used to derive this spatially lagged slope. The values of the kernel are artificial and only intend to smooth the surface of the slope with a decreasing influence of locations further away. Zero values in table 3.5 identify cells, which are not used for the calculation of the spatial lag ([ESRI, 1999](#)). Therefore, influential observations for the center cell i are up to four cells or 200 meters away. The choice of this kernel size ensures that no value of cells selected by the regular spatial sample is used twice for the calculation of the spatial lag.²³ Kernel values greater than one specify the spatial weights, which make slope dependent on the distance from the actual observation i . When processing the neighborhood of a cell from the slope grid with the spatial weights matrix in table 3.5, the value at each cell position in the neighborhood of the slope grid is multiplied by the value of the corresponding cell position in the kernel file. The products are added for the nine by nine neighborhood. A zero within the kernel file is treated like any other value in the file. The results define the values for the new spatial lag variable, subsequently called *spatial lag slope*, which is determined by neighborhood influences ([ESRI, 1999](#)).

After applying the spatial sample, the data from the socioeconomic survey was merged with geophysical and land use data based on the village accessibility catchments (section 3.3.2). These sampling procedures resulted in a data set of 38,000 observations, which were used for subsequent econometric modeling in chapter 5.

3.5 Expected effects of regressors

Regressions are estimated for four types of regressors. *First*, time-invariant geophysical variables are used. One exception is rainfall, which is available in daily

²³The author did not find empirical evidence to verify this aspect, but assumes that the selection of this kernel size might further reduce spatial autocorrelation in the final data set.

Table 3.5: Kernel to derive spatially lagged slope

0	0	0	0	1	0	0	0	0
0	0	0	1	2	1	0	0	0
0	0	1	2	3	2	1	0	0
0	1	2	3	4	3	2	1	0
1	2	3	4	5	4	3	2	1
0	1	2	3	4	3	2	1	0
0	0	1	2	3	2	1	0	0
0	0	0	1	2	1	0	0	0
0	0	0	0	1	0	0	0	0

Source: Author

time series since liberation day and calculated for both periods. *Second*, historical variables are employed, which describe socioeconomic characteristics of the villages. Here, population as a time-variant, but lagged variable referring to the state in the previous period is included. *Third*, the effects of exogenous policy variables to describe government investments and macro-level policies are calculated in the model (see also the discussion on endogeneity in section 3.6). *Fourth*, the coordinates in X and Y direction are used to correct for potential spatial dependence.

Hypothesized effects of regressors on land-use categories can be anticipated based on previously outlined theoretical considerations, research findings, and field experience. Table 3.4 summarizes the variables and table 3.6 on the facing page their expected effects where a plus sign indicates a positive expected influence of that variable on the respective land-cover class and a minus indicates a negative effect. Question marks indicate potentially ambiguous outcomes. In the following, the hypotheses are stated for the four types of regressors.

3.5.1 Geophysical data

On higher *elevations*, agricultural land use is less feasible and less likely due to more difficult climatic conditions such as lower temperatures and terrain, which is more difficult to cultivate. Equally, agriculture on steeper *slopes* and poorer *soils* is usually less profitable and more difficult. Accordingly, the hypotheses are that the coefficients for elevation, slope, and soil are negative for agricultural land uses and positive for forested areas.

Table 3.6: Predicted signs of exogenous variables

	Mixed agriculture	Paddy	Closed forest	Open forest	Mixed grass
Geophysical					
Slope degrees ^a	-	-	+	+	?
Elevation	-	-	+	+	?
Soil suitability for mixed agriculture	-	-	+	+	+
Soil suitability for paddy	-	-	+	+	+
Mean rainfall amount	+	+	+	-	-
Mean variance of rainfall	-	+	+	+	+
Land suitable for conversion	+	+	-	-	-
Socioeconomic history of village					
Years since village establishment	+	+	-	-	-
Ethnic composition	+	-	-	-	+
Lagged population	+	+	-	-	?
Hero village	-	+	+	+	+
Policy Variables					
Distance and travel time to all-year road	-	-	+	+	?
Years since primary school	+	+	+	-	-
Years since introduction of NPK	?	?	?	?	?
Years since reservoir construction	-	+	+	+	-
Increase in irrigated area	-	+	+	+	-
Nature reserve areas	-	-	+	+	-
Spatial Effects					
X and Y coordinates	?	?	?	?	?

Source: Author^aThe same effects are anticipated for spatially lagged slope.

The *amount and variance of rainfall* critically affect agricultural activities in several ways. In areas with high absolute rainfall, more water-demanding crops can be cultivated. To decrease risks from highly variable and, therefore, more insecure rainfall, more drought-resistant crops and a higher crop diversification should be expected. However, since the land-cover data at hand does not permit to explain crop choice, only the following hypotheses can be deduced: More absolute amounts of rainfall in the moderate climate prevailing in the research area can be expected to increase the natural potential for agriculture, while, due to increasing risks, one would expect that higher rainfall variation to reduce land under mixed, rain-fed agricultural land in the absence of irrigation facilities and to increase the likelihood of other uses.²⁴

3.5.2 Socioeconomic village characteristics

Rural development is essentially path-dependent. Hence, initial states and historical events influence the current and future socioeconomic situation of villages, including land use. In this context, several variables will be considered and explained in the following.

More agricultural area and less forest and grass is anticipated in *older villages* since more permanent conversion of land into agricultural uses is likely. A positive relation is expected between *ethnic minority villages* and former shifting cultivation areas, be they fallow or currently cultivated. Fallow fields are either interpreted as grass land or as agricultural land, depending on the time of fallow.¹¹ Some fields might have already recovered to become secondary, open canopy forest. Due to these traditional land intensive cultivation techniques and presumably less political power, ethnic villages might be associated with more agricultural area (cultivated less intensively), less irrigated paddy area, and more grass land. Common belief also attributes deforestation to ethnic people; therefore, less forest of either type would be expected.

²⁴In more tropical settings with higher amounts of rainfall other authors found the opposite effect with a proportion of land in agriculture declining with increasing rainfall, reaching near zero at rainfall amounts of 3,200 mm (Chomitz & Thomas, 2001). In the research area rainfall averaged 1,700 mm between 1976 and 1999. Thus, in the relatively drier climates of Dak Lak province, absolute rainfall is a restricting factor at least part of the year. Therefore, an increase in agricultural area is expected for locations with higher amounts of rainfall.

Available *convertible land* provides the potential for agricultural expansion. A higher lagged value should increase the probability that these patches will be converted for agricultural land uses. A strong positive relationship for agriculture and paddy land uses and negative signs for other land uses would be anticipated.

Higher lagged *population* should be associated with an increasing demand for agricultural land and is often cited as a major factor influencing deforestation given that land is available for conversion into agricultural uses (see e.g. [Kaimowitz & Angelsen, 1998](#); [Bilsborrow & Carr, 2001](#); [Tachibana et al., 2001](#)).

The status of *hero villages* is used as an operational variable to proxy the political capital of a village. Higher political capital ([Birner & Wittmer, 2003](#)) is hypothesized to improve villagers' access to government programs and outside assistance. This may well lead to a better access to extension services and credit supplies, which could not be measured directly. Subsequently, the intensification effect might be associated with a reduction of land expansion, therefore reducing the pressure on primary forests and allowing secondary forests and grass land to recover. Area under paddy production is assumed to increase as government investments in irrigation services and terracing are anticipated.

3.5.3 Policy variables

With respect to infrastructure investments by the government, several operational variables derived from the village survey are employed. Since no farm-gate prices for inputs or outputs are available with spatial variation, *distance and travel time* from the village to all-year roads serve as a proxy for prices as well as transaction costs ([Deininger & Minten, 2002](#)) and access to political centers. Lower values of these variables might be associated with lower forest cover because of higher producer prices, lower input prices and lower transaction costs compared to remote locations. Yet, [Deininger & Minten \(2002\)](#) find that these variables are highly correlated with population density, soil fertility, and physical capital such as irrigation facilities. However, that was not the case in the our data set where these explanatory variables all showed correlation coefficients below 0.5. Agricultural land use is expected to be more likely to occur closer to all-year roads, while forest cover should be less likely ([Chomitz & Gray, 1996](#); [Nelson & Hellerstein, 1997](#); [Pfaff, 1997](#)). Again, the effect on mixed grass land cannot be hypothesized a priori. Most of the all-year road network, which is employed as explanatory variable in the estimation, is assumed to be exogenous as it was build by the French before liberation day. Agricultural

cultivation started off from this road where New Economic Zones (NEZ) are located and especially *Kinh* migrants settled after liberation day. Still, many immigrants have settled in areas with favorable natural conditions and government investments into the road network might have provided better access to these regions, particularly in the last decade.

Access to education is represented by the availability of primary schools in years since their opening. The a priori effect is uncertain. Common sense tells that farmers with better access to education should be expected to use land more intensively and adopt new technologies earlier, thus saving land to achieve a given income or farm output. However, there is also empirical evidence that educated households might engage more in land clearing (Pichon, 1997). Here, better access to education is likely to be associated with less shifting cultivation and a more permanent agricultural land use. Therefore, grass land should be less and primary forest more likely.

Technological change, which has been promoted by a variety of government agencies, is incorporated via the introduction of NPK fertilizers, one example for a yield-increasing technology. NPK increases first and foremost the productivity of land, and one might expect land expansion to become more attractive with increasing availability of NPK, provided farmers have sufficient labor and the ability to finance the purchase of agricultural inputs. However, given labor constraints and incomplete rural financial markets, increasing yields and more input in production on existing fields should be more likely. The longer ago the introduction of NPK, the more likely the possibility of forest cover. On the other hand, more profitable agriculture could increase the incentive to expand area under cultivation and, therefore, lead to a lower forest cover and a positive coefficient for agriculture (Angelsen, 2001). Therefore, the overall expected sign of the effect is ambiguous. Other variables related to technological change are the years since the construction of a reservoir for irrigation and the increase in irrigated area of the villages. It is anticipated that these labor-intensive technologies, which facilitate intensification on existing agricultural areas, will induce households to allocate less time to land clearing and land expansion into upland areas. Accordingly, upland agricultural area is expected to recover and forested pixels will be more likely (Shively & Martinez, 2001; Pender et al., 2001b).

Inside a protected area, it is impossible to establish property rights for long-term secure cultivation, thereby decreasing the profitability of agriculture (Deininger &

Minten, 2002). A significant positive coefficient is expected for non-agricultural land uses and a regeneration of grass into secondary forest. Protected areas may be located on more marginal lands with steeper slopes and on less suitable soils. The delineation of a protected area, which often restricts land and forest use, will cause less opposition, have a higher chance to protect environmental resources and inhibit reduced monitoring costs, if established on more marginal lands with lower population pressure. For that reason, the protection dummy is interacted with slope and soil suitability for mixed agriculture. If the hypothesis holds, the introduction of these interaction terms should reduce the significance and the effect of the protection dummy on preserving forest land.

3.6 Methodological constraints and data limitations

Econometric constraints and data limitations are a common issue in empirical studies. Difficulties in this study arise especially from the long time horizon and the integration of the various data sources. In the following the author briefly discusses the weaknesses in the treatment of endogeneity and data constraints arising from the long time horizon.

3.6.1 Endogeneity issues

Endogeneity issues are widespread in many econometric studies. In econometric terms *endogeneity* arises, if an explanatory variable x is correlated with the disturbance term u . If x and u are uncorrelated then x is said to be *exogenous*. Possible causes are omitted variables, measurement errors or simultaneity, which can also arise simultaneously (Wooldridge, 2001). Still, the distinction of endogenous and exogenous variables is a controversial issue in many empirical studies (Greene, 1997).

This study as well potentially suffers from these three sources of endogeneity. First, unobserved factors, e.g. on socioeconomic characteristics, might lead to the problem of omitted variables. However, based on field experience and numerous discussions, the author believes that most relevant indicators were indeed collected and included in the estimation. Second, measurement errors might be present in both land-cover data as well as other spatially explicit and survey data (see also section 3.6.2 below). And, third, simultaneity problems could arise when changes in the explanatory (independent) variables are caused in part by the dependent variable. For example, the spatial placement of government programs is likely not to be random, but a response to unmeasured attributes of the targeted population

(Pitt et al., 1993). A correlation of program placement with an unobserved variable will result in biased estimates. Roads, for example, might be built in areas, which are suitable for agricultural development. As Chomitz & Gray (1996) point out, “. . . a plot of land may be undeveloped not because it is far from the road – it may be far from the road because it is not suitable for development”. Along these lines, the inclusion of topographic and soil suitability indicators is assumed to partly eliminate for the potential bias arising from road endogeneity (Chomitz & Gray, 1996). Similarly, the establishment of irrigation facilities is affected by the quasi-fixed variables slope and soil suitability, which in turn determines their placement. Consequently, government investments are assumed to be exogenous for the research area.

To control for potential problems due to the inclusion of endogenous variables only reduced-form models are estimated. Potentially endogenous time-variant variables are included in lagged form as predetermined variables, which can be treated, at least asymptotically, as exogenous. Consistent estimates can be achieved in that way (Greene, 1997). However, in dealing with potentially endogenous explanatory variables, lagged values might not always solve the endogeneity problem. For example, temporal autocorrelation could be present in the process generating the endogenous RHS variable (John L. Pender, personal communication, August 2002).

3.6.2 Data issues

This study intends to cover the period since Liberation Day, i.e. the last 25 years (see the section 3.2.3 on employed survey methodology). Consequently, survey data had to be collected for this period by employing recall techniques. Recall is a frequently used technique for data collection in order to enhance longitudinal analysis when secondary panel data does not exist. This technique of gathering data with retrospective questions is prone to measurement errors like misrepresentations, underreporting, omission and intentionally omitted or distorted answering and must therefore be treated with care. Smith & Thomas (2003) found that recall quality is higher among the better educated. Respondents better remember more salient moves, which can be linked with important events. In the research area we could relate questions for 1980 with a large flood that occurred in 1981, which nearly all of the villagers remembered (see also the rainfall data in figure A.2 in the appendix). The recall information from 1990 could be associated with a drought (figure A.2) as well as with the period characterized by the actual enforcement in the research area

of the economic renovation policy (*doi moi*) in the late 1980s, which also provided another suitable point of reference.

In addition, the quality of the recall data depends crucially on the questioning techniques of the researcher and on the selection of participants in the group interview, e.g. across social classes and gender. In the research area people are of divers ethnic and cultural backgrounds. The author suspects that, despite thorough explanations in the village interviews, the understanding of some terms like e.g. “all-year road”, “primary forest” or terms related to “quality of housing” was not always identical and might have been culturally biased. Questions were designed to validate information and survey enumerators were encouraged to conduct a variety of cross-checks in the interviews to facilitate a more accurate collection of the retrospective data.

The author conjectures that the figures for 2000 might in some cases underestimate the actual yield levels due to strategic behavior of the responding villagers. Recall data, especially for 1980, is somewhat questionable and can provide only a rough estimate due to the long time horizon. Data from the [Dak Lak Statistical Department \(1981\)](#) are probably closer to reality for the yield levels in 1980, which is also confirmed by personal communication with district and provincial officials. In the econometric estimations of chapter 5 variables derived from the village survey merely enter as metric variables, i.e. as the number of years since the occurrence of a certain event. To better match remote sensing data with other recall information, satellite images from 1980 and 1990 should have been used. However, no cloud-free satellite data for these years were available within the budget capacities of the research project. The combination of survey data containing recall information for the years 1975, 1980 and 1990 with the land-cover maps for 1975 and 1992 as dependent variable does not pose serious econometric problems considering the mostly metric variables employed. The author believes that the dating of the year of certain events is achieved with relatively high consistency and accuracy. Other recall information included in the estimations are the two dummy indicators for *ethnic composition* and *hero villages*. The remaining survey data are mainly employed for the descriptive chapter 4 of this thesis. Some of these indicators have to be treated with care keeping in mind the collection techniques described previously.

3.7 Summary

The influence of exogenous and predetermined driving forces of land-use change are estimated with a multinomial logit model for discrete unordered states of the dependent variable. The fundamental specifications of the estimation of MNL models using maximum likelihood estimation is outlined and serves as a basis for the empirical application in chapter 5. The spatially explicit data analysis requires some additional elaboration on the basic concepts of spatial econometrics including techniques available to correct for spatial effects in a limited dependent variable framework. The assessment of estimation power and model accuracy concludes the section on the econometric modeling approaches chosen for this thesis.

A variety of data sources are employed in this thesis. The explanation of the data collection procedures starts with the generation of the land-cover data derived from the Landsat satellite images, elucidates the collection and manipulation of geophysical indicators and outlines the implementation of the village survey. The combination of these data sources requires special techniques for the integration of the spatially explicit GIS data based on grid cells with the survey indicators, which are based on the village interviews. The approach taken is the construction of village accessibility catchments, which approximate village areas based on transport costs. The chapter concludes with a discussion of expected effects of the exogenous and predetermined variables based on previous research and knowledge of the research area. In the discussion of the estimation results, references will be made to the expected effects and deviations between actual influences of the independent variables from a priori expectations will be discussed in more detail in the interpretation of the results.

CHAPTER 4

Descriptive Analysis of Historical Changes and the Current Situation

This chapter first gives a brief introduction to Dak Lak province. The two districts where the research project took place will then be examined in detail using mainly graphical display techniques to present available data and trends therein. A number of comparisons, which are believed to be central to the objectives of this study, are highlighted employing the various data sources described in the previous chapter.

The techniques of explorative data analysis applied in this chapter aim to reveal influences of and relationships between relevant indicators in the research area. The exogenous variables employed in the econometric estimations of chapter 5 and 6 are illustrated including the natural potential, the socioeconomic history of the villages and major policy changes. Land-cover changes are investigated in detail including the changes in the spatial fragmentation of landscape patches. An examination of changes in agricultural productivity, wealth, education and perceptions of the villagers about their natural environment intends to yield descriptive insights into the alterations of indicators related to endogenous outcome variables such as economic growth, human welfare and environmental sustainability .

The descriptions in this chapter are mainly based on GIS data for geo- and biophysical variables. Most socioeconomic indicators are derived from the village interviews of the quantitative survey (see subsection 3.2.3) and from interviews with

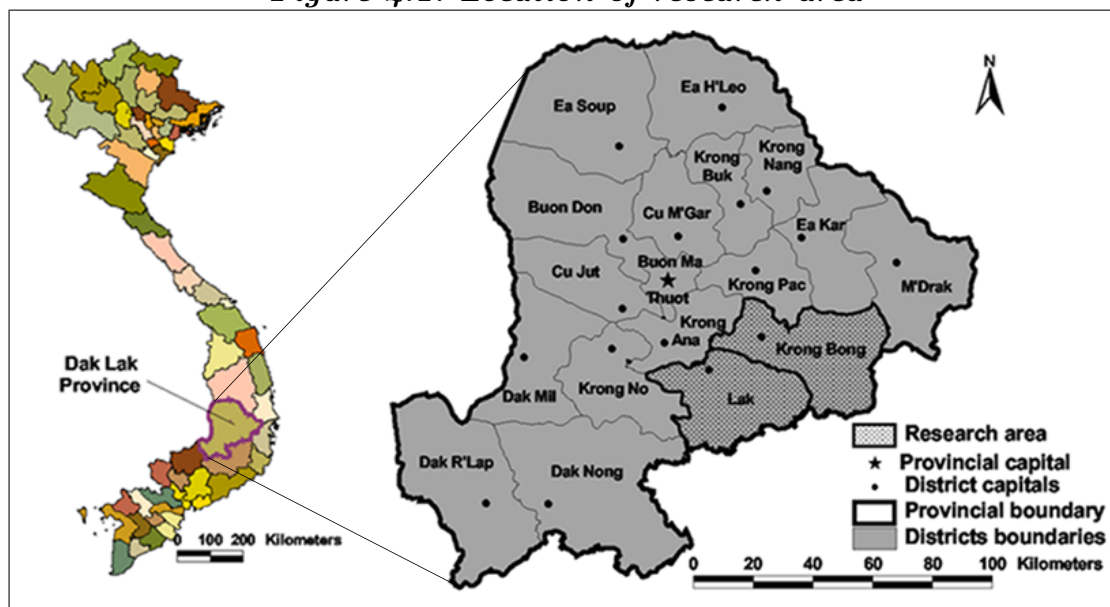
commune and district officials. A focus in the descriptive and explorative analysis is on the differences between ethnic Vietnamese and other indigenous ethnic groups as significant differences in past socioeconomic history and in land use strategies between these groups are assumed to affect influencing factors and outcomes. In addition, some indicators will be highlighted for changing levels of market access, which in many parts of the world have a significant influence on the labor, capital and land intensity of agricultural production.

4.1 Dak Lak province

Dak Lak is the largest province of Vietnam and is located in the Central Highlands. Dak Lak was a part of former South Vietnam and borders Cambodia to the West (see figure 4.1). Dak Lak is the province with the most abundant forest resources in the country. The province has experienced high immigration rates since liberation day in 1976. Yearly population growth averaged 7.7% in the period 1977 to 1990 with growth rates above 10% in the years 1987 to 1989 (Dak Lak Statistical Department, 1991). Between 1991 and 2001, population growth per year dropped slightly to the still substantial growth rate of 6.6% with peaks from 1996 to 1999 (Dak Lak Statistical Department, 2002). In 2001, population was at 1,950,000 for the 19,600 square kilometers, thus resulting in a population density of roughly 100 persons per square kilometer (Dak Lak Statistical Department, 2002).

After liberation day in 1976 Dak Lak had a population of 363,000 of which almost 300,000 were living in rural areas. At that time, provincial population density was at 18.5 persons per square kilometer (Dak Lak Statistical Department, 1981). Since 1990, the population of Dak Lak doubled, especially due to immigration of ethnic Vietnamese from Central and Northern Vietnam (see figure 4.2).¹ In 2000, 1.5 million people are classified as rural, half of the rest live in the provincial capital Buon Ma Thuot. Unfortunately, the author could not obtain official figures on ethnic composition in both the province and the research districts, as these data is not listed in official statistical data sources.

¹Chapter 4 predominantly contains graphs employing data from official maps and statistics as well as from the village survey. Graphs are used as they are better suited to present trends in the data and to make comparisons. In addition graphs save space and facilitate quicker interpretation (Gelman et al., 2002). The version 8.0 of the statistical software package Stata[®] (StataCorp, 2003) provides great graphing facilities, which are exploited for most graphical displays of this chapter to efficiently explore and present the data and research results.

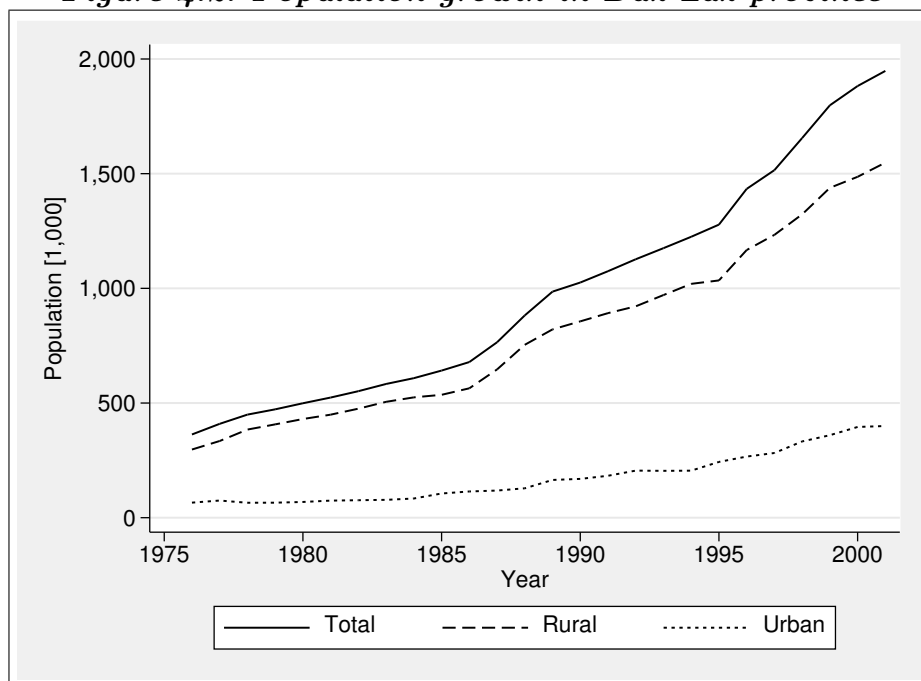
Figure 4.1: Location of research area

Source: Data from Department for Agriculture and Rural Development (DARD), Dak Lak

Dak Lak benefited from rapid economic growth during the last decade. Increasing cash crop production, mainly coffee and rubber and more recently pepper, are the major factors for the growth in export volume of agricultural production. Coffee production, for example, exploded in the last decade making Vietnam the second largest coffee exporter in the world after Brazil with 11% of total world coffee production and the biggest exporter of Robusta coffee.² In 1976, coffee was produced on 20,000 hectares in Dak Lak,³ which increased to 258,500 hectares under coffee production in 2001 (Dak Lak Statistical Department, 2002). Dak Lak's total harvest of 456,000 tons amounts to more than half of the Vietnamese coffee production (Dak Lak Statistical Department, 2002, and VICOFA). However, coffee expansion stagnated since 2000 as prices plummeted from 2,633 USD in 1995 to 436 USD in 2001 for a ton of green beans of Robusta coffee (VICOFA). The basaltic soils suitable for coffee cultivation cover approximately 40% of Dak Lak province (National Institute for Agricultural Planning and Projection, 2000). Particularly where soil conditions

²International Coffee Organization (ICO), www.ico.org.

³Vietnam Coffee and Cocoa Association (VICOFA), www.vicofa.org.vn.

Figure 4.2: Population growth in Dak Lak province

Source: Dak Lak Statistical Department (2002)

allow for big-scale coffee production, forest cover has declined steadily from 75% at liberation day to 64% in 1992 and 51% in 2002.⁴

4.2 Research area

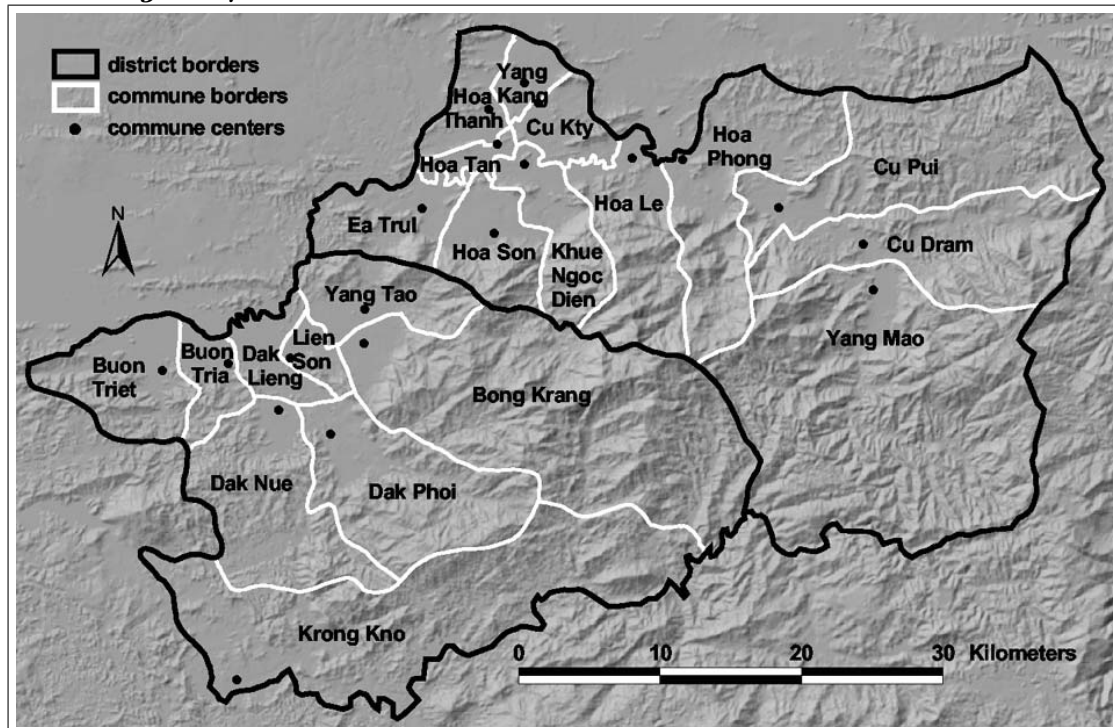
The research area is situated in the southeastern part of Dak Lak province (see figure 4.1) and covers 2,327 km². The two districts border Lam Dong province to the south and Khanh Hoa to the east. They contain 21 communes as shown in figure 4.3 with 191 villages. Communes are the smallest administrative units in Vietnam with communal People's Committees as executive organs of local state administration (*The Socialist Republic of Vietnam, 1992*).⁵ The center of Lak district is Lien Son and Krong Bong's center is Khue Ngoc Dien.⁶

⁴Data from the Forest Development Department (FDD) and Forest Protection Department (FPD) of Dak Lak province.

⁵The responsibilities of the PCs include the implementation of the constitution, the law, the formal written orders of superior state organs and the resolutions of the People's Council (*The Socialist Republic of Vietnam, 1992*).

⁶Khue Ngoc Dien was recently divided and the district center is now called Krong Kmar.

Figure 4.3: Research area with administrative boundaries



Source: Department for Agriculture and Rural Development (DARD), Dak Lak, and DEM from Mekong River Commission (MRC)

4.2.1 Geophysical indicators

Figure A.2 in the appendix illustrates the climatic conditions in the research area. The graphs are based on daily rainfall time series since 1976 and monthly temperature data since 1987. Average temperature in the last 25 years was 24 degrees with a minimum of eight and a maximum of 38 degrees. The warmest month is April with a monthly average of 26 and the coldest is January with 20 degrees on average.

The rainy season usually lasts from May until October with the highest risks for floods in September and October.⁷ Average yearly rainfall since 1976 was 1,655 millimeters. The upper left part of figure A.2 in the appendix with the yearly averages shows the high variability of rainfall in the research area indicating the high output uncertainty, which farmers face. To account for this variability in rainfall, the variance of rainfall over time is included in the empirical estimations in chapters

⁷In 2001 several people died in the two research districts due to heavy flooding.

5 and 6. The unusually low rainfalls in 1991 were also reported by farmers in the village interview.

The topography in the research area is illustrated in figure 4.3 and indicates that commune centers are situated in the plains on elevations below 450 meters above sea level. Elevation in the entire area ranges from 419 to 2,415 meters with a mean of 807 meters.

Comparing agricultural land uses for different elevations reveals that significantly more agricultural cultivation can be noticed at higher elevations in 1992 than in 2000 (table 4.1). In 1992, 19% of agricultural land use is concentrated on elevations above 500 meters compared to only 8% in 2000, another manifestation for higher land-use intensity on more marginal lands one decade ago. Virtually no agricultural cultivation is found above elevations of 550 meters in the year 2000.

Table 4.1: Agricultural land use and elevation

Cumulative elevation [m]	Agricultural land use			
	km ²		Percent	
	1992	2000	1992	2000
Above 600	31.8	0.8	7	0
Above 550	44.4	3.0	9	1
Above 500	89.8	35.1	19	8
Above 450	193.8	165.5	41	37

Source: DEM from MRC and satellite image interpretation by Nguyen Thi Thanh Huong and author.

Mean slope in the research area is 13 degrees with a maximum of 53 degrees. Interestingly, agricultural land uses in 2000 are much more concentrated on less sloping land compared to 1992 and 1975 (figure 4.4, 4.5 and 4.6; for a description of the land-cover categories see section 3.2.1 on page 42).

In all three figures the highest slope category is mostly covered by closed canopy forest (C). In 1975 and in 2000 slopes above ten degrees were almost entirely covered by open (O) and closed canopy forests. In 1992 a considerably higher amount of mixed agriculture (M) and grass land (G) is observed on slopes above five degrees. Again, this points to agricultural cultivation on more marginal lands in 1992, which might also be an indication for the existence of rotational farming practices like

shifting cultivation. Most of the cultivated land on high elevations and steep slopes apparently recovered into forest between 1992 and 2000.

Figure 4.4: Slope categories by land-cover class, 2000

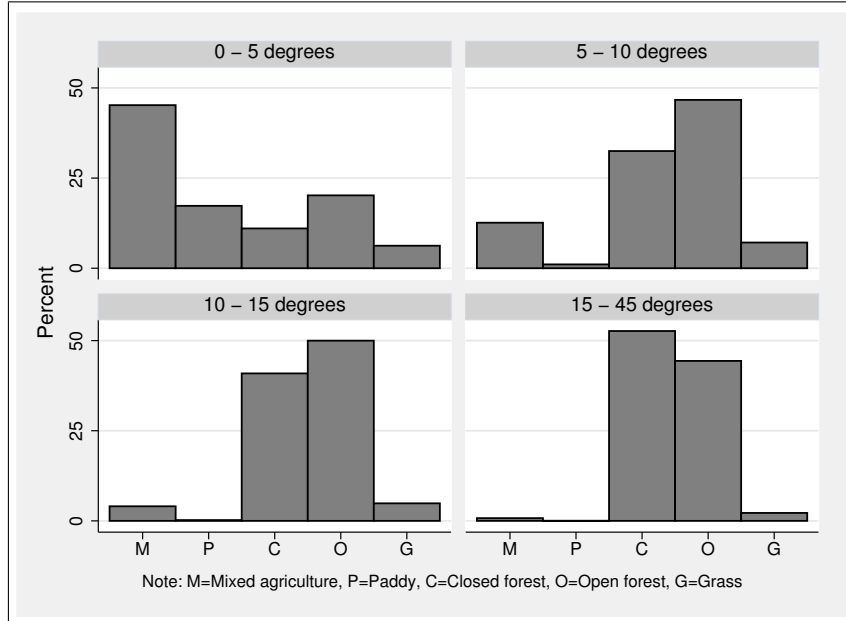
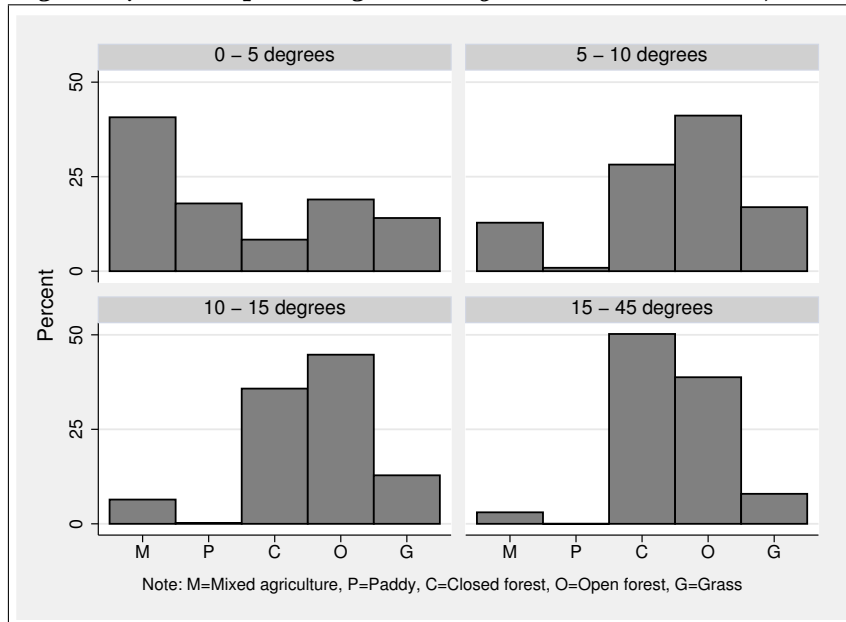
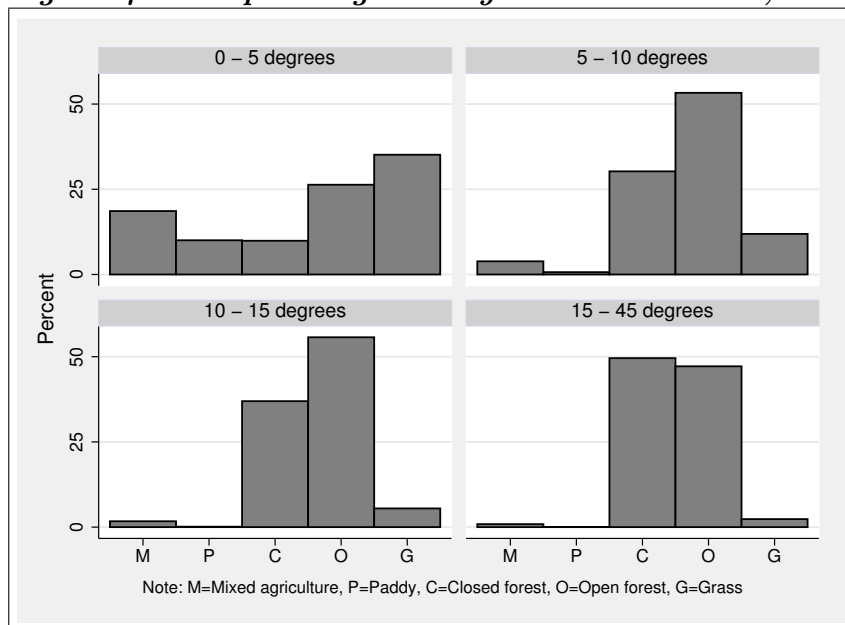


Figure 4.5: Slope categories by land-cover class, 1992



Source: DEM from MRC and satellite image interpretation by Nguyen Thi Thanh Huong and author.

Figure 4.6: Slope categories by land-cover class, 1975

Source: DEM from MRC and satellite image interpretation by Nguyen Thi Thanh Huong and author.

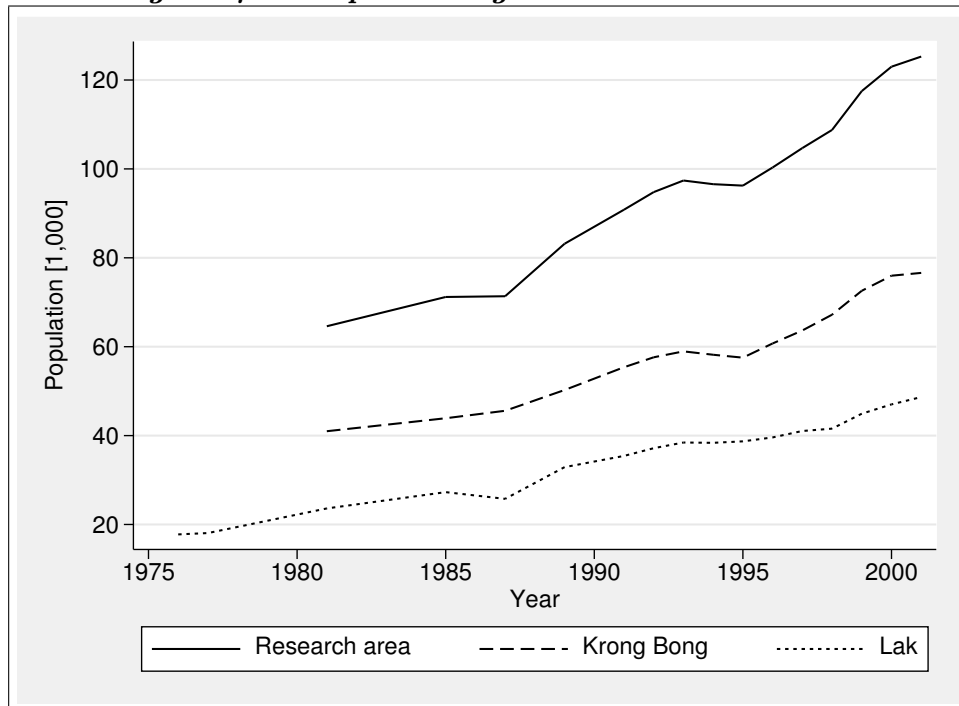
4.2.2 Socioeconomic history

According to official statistics population almost doubled in the research area from 64,000 in 1981 to 125,000 in 2001 (see table 4.2 and figure 4.7, data from [Dak Lak Statistical Department, 2002](#)). According to village interviews, population in the survey villages increased by more than 300% over the last 20 years and by 150% in the last decade. Population growth was higher in Krong Bong district, especially in the last five years. Most of this increase is presumably due to uncontrolled immigration from outside the province although no consistent data on both government controlled and uncontrolled immigration could be obtained.⁸ According to the commune survey roughly 20% of the population growth are estimated to be due to government-controlled resettlement programs into areas delineated as New Economic Zones (NEZ) following the Fixed Cultivation and Resettlement Program (FCRP). Controlled migration was presumably also higher in Krong Bong, which

⁸In contrast to government controlled migration, the phenomenon called *uncontrolled migration* in Vietnamese terminology refers to migration patterns outside the control of resettlement programs administered by the government.

has more NEZs where the government settled migrants.⁹ Uncontrolled immigration was strongest during the last decade when high coffee prices further stimulated migration from Central and Northern Vietnam to the Central Highlands.

Figure 4.7: Population growth in research area



Source: [Dak Lak Statistical Department \(2002\)](#)

Impoverished and landless migrants from lowland regions might have followed their resettled families, neighbors or friends, who were moved to Dak Lak in the planned migration programs. The research area did not participate as much as other districts of the Central Highlands in the *coffee boom*, because only small areas of the two selected districts have the basaltic soils suitable for large-scale coffee production. Therefore, immigration was lower in the research area than in the rest of the province. Nevertheless, the average population growth in the research area of 5% per year over the last two decades is still considerable ([Dak Lak Statistical Department, 1981, 1991, 2002](#)).¹⁰

⁹The figures on immigration are based on personal communication with government officials from Dak Lak province.

¹⁰Krong Bong district was established in 1981. Before 1981 no statistical data exist for Krong Bong district.

Table 4.2: Population and area of Lak and Krong Bong districts, 2001

Name of district	Area [km ²]	Inhabitants	Inhabitants/km ²
Krong Bong	1,250	76,609	61
Lak	1,077	48,668	45
Total	2,327	125,277	54

Source: [Dak Lak Statistical Department \(2002\)](#)

Population density is higher in Krong Bong with 61 persons per km² than in Lak with 45 persons per km² (table 4.2). Subtracting areas, which are protected or on slopes steeper than 15 degrees, population density amounts to 107 persons per km² of land potentially suitable for agricultural production.

Population per commune varies between 3,000 and 12,000 persons and population densities between 0.11 and 1.53 per hectare (table 4.3). More important in rural areas, where the population almost entirely depends on agricultural production, is the population density per agricultural area. Table 4.3 reveals that population density per hectare of agricultural land ranges from 0.64 (Krong Kno) to 5.62 (Lien Son). Krong Kno in Lak district (see figure 4.3) is the commune, which is most difficult to access. It is also the commune with one of the lowest wealth levels, the highest proportion of food shortages and of people from ethnic minority groups. Apparently, Krong Kno did not benefit as much as other communes in the improvements of rural development indicators.

Ethnic composition in the 101 interviewed villages remained relatively stable between 1980 and 2000 (see figure 4.8). The *Mnong* are indigenous to the research area while the *Ede* are indigenous to Dak Lak province, but originate from areas further to the North. The group “Others” in figure 4.8 consists of ethnic minority groups, which immigrated from other parts of Vietnam, most notably the *Tay*, which are the biggest ethnic minority group in Vietnam. Figure 4.8 illustrates that the share of the *Mnong* dwindled from 39% to 31% while the share of the *Kinh* increased by 6% in the last two decades. A large number of the *Kinh* immigrated after the end of the war in 1975 and before 1980, a period with intensive resettlements to the NEZs of the research area. Data from our commune interviews reveal that on average more than 200 Vietnamese households immigrated to each commune in that

Table 4.3: Commune population and population density, 1999

Commune	District	Population	Population density per hectare	
			total area	agricultural area
Lien Son	Lak	5,382	1.53	5.62
Yang Tao	Lak	5,653	0.44	2.81
Dak Phoi	Lak	5,990	0.21	2.67
Dak Nue	Lak	3,744	0.15	3.01
Krong Kno	Lak	3,105	0.11	0.64
Buon Triet	Lak	5,781	0.38	2.13
Dak Lieng	Lak	8,118	1.22	3.15
Buon Tria	Lak	2,945	0.47	2.02
Bong Krang	Lak	4,843	0.07	2.89
Khue Ngoc Dien	Krong Bong	12,007	0.97	5.31
Hoa Son	Krong Bong	8,549	0.73	3.24
Ea Trul	Krong Bong	9,334	0.79	2.89
Cu Kty	Krong Bong	4,278	0.62	1.73
Hoa Tan	Krong Bong	2,575	0.91	2.07
Hoa Thanh	Krong Bong	3,812	0.48	1.23
Hoa Le	Krong Bong	6,604	0.34	3.15
Hoa Phong	Krong Bong	5,052	0.17	1.79
Cu Pui	Krong Bong	5,594	0.15	3.21
Cu Dram	Krong Bong	5,161	0.18	3.85
Yang Mao	Krong Bong	3,668	0.04	3.64
Yang Kang	Krong Bong	5,034	1.15	2.39

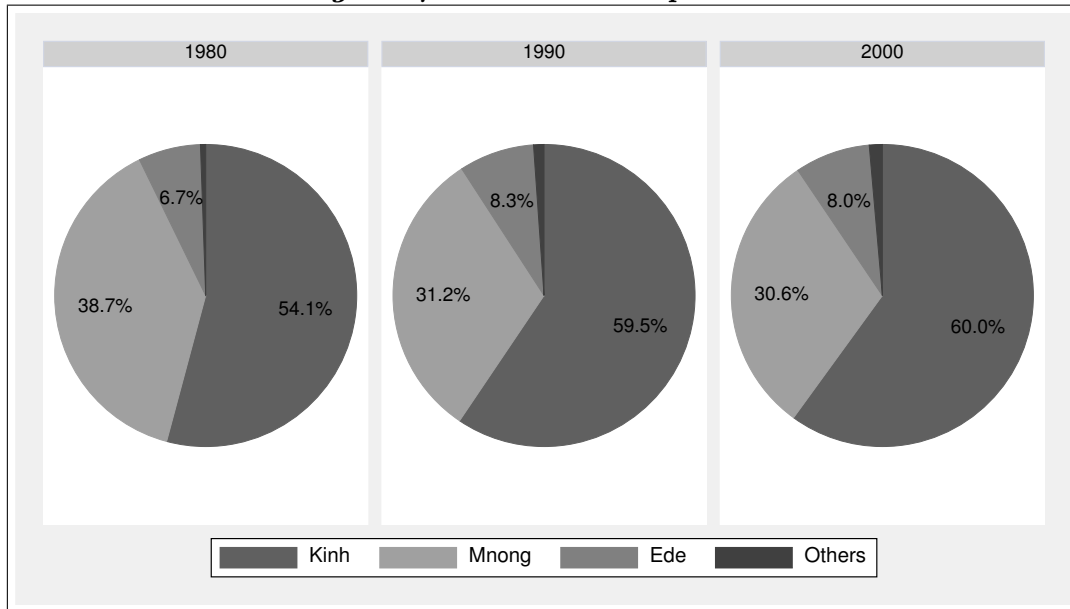
Source: [Lak District Statistical Office \(1999\)](#); [Krong Bong District Statistical Office \(1999\)](#) and interpretation of Landsat satellite images by Nguyen Thi Thanh Huong and author.

period. Before 1975 the share of the *Kinh* was probably considerably lower in the research area.¹¹ In total, 53 out of the 101 survey villages are ethnic villages with an average of 84 households per village while an average *Kinh* village is inhabited by 111 households.

4.2.3 Policy review

The Fixed Cultivation and Resettlement Program (FCRP) started in 1968 with following objectives (cited from [Vuong Xuan Tinh and Bui Minh Dao, 2000](#)):

¹¹No official data on that matter are available and this estimation stems from personal communication with district and provincial officials.

Figure 4.8: Ethnic composition

Source: Village survey in Lak and Krong Bong districts, 2000/2001

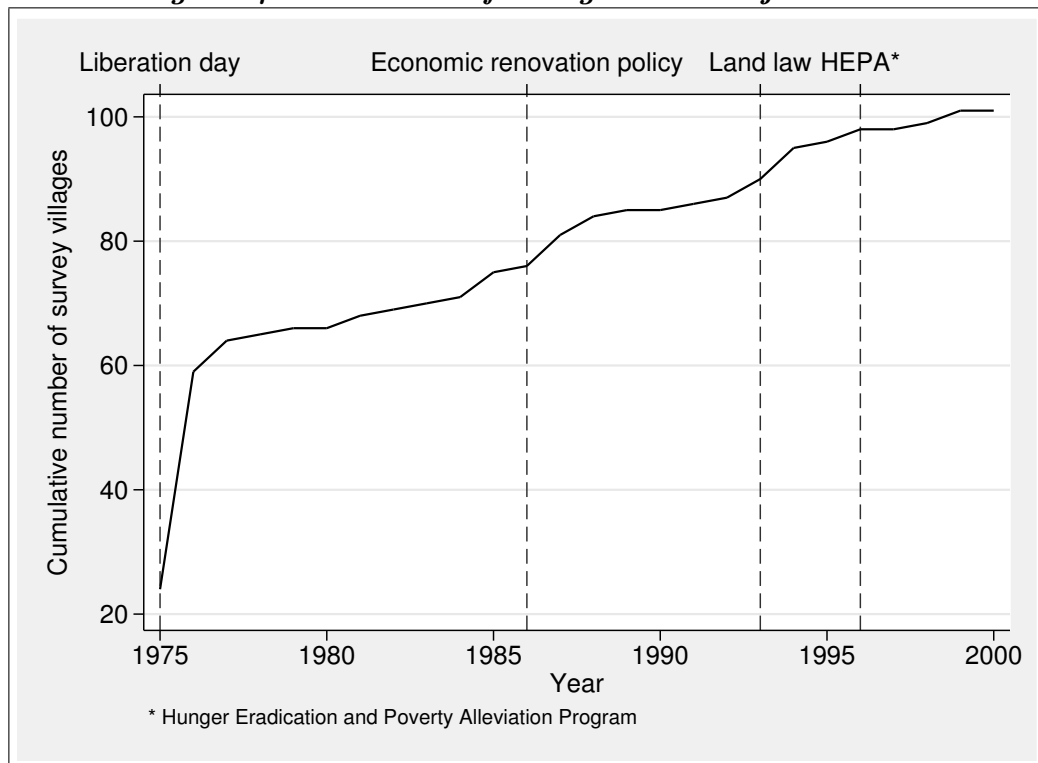
1. Ending nomadic life and farming to protect forest resources;
2. Eradicating hunger and alleviating poverty as well as an improvement of the living conditions of ethnic farmers;
3. Consolidating the agricultural collectivisation in ethnic minority areas.

Communal ownership was banned by the communist government after the war. The areas designated for shifting cultivation by the indigenous ethnic groups were handed to the state administration. Numerous ethnic villages were resettled to lowland valleys. After liberation day coffee and rubber plantations were nationalized and more state farms and state forestry enterprises established. According to [Vuong Xuan Tinh and Bui Minh Dao \(2000\)](#) 450,000 persons were resettled to the NEZs in the three provinces of the Central Highlands between 1976 and 1980 and another 260,000 from 1981 between 1988.¹² Unfortunately, no official data on controlled immigration could be obtained for the research area. However, the increase in the number of villages (see figure 4.9) and the population increase during this period

¹²Other sources talk of nearly one million ethnic Vietnamese who settled in the Central Highlands in the last decade ([Johnson & Brieng, 2001](#)).

(see figures 4.2 and 4.7) suggests the occurrence of significant controlled immigration, also to the NEZs in the research area.

Figure 4.9: Number of villages and major events



Source: Village survey in Lak and Krong Bong districts, 2000/2001

In 1986 Vietnam opened its economy with the implementation of the Economic Renovation Policy (*doi moi*), which was issued by the central government. According to [Vuong Xuan Tinh and Bui Minh Dao \(2000\)](#) a main reason for the introduction of the economic renovation policy was low labor productivity resulting from inadequate central planning:

“The administration and use of agricultural land in the mechanism of centralized planning, state subsidies and bureaucracy in the past did not tap the labor potential of the people. That was the main cause leading to low labor productivity and chronic food shortages in both the deltas and mountainous regions.”

The erosion of the collective agricultural system was initiated by *Decree 100*, which allowed households to exercise a limited control over labor and land. Other aspects

in the production process still remained under collective control (Sikor, 1999). The Decree 100 thus gave people the right to use, but not to own the land.

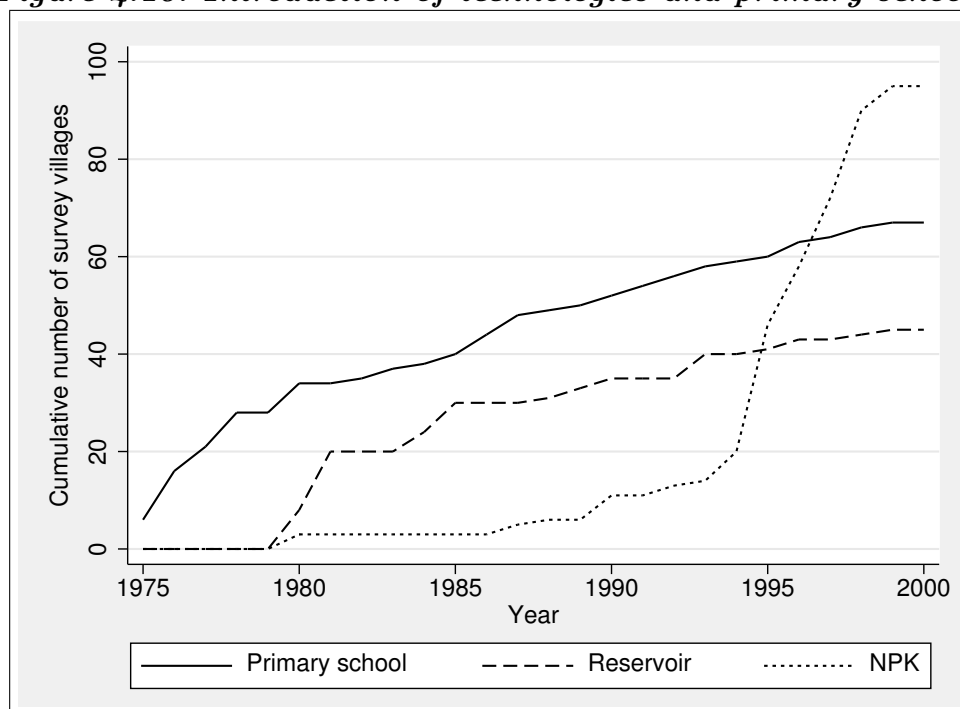
Resolution 10, issued by the Central Party Committee in 1988, further shifted control over productive assets to households, thereby effectively ending the cooperative agricultural system and paving the way for a reorganization of agricultural production. In 1993 the *Land Law* was promulgated to provide agricultural producers with stable and long-term rights to exchange, transfer, lease, mortgage and inherit agricultural land (the so-called *five rights*). The Land Law aimed at providing incentives for more productive use of the land by transferring the responsibility of land management to individuals, households, groups, organizations and institutions (Christ & Kloss, 1998). The land-use right certificates (LURC, also called *Red Books* due to the color of the paper they are written on) grant the allocation of land to households for 20 years for annual crop production and 50 years for perennials. In 1994, the Decree 64 extended the land law to the allocation of forest land.

It was not until 1995 when the first households in the research area obtained their *Red Books*. At the time the village survey was carried out, approximately 65% of the households in the interviewed villages have already obtained their *Red Books* for long-term use of agricultural land. This low figure is mainly a result of the unclear land ownership situation due to the virtual absence of consistent land use planning activities. Besides, the inferior standard of technical equipment contributed to the slow issuance of the *Red Books* as well as the limited administrative capacities in district and commune offices. The actual certificates are issued in the district People's Committees. For this study, the impacts of this relatively recent devolution of secure property rights for agricultural and forest land are of little measurable importance for land-use changes as regressions showed. This is due to the temporal and spatial resolution of the land cover data, which does not allow to detect changes in agricultural crop composition. More recent satellite images would be required to draw inferences from tenure changes on changes in land use.

Another government program, which focused on assistance for the rural poor, was the Hunger Eradication and Poverty Alleviation (HEPA) Program (see figure 4.9), initiated in 1995/1996. The HEPA program is a nationwide program under the mandate of the Ministry of Labor, Invalids and Social Affairs (MOLISA, People's Committee of Dak Lak Province, 1996). Loans are provided to families classified

as hungry and poor by the Vietnamese government.¹³ The HEPA program reached 25% of the interviewed ethnic minority households and 15% of the ethnic Vietnamese households mainly with subsidized credit, inputs and tree seedlings as well as with support in kind. According to survey data the HEPA program was active in 90% of the villages where households suffer from food shortages at least part of the year. Within these villages every second household was reached by the program.

Figure 4.10: Introduction of technologies and primary schools



Source: Village survey in Lak and Krong Bong districts, 2000/2001

The years since the introduction of NPK, the construction of major reservoirs and a primary school for the respective village is sketched in figure 4.10. Striking in figure 4.10 is the rapid adoption of NPK fertilizers in the 90s, especially after 1995. More gradual, but still evident, is the increase in government investments in the construction of schools and reservoirs.

The Euclidean distance to four road categories are depicted in table 4.4 for *Kinh* and ethnic minority villages. The road categories are based on field experience and numerous travels in the area (see also section 3.3.1 in chapter 3). All villages

¹³ Again, no official data could be obtained for the wealth classification as well as for the cash flow to rural households from the HEPA program.

Table 4.4: Average Euclidean distance to road categories

Distance [m] to... ^a	Kinh villages	Ethnic minority villages
...nearest road	97	99
...nearest seasonal road	337	226
...nearest all-year road	2,410	4,367
...nearest tarred road	5,485	11,428

Source: Village survey in Lak and Krong Bong districts, 2000/2001, and road network

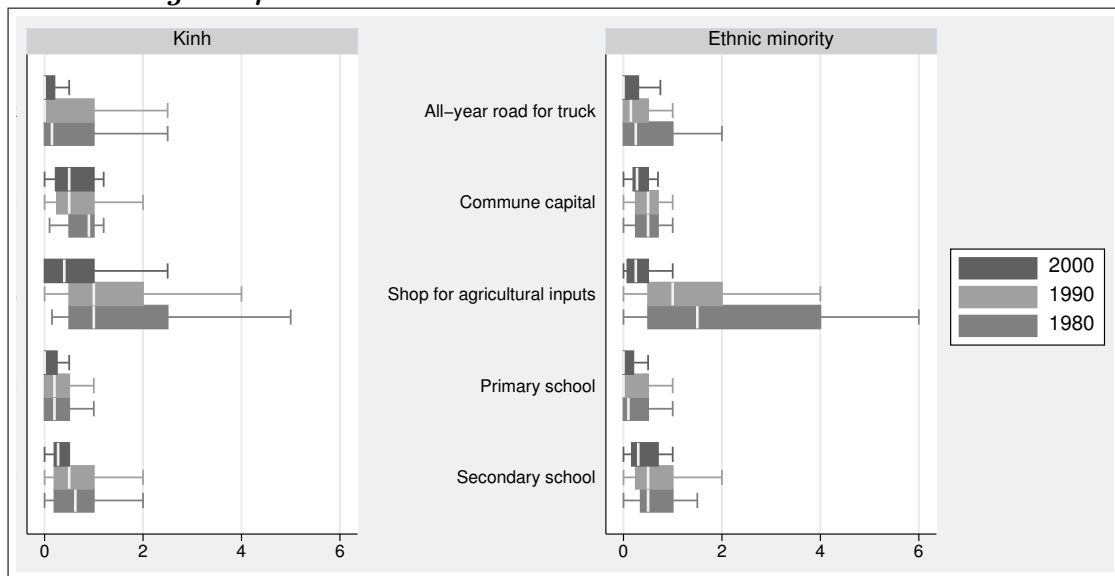
^aNote: Road categories are in descending order of road quality. Upper categories include lower categories, i.e. the calculation of the distance to the nearest all-year road includes the distances to the nearest tarred road.

irrespective of ethnic group are located on average 100 meters away from the nearest road. Differences in access to roads become evident when different road categories are accounted for. On average, *Kinh* villages are located almost six kilometers closer to the nearest tarred road in 2000. Especially in Krong Bong district many *Kinh* villages were established after liberation day along the main roads built by the French before the war. Ethnic villages, on the other hand, are situated closer to roads, which are not accessible throughout the entire rainy season.

In the village survey access to markets, services and political centers is measured by the travel time for villagers to reach these infrastructural facilities. In figure 4.11 the travel time to the nearest all-year road, to the communal centers, to the closest shop for agricultural inputs and to the nearest primary as well as secondary school is depicted. Reported travel time to the nearest all-year road decreased from a mean of 45 to 20 minutes in the last two decades. Villagers clearly have better access to roads today, which to some extent is due to improved means of transport like tractors and motorbikes, but also to significant road upgradings. Remarkable in figure 4.11 is the decrease in travel time to the nearest shop for agricultural inputs from an arithmetic mean of two hours to 35 minutes. Nearly all villages are now located close to primary and secondary schools, which were accessible at a mean travel time of 20 and 50 minutes 20 years ago. By 2000, travel time to schools decreased by 15 and 30 minutes, respectively. Variation between villages declined almost consistently for all infrastructural facilities and differences across ethnic groups in figure 4.11 are small. This seems to contradict the evidence presented in table 4.4. The author suspects

that, despite thorough explanations in the village interviews, the understanding of the term “all-year road” might carry a measurement error due to cultural influences on the degree of understanding (see also section 3.6 for a concise discussion on data limitations). The impression that the road network created with the GIS software is more consistent than the survey results for estimated travel time resulted in the use of the continuous GIS surface for the *distance to the nearest all-year road* in the econometric estimations presented in chapter 5 and 6.

Figure 4.11: Travel time in hours to various locations



Source: Village survey in Lak and Krong Bong districts, 2000/2001

The two existing *nature reserves and protected areas* comprise 15% of the total research area. *Nam Ca* was established in 1986. The management board and investment plan was approved by the Dak Lak Provincial People’s Committee in 1991 when the total area of the nature reserve comprised 24,555 hectares of which 11,410 are within the research area. The other reserve, *Chu Yang Sin*, was established as a 32,328 hectares nature reserve in 1994, of which 19,401 hectares are allocated for strict protection and 12,907 hectares for forest rehabilitation (Birdlife International, 2001). However, the provided map of nature reserve areas indicate the protected area of *Chu Yang Sin* as 22,291 hectares. This map is used for subsequent calculations. Consequently, about 14% of the research area belongs to nature reserves and protected areas designated for forest rehabilitation and biodiversity conservation. Stricter enforcement of forest protection in recent years might have facilitated

a restoration of forest cover in these areas to a fairly homogenous, closed canopy cover.

4.3 Outcome Variables

In this section the outcome variables are explored as described in the conceptual framework (section 2.1 and figure 2.1). Land-cover changes in the research area will be explored including measures of **landscape** metrics. Selected outcome indicators of agricultural development, data on economic growth and environmental indicators will be presented in detail.

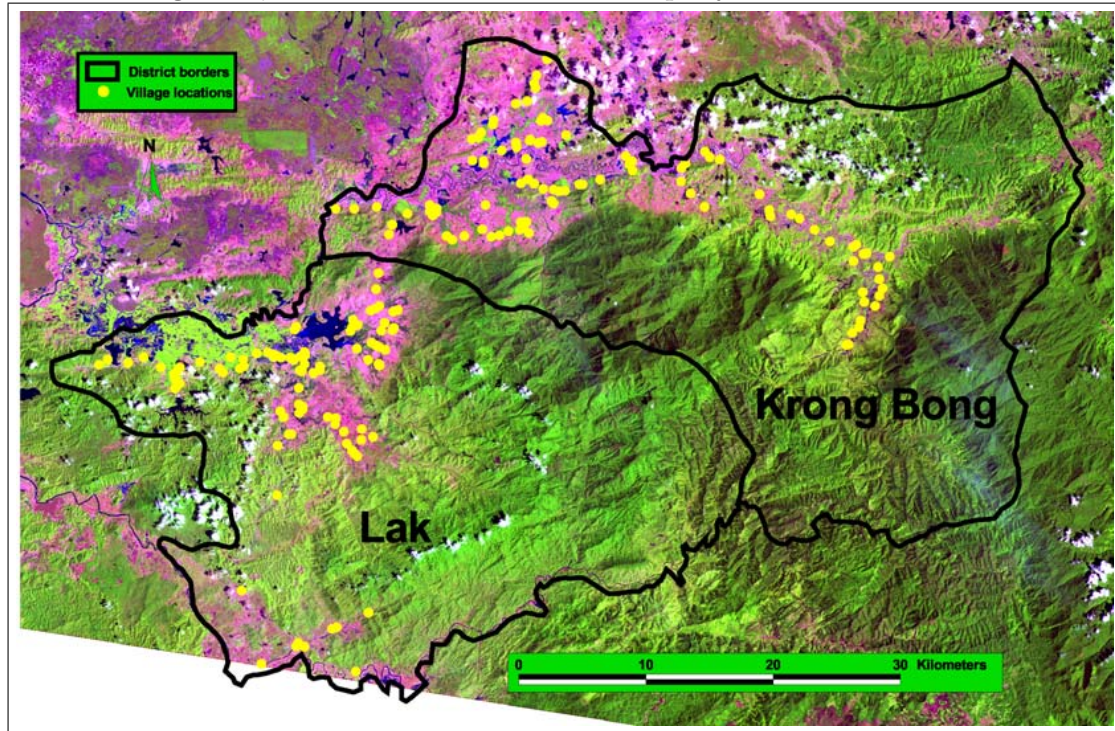
4.3.1 Land-cover changes

Countrywide forest cover was estimated at 30% for 2000 (World Bank, 2002c).¹⁴ In the same year, 77% of the research area was covered by forest of different qualities (figure 4.12).¹⁵ At lower elevations and in locations close to villages and roads, bush land, open canopy, and bamboo forests prevail. Mountainous evergreen forests dominate areas with altitudes above 600 meters. Bush and bamboo forest as well as young open canopy forest are found as secondary vegetation on areas formerly used for upland cultivation.

The land-cover maps of 1975, 1992, and 2000, derived from the satellite image interpretation (see the section 3.2.1 on land-cover data collection), are depicted in figure A.1 of the appendix. Agricultural land increased by 9% from 1975 to 1992 and has stagnated since then (see table 4.5). Agricultural area mainly developed from open canopy forest and grass in that period. Mixed grassland decreased in the second period from 11% to 4%. Paddy area also increased between liberation day and 1992 and stagnated in the last decade. Interesting is a closer look at total forest cover, which declined by 10% in the first period, but grew by 8% since 1992. This growth in the last ten years is accompanied by a decrease in mixed grass land. Thus, no severe deforestation took place in the research area over the last 25 years with a decline in total forest cover (closed and open canopy forest in table 4.5) of about 2%.

¹⁴Forest classification differs widely in Vietnam depending on the institution, which conducts the classification. Which data source, and therefore classification, the World Bank utilizes to arrive at the cited figure for country-wide forest cover is not stated in the *World Development Indicators*.

¹⁵All land-cover data for the research area stem from the interpretation of Landsat satellite images from Nguyen Thi Thanh Huong from Tay Nguyen University, Buon Ma Thuot, and the author.

Figure 4.12: Research area on top of Landsat ETM+

Source: Landsat ETM+ satellite image (false color composite with band combination 5-4-3) and village positions (yellow dots) collected by author and colleagues

Land-cover changes are often reversible and follow certain time sequences of successive cover types.¹⁶ *Land-cover trajectories* refer to such generic paths of land-cover changes over more than two observations in time (Mertens & Lambin, 2000). Figure 4.13 illustrates these trajectories in the research area by combining the three land-cover maps in figure A.1 of the appendix. Land-cover data for this figure is aggregated into two classes:

- ▷ Agricultural land uses consisting of mixed agricultural land and paddy;
- ▷ Other land-cover classes comprise open and closed canopy forest as well as mixed grass land.

For the subsequent discussion of land-cover trajectories in figure 4.13 the category of other land will be referred to as forested land, because, abstracting from anthro-

¹⁶As the econometric analysis (chapter 5) in this study is done in a static fashion, sequences of land-cover changes are not included as an exploratory variable.

Table 4.5: Land cover in km^2 and percent of total land

Cover class	1975		1992		2000	
	km^2	%	km^2	%	km^2	%
Mixed agriculture	146.7	6.1	352.0	14.6	339.8	14.2
Paddy	67.2	2.8	117.0	4.9	111.0	4.6
Closed forest	836.2	34.9	828.2	34.5	866.6	36.4
Open forest	1052.3	44.0	827.9	34.6	962.6	40.5
Grass	295.7	12.2	273.5	11.4	101.5	4.3
Total	2398.2	100	2398.6	100	2381.6^a	100

Source: Interpretation of Landsat satellite images by Nguyen Thi Thanh Huong (Tay Nguyen University, Buon Ma Thuot) and author for Lak and Krong Bong districts, Dak Lak Province, Vietnam.

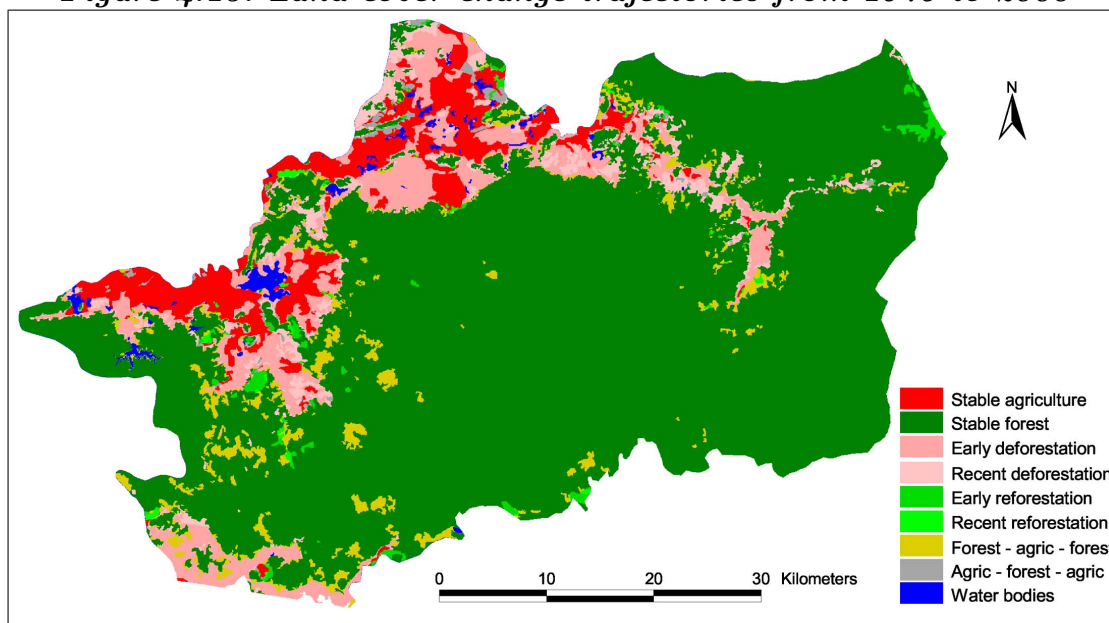
^a17 km^2 missing in 2000, because part of the research area was not covered by the Landsat ETM+ scene of 2000.

pogenic influences on vegetation cover, most of it would regenerate into forest in a relatively short period of time.

A clear pattern can be observed for agricultural expansion, which started from the lowland areas with suitable soil conditions for paddy and mixed agricultural production (northern, western and southwestern parts) and close to district centers (see figure 4.3 for the location of the district centers). Areas where forest regenerated in the last decade are typically further away from major roads and on steeper slopes. The isolated patches of land, which were converted to agriculture after liberation day and partly regenerated into open canopy forest in the last decade, probably belonged to upland shifting cultivation fields. Relatively few of these areas regenerated into forest and were then converted again for agricultural cultivation. Given our land-cover data we only observe very little evidence of traditional fallow-based farming systems.

The biggest areas of agricultural expansion are detected in locations with relatively good soil conditions and adequate market access. Most roads leading to these regions were built by the French long before liberation day in 1975. Available road access to these regions has, together with favorable natural conditions, continued population growth and an increased potential of market-oriented production, stimulated land expansion especially in the 15 years after liberation day.

Figure 4.13: Land-cover change trajectories from 1975 to 2000



Source: Interpretation of Landsat satellite images by Nguyen Thi Thanh Huong and author.

Another notable feature from the land-cover maps in figure A.1 of the appendix are the spatial patterns of land fragmentation. Clearly visible is the high spatial fragmentation for the year 1992. Land-cover fragmentation can be measured as the *number of patches* under a specific land cover (table 4.6). *Landscape* metrics can be used to explore and quantify the spatial patterns of a landscape structure. It is based on the recognition that ecological processes are likely to be affected by the interspersion and juxtaposition of patches within a landscape (McGarigal & Marks, 1995).¹⁷ The measures of landscape metrics illustrated in table 4.6 include the number of patches, the mean patch size and the mean perimeter-area ratio. These indicators are computed for individual patches, then summarized for the whole landscape, i.e. the research area.

Landscape fragmentation in the last eight years decreased dramatically from 614 patches of land cover in 1992 to 357 in 2000.¹⁸ Most of that change is due to the high fragmentation of agriculture and grass in 1992 compared to 2000. This reflects the

¹⁷Landscape metrics are calculated using the freeware program FRAGSTATS for ArcView[®] available at www.umass.edu/landeco/research/fragstats/fragstats.html.

¹⁸Landscape statistics in table 4.6 are only calculated for the images of 1992 and 2000, which have the same spatial resolution of 30 by 30 meters (see section 3.2.1).

Table 4.6: Landscape metrics

<i>Cover class</i>	Number of patches		Mean patch size		Mean perimeter-area ratio	
	1992	2000	1992	2000	1992	2000
Mixed agriculture	287	107	123	317	280	214
Paddy	44	54	265	206	158	133
Closed forest	10	8	8,216	10,831	292	26
Open forest	59	41	1,403	2,280	284	211
Grass	214	147	128	69	160	162
Total/average^a	614	357	2,027	2,741	235	149

Source: Interpretation of Landsat satellite images by Nguyen Thi Thanh Huong and author.

^aThe total number of patches as well as the averages of the mean patch size and the mean perimeter-area ratio are calculated.

virtual disappearance of shifting cultivation in the study region, until today often wrongly blamed responsible for large-scale deforestation. Table 4.6 also exhibits the *mean patch size*, which shows that agricultural patches are significantly bigger in 2000 while the (relatively few) grass patches are smaller. At both points in time there are few, but large patches of closed forest (see also table 4.5). The *mean perimeter-area ratio* in table 4.6 is a measure for shape complexity and calculated as the sum of the ratio between the perimeter of each patch and its area divided by number of patches (McGarigal & Marks, 1995). Square objects¹⁹ have the smallest perimeter-area ratio while skinny, fragmented and elongated object have larger values. In table 4.6 the land cover in 1992 has the highest complexity for most land-cover classes and on average shapes are substantially more complex in 1992.

The spatially explicit measures of landscape metrics presented here at the level of individual patches reveal that – given the data at hand – land-cover fragmentation in 2000 is more homogeneous, less complex and less patchy. This could give rise to possible benefits for the ecological functions of the forests as well as potentially better economic prospects as fields are less dispersed, less fragmented and closer to the populated areas.

¹⁹Or (approximately) circular objects in the case of polygon analysis. The analysis presented here is done for square grid cells.

4.3.2 Agricultural development indicators

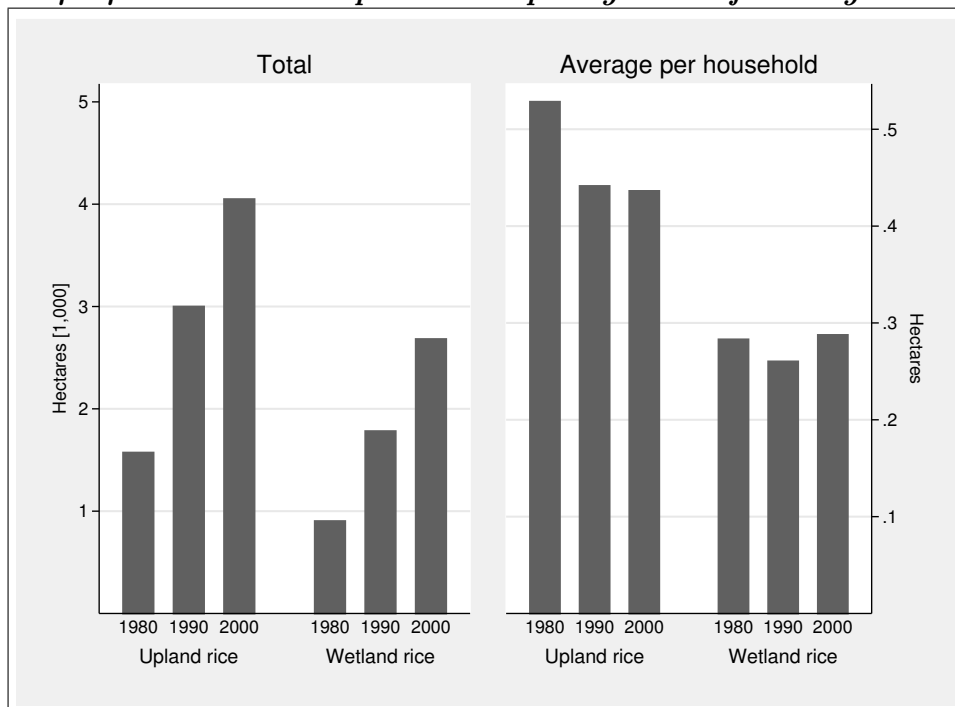
Agriculture is the main source of income in the research area, with paddy being the most important crop. Ethnic Vietnamese (*Kinh*) and the indigenous ethnic groups of the research area traditionally have different farming systems. Indigenous ethnic groups practice sedentary shifting cultivation while the *Kinh* have a long tradition of intensively growing wetland rice and marketing surplus production to local traders. *Kinh* villages have a tendency to be located closer to better roads (see table 4.4) and they usually market a bigger share of their agricultural production.

In the nineties, the adoption of new crops (particularly maize and coffee), fertilizers (Urea and NPK) and pesticides improved agricultural productivity. Today, the second most important crop is maize, with hybrid varieties cultivated by 65% of households in the study area.²⁰ Increasing market prices turned maize into a major source of cash income in the last five years, especially for households which are not able to produce rice above subsistence needs and lack the suitable natural conditions for coffee production.

The area allocated to permanent agricultural cultivation as reported from the surveyed households is graphed in figure 4.14. It is important to note that the questions focused on *permanent* fields, thereby excluding shifting cultivation plots.²¹ Wetland rice is intensively cultivated in the lowlands with high applications of fertilizer, pesticides and high labor input. Upland fields are mainly allocated to maize and, to a lesser extent, to sugarcane, cassava and beans. Total permanent area cultivated on upland and lowland fields increased more than twofold in the last two decades according to the interviewed households. Average upland area per household stagnated in the last decade after decreasing by almost 20% between 1980 and 1990. Wetland rice cultivation per household in 2000 was similar to the levels in 1980 after a slight drop in 1990. On average one household cultivated a little less than 0.3 hectares of paddy rice and 0.42 hectares of upland fields in 2000. These figures correspond fairly well to the land-cover data derived from the satellite images (see the section 4.3.1 on land-cover changes).

²⁰If not otherwise specified all data for the research area is derived from the village survey in Lak and Krong Bong districts conducted by the TÖB research project.

²¹As the Vietnamese government issued several policies to discourage shifting cultivation most villagers felt uncomfortable with questions concerning this farming practice. That being the case, the author suspects that answers concerning shifting cultivation were largely doubtful and, consequently, are not reported.

Figure 4.14: Permanent upland and paddy area of survey households

Source: Village survey in Lak and Krong Bong districts, 2000/2001

Higher land-use intensity is also manifested by the number of interviewed households, which harvest their paddy fields two times per year. For 1980, 1,500 reported two paddy harvest per year, while 3,500 did so in 1990 and 4,900 in 2000. This is equivalent to 35% (1980), 48% (1990) and 53% (2000) of all households in the interviewed villages. Reasons are improved irrigation channels, better technologies to pump water and investments in the construction of reservoirs by the external sources.

According to farmer's estimates, agricultural productivity per hectare of paddy rice increased over the last 20 years from a mean of 3.3 to 3.9 tons per hectare (t/ha) while the median of paddy yield rose from three to four t/ha (figure 4.15). These figures correspond to official statistics for the two districts, which estimate yields at 4 t/ha in 2001 (Dak Lak Statistical Department, 2002). However, there are significant differences across ethnic groups. In the last two decades, the mean paddy yield of the *Kinh* increased from 3.6 to 4.6 t/ha while in ethnic minority villages yields only improved by 0.3 t/ha to 3.5. For all groups the reported variation in yields are very high. In ethnic villages yield variation for the year 2000 spans over a

Figure 4.15: Paddy rice yield by ethnic group and in total

Source: Village survey in Lak and Krong Bong districts, 2000/2001

range of almost six t/ha. Data from official statistical sources report for the research area similar yield levels for 2000, but average paddy yields of below 2 t/ha for the years between 1976 and 1980 and 3 t/ha in 1985 (Dak Lak Statistical Department, 1981, 1991).²²

Upland rice is almost entirely grown by ethnic villages. In 2000 upland rice was only cultivated in 20 of all 101 surveyed villages while in 1980 approximately 80% of the villages were involved in the cultivation of upland rice. Only two *Kinh* villages reported to grow upland rice in 2000. Formerly, upland rice was a main staple food crop and cultivated on the shifting cultivation fields by the majority of the households. A large part of the abandoned upland rice fields are now used for maize cultivation for sale on local markets. Other fields regenerated into mixed grassland or open canopy forest. Declining productivity of upland rice is rarely compensated by an increase in fertilizer use as the profitability of upland rice was decreasing over time relative to other crops. Today, upland rice is an example for an extensive cultivation practice with low applications of purchased inputs. Mean reported upland rice yields dropped from an average of 1.4 to 0.85 t/ha, which,

²²These figures have to be treated with some care as discussed in the subsection 3.6.2 on data limitations.

in the absence of fertilizer application, might be an indication for decreasing soil fertility due to reduced fallow periods (figure 4.16).

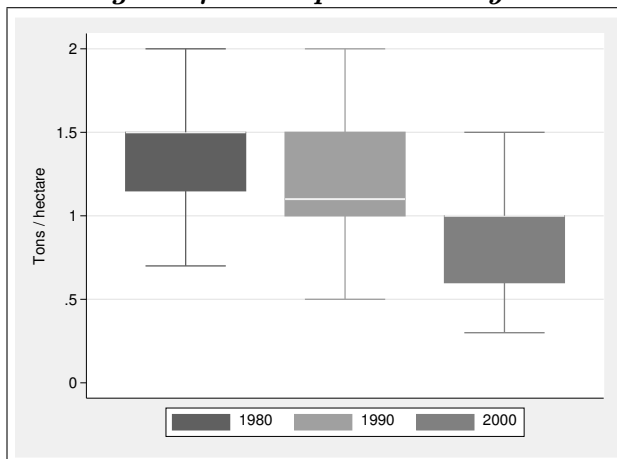
Crop composition shifted to more market-oriented crops in the last decade as shown in figure 4.17. In the last five years households rapidly adopted improved rice varieties and hybrid maize as well as coffee. In 2000, 20% of the households cultivated improved rice varieties and 18% hybrid maize. Coffee is cultivated by almost 14%, mainly to a small extent in home gardens as soil conditions in the research area do not allow for profitable coffee production on a bigger scale. The percentages in figure 4.17 imply a significant

absolute increase as the population in the survey villages soared in that period (see figure 4.7).

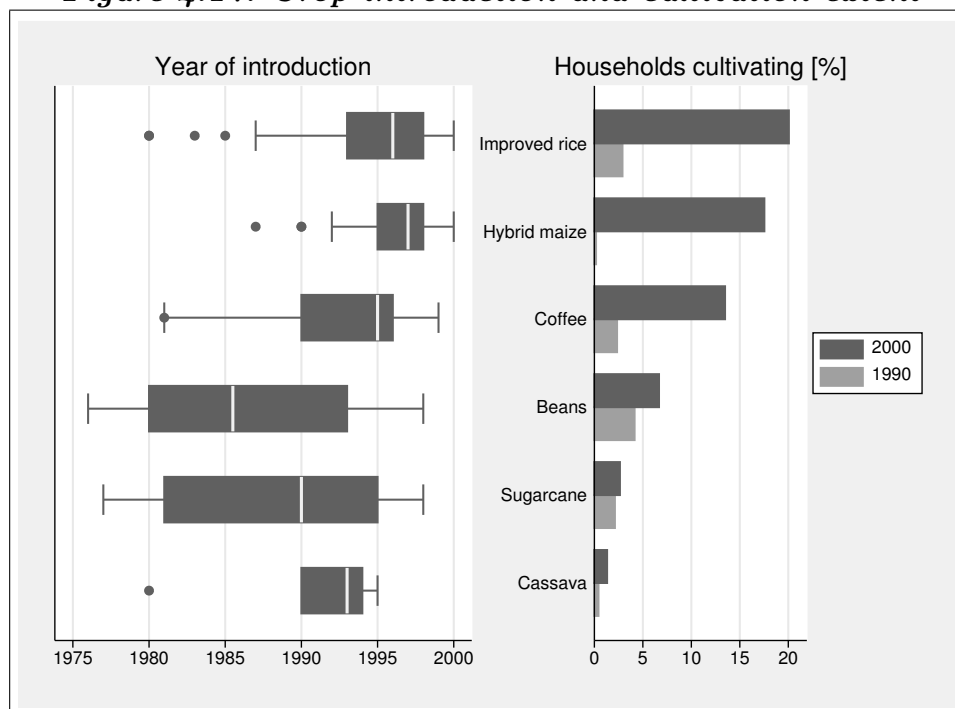
4.3.3 Economic growth and welfare

In the past decade Vietnam was one of the fastest growing economies of the world with a gross domestic product (GDP) rising at rates averaging 7.7% in the period 1991 to 2001. Per capita growth rates of GDP in that period were 6.1% (World Bank, 2002a). In the *Vietnam Development Report 2003*, the World Bank expects growth rates over the next few years likely to be in a similar range at around 7% (World Bank, 2002b). According to the *Vietnam Living Standards Measurement Survey* (VLSS) of 1992/1993 and 1997/1998, carried out by the General Statistics Office (GSO) and the most detailed socioeconomic information available country-wide, poverty incidence declined from 58% in 1993 to 37% in 1998 (The Socialist Republic of Vietnam, 2002). In 2001 the poor represented 32% of the population (World Bank, 2002a). By international poverty standards Vietnam's poverty incidence reduced to just half of the 1990 rate. Vietnam is considered by the international community to rank among the best performing countries in terms of poverty reduction (The Socialist Republic of Vietnam, 2002). The three poorest regions

Figure 4.16: Upland rice yield



Source: Village survey in Lak and Krong Bong districts, 2000/2001

Figure 4.17: Crop introduction and cultivation extent

Source: Village survey in Lak and Krong Bong districts, 2000/2001

in Vietnam are the Northern Uplands, the North Central Coast and the Central Highlands (World Bank, 2002b).

In 2001, Dak Lak had a GDP per capita of USD 421 (Dak Lak Statistical Department, 2002).²³ The two research districts are among the poorest in Dak Lak Province. In the village survey, the villagers were asked about the number of several consumables in the village. To compute *wealth indicators* selected consumables were aggregated by combining the percentage of households owning television sets, motorbikes and improved roofs of their houses. Based on discussions with the villagers it was decided that these variables are best suited to distinguish wealthier from poorer households. Other indicators as the number of stone floors were not included in the wealth index as they differ due to cultural peculiarities and not necessarily because of differences in wealth. For example, most ethnic minority groups live in traditional houses built on wooden stilts, obviously an unsuitable foundation to construct stone floors.

To summarize the wealth indicators *violin plots* were used. The violin plots in figures 4.18, 4.19, and 4.20 combine box plots with vertical density traces into one

²³The author could not obtain reliable figures on earlier GDPs. Johnson & Brieng (2001) reports an average growth of GDP of 12% in the 1990s without providing a source.

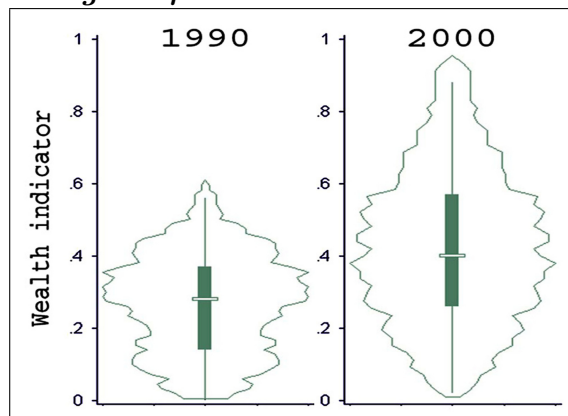
diagram to add information in data exploration (Hintze & Nelson, 1998). The violin plots highlight the peaks and valleys of a variable's distribution and, derived from a box plot, show the median, spread, asymmetry as well as outliers of the data. The density traces augment information by showing the distributional characteristics. One density trace extends to the left while the other extends to the right. Two density traces are plotted to add symmetry, which makes it easier to compare the magnitude of the distribution. The median is represented by the horizontal bar, so that quick comparisons can be made. Values within the second and third quartile are illustrated by the vertical box.

Village data reveal that this *wealth index* improved significantly in the last decade with a higher median and a significantly higher second and third quartile (figure 4.18). The spread, however, extends over the whole range in the year 2000, which points at increasing inequality. Separating the wealth indicators by ethnic composition (figure 4.19) uncovers that differences between ethnic Vietnamese and ethnic minorities widened in the last decade and, according to the data at hand, households in *Kinh* villages own substantially more of the considered consumables than households belonging to ethnic minorities in 2000.

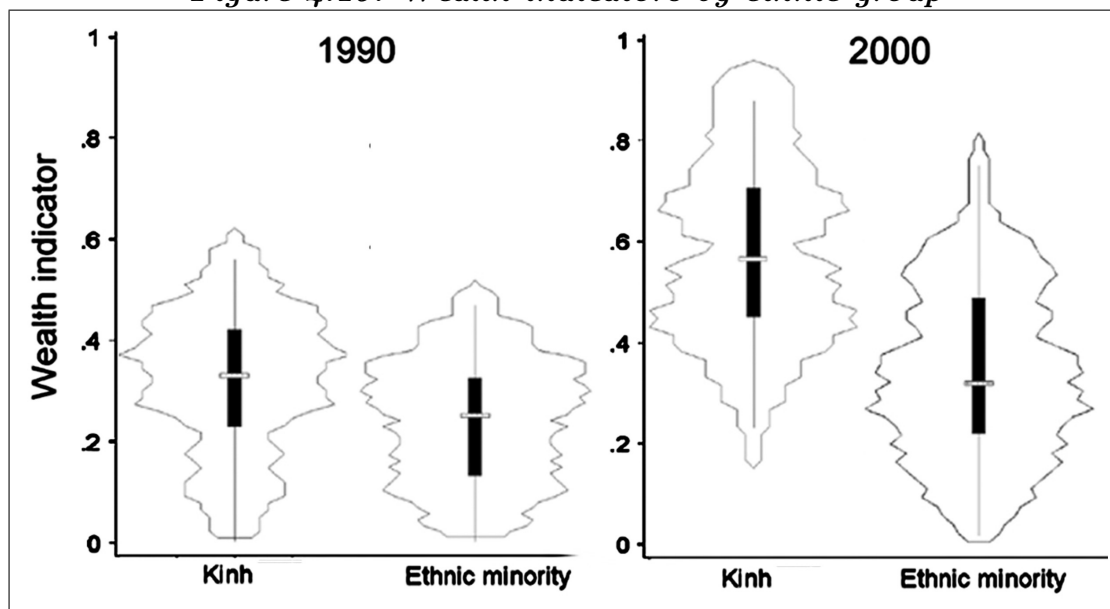
These findings are in line with results from the two VLSS of 1993 and 1998, both for poverty reduction and for increasing differences in wealth between *Kinh* and ethnic people (The Socialist Republic of Vietnam, 2002).

In figure 4.20 the wealth indicator is plotted by three distance intervals from the villages to the nearest all-year road in 1990 and 2000. The intervals for distance are created by specifying three grouping intervals with equal frequency. The interval groups are labeled *close*, *medium* and *far* to describe the increasing distance from the nearest all-year road as a categorical indicator. Good road access seems to have a positive influence on wealth at both points in time. The effect herein is much smaller in 1990, presumably as a result of less market orientation with a smaller share of the agricultural production sold on markets. Particularly, villages located

Figure 4.18: Wealth indicators



Source: Village survey in Lak and Krong Bong districts, 2000/2001

Figure 4.19: Wealth indicators by ethnic group

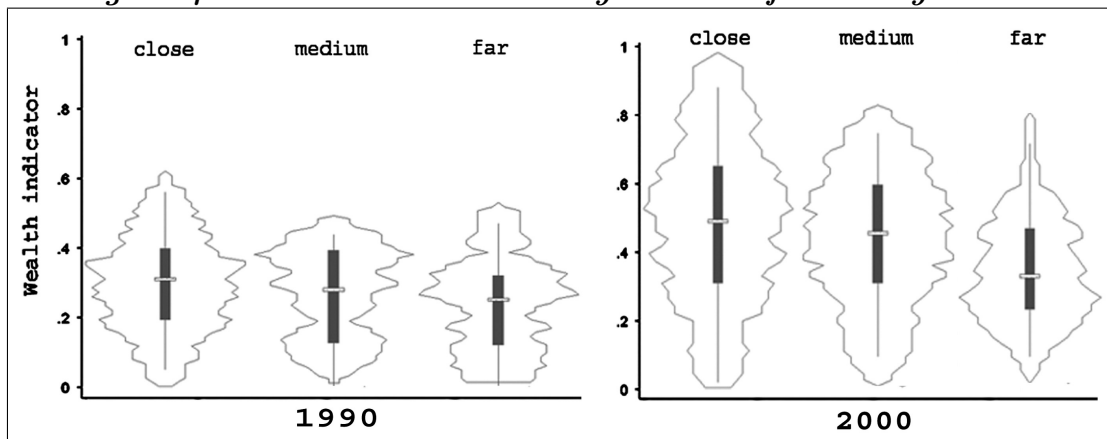
Source: Village survey in Lak and Krong Bong districts, 2000/2001

far away from all-year roads in 2000 have lower wealth levels with a median equal to villages, which were located close to all-year roads a decade earlier.

Food security constantly improved in all villages (figure 4.21). The average number of households reportedly suffering from food shortages for more than one month per year decreased from 68% to 50% in the last two decades while the number of households with food shortages longer than three months dropped from 49% to 28%. According to our village interviews half of the households are still affected by a lack of food.²⁴ Striking is the difference between ethnic and *Kinh* villages. While in 2000 57% of the ethnic villages reported food shortages, only 35% of the *Kinh* did so. The figure also reflects the improvement in household's livelihoods over the last two decades. The share of *Kinh* households with a lack of food for at least one month dropped from 77% to 35%. In ethnic villages the same indicator only decreased from 64% to 58%. Overall improvements of reported food security are largely on the account of *Kinh* villages. Apparently, ethnic villages participated only modestly in improvements in food security and rural livelihoods.

²⁴The author suspects that to some questions – like the one on food shortages – the villagers gave strategic answers to the questions in (justified) expectation of future government support. Therefore, some figures might be overly pessimistic.

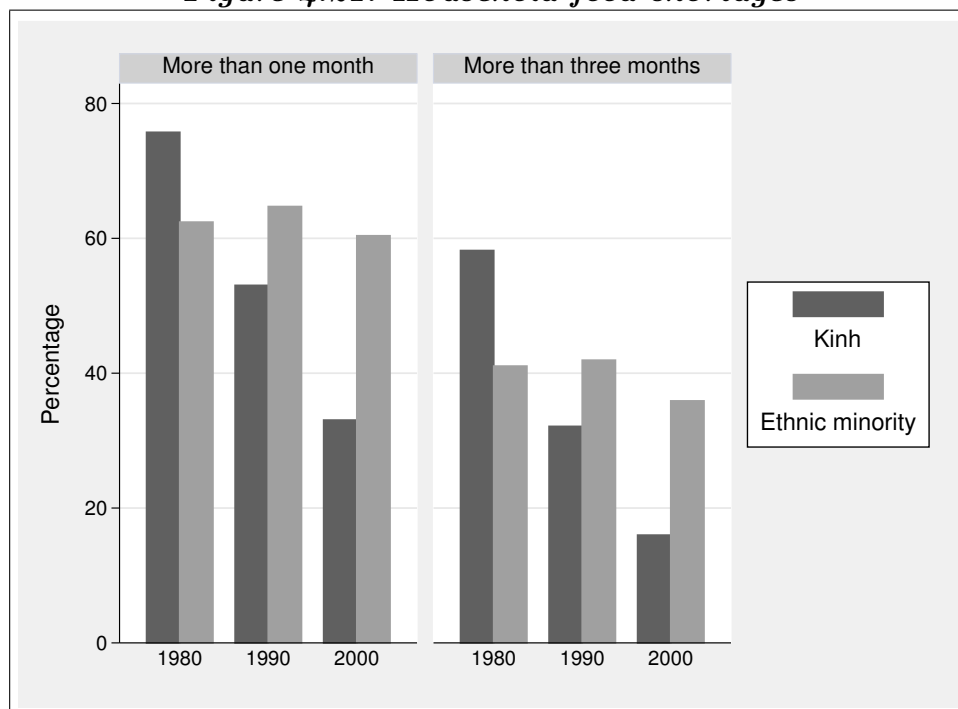
Figure 4.20: Wealth indicators by distance from all-year road



Source: Village survey in Lak and Krong Bong districts, 2000/2001

One reason for improved food security might be better *access to credit*. In 1990, only 3% of the households in the interviewed villages used formal sources of credit without any differences across ethnic groups. At the time of the survey 48% of the *Kinh* and 28% of ethnic minority groups reported to have borrowed money from formal sources, which include government programs like the HEPA program as well as the *Bank for the Poor* and the *Vietnamese Bank for Agriculture and Rural Development*. Better access to credit enabled farmers to invest in inputs like improved seeds and fertilizer to boost output. Many farmers also stated that they use a significant share of the credits for the education of the children.

Overall *school enrollment* is traditionally high in Vietnam for a country with a GDP of USD 420 per capita (World Bank, 2002a). The *World Development Indicators Database* state for Vietnam a rate of 99% for primary and 52% for secondary school enrollment in 1997. In the research area, education significantly improved in all villages and across all ethnic groups. In figure 4.22 school enrollment is graphed according to ethnicity and separated for primary and secondary school enrollment. On average 88% of all children visit primary school and 53% secondary school in 2000. Of the *Kinh* 96% of the children in the respective schooling age are enrolled in primary schools and 70% in secondary schools. Secondary school enrollment of ethnic Vietnamese increased from 18% to 70% between 1980 and 2000. School attendance is considerably lower in ethnic villages. At present, this difference is especially large for secondary school enrollment, which only 53% of children in ethnic villages visited in 2000 and 25% in 1980. Despite the differences across ethnic groups

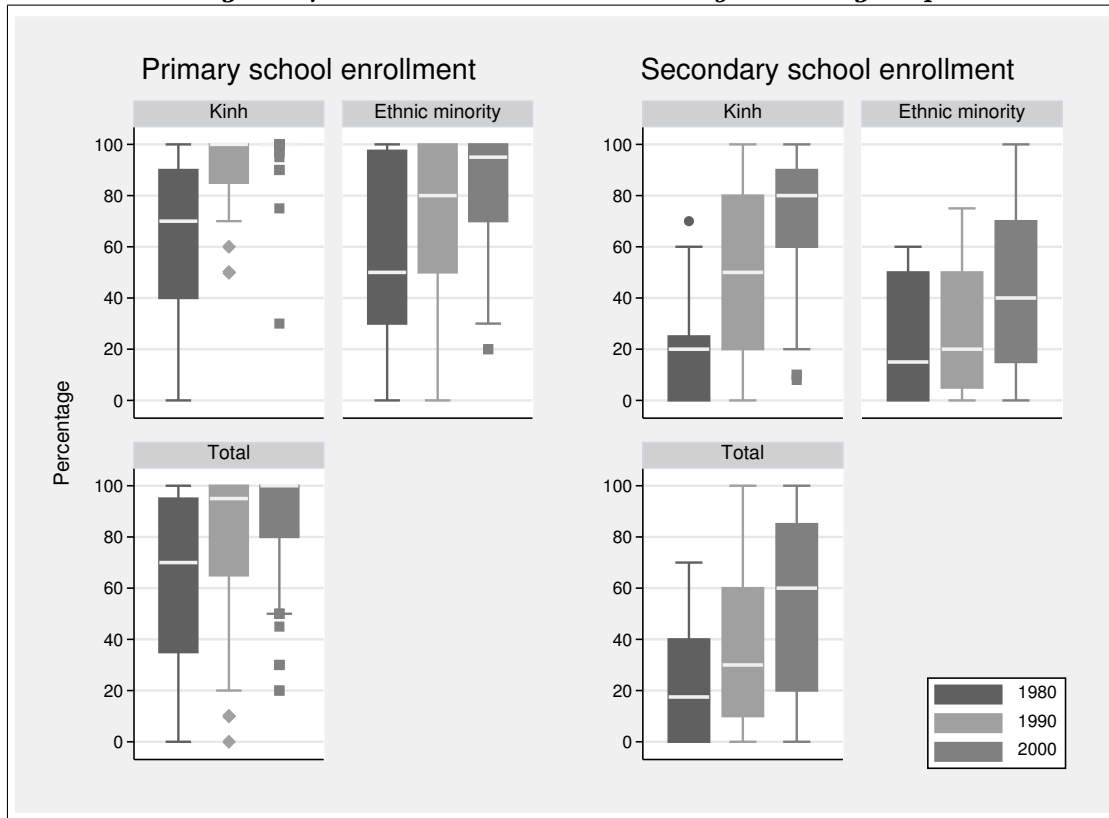
Figure 4.21: Household food shortages

Source: Village survey in Lak and Krong Bong districts, 2000/2001

school enrollment in the research area is high compared to other areas with similar economic indicators. This is the case for most of Vietnam and might partially explain the high growth rates especially in agricultural production and Vietnam's good performance in poverty reduction after the *economic renovation*.

Relationships for selected indicators of the conceptual framework (see figure 2.1) are further examined with standard pairwise correlation coefficients, depicted in table 4.7. Correlation coefficients examine the linear relationship between variables, without providing insights into underlying causalities. Off-diagonal elements in table 4.7 indicate the correlation coefficients and diagonal elements are unity, i.e. perfectly correlated.

As a visual alternative to standard correlation matrices, figure A.3 in the appendix sketches pairwise scatterplot graphs (or scatterplot matrices) to visualize the direction and strength of association among the same set of selected survey

Figure 4.22: School enrollment by ethnic group

Source: Village survey in Lak and Krong Bong districts, 2000/2001

variables for the year 2000 including their dispersion (StataCorp, 2003).²⁵ The scatterplot matrix reveals the expected strong negative relation between food shortages and the wealth indicator with a correlation coefficient of -0.810 (table 4.7), significant at the 1% level. Rice yields show equally strong and significant correlations with wealth and food shortages in the expected direction (0.628 and -0.523). Most of the other variables do not unveil very strong, but partly still significant ties. A positive relation can be observed between the percentage of households using credit and paddy yields as well as the wealth indicator. Wealthier households tend to send more of their children to a secondary school (correlation coefficient = 0.498) and suffer less from food shortages (-0.423). A stronger negative correlation coef-

²⁵In figure A.3 of the appendix the lower left and upper right triangle of the scatterplot matrix are mirrored on the diagonal. This provides more room to label the axes making the graph more readable, but does not convey additional information.

Table 4.7: Correlation coefficients of selected variables

	Wealth indicator	Food shortages	Paddy yield	Two harvests ^b	Paddy area	Credit use	School enrollment
Wealth indicator	1						
Food shortages	-0.810* ^a	1					
Paddy yield	0.628*	-0.523*	1				
Two harvests ^b	-0.091	-0.028	-0.100	1			
Paddy area	0.089	-0.077	0.302*	-0.042	1		
Credit use	0.279*	-0.129	0.286*	-0.159	0.074	1	
School enrollment	0.498*	-0.423*	0.347*	-0.006	-0.068	0.343*	1

Source: Village survey in Lak and Krong Bong districts, 2000/2001

^aNote: Coefficients significant at the 5% level or better are marked by an asterisk.

^bTwo harvests of paddy rice per year.

ficient would, for example, be expected between secondary school enrollment and food shortages, which the data at hand does not prove. Household credit use is positively correlated with secondary school enrollment, which gives some evidence for the link between higher educational levels and better access to formal capital sources. Somewhat surprising, given the labor constraints of most households, is the positive correlation of the cultivated paddy area with paddy yield. This might point to existing economies of scale. However, the effect is relatively small.

In addition to the linear association between variables, figure A.3 also reveals their dispersion. In the research area, only 10% of the households cultivate more than half a hectare of paddy rice. Average paddy yield was 3.9 tons per hectare in 2000 and most of the households lie in the range of 3.0 to 5.5 tons. Especially the yields of the *Kinh* households are uniformly distributed around a mean of 4.6 tons per hectare.

4.3.4 Environmental indicators

Figure 4.23 plots villager's perception about the change in the state of natural resources from 1980 to 1990 and figure 4.24 the change from 1990 to 2000. The survey questionnaire focused on variations in soil fertility on upland fields, ground water levels, timber availability and the availability of non-timber forest products (NTFP).²⁴ While between 1980 and 1990 (figure 4.23) the indicators are almost normally distributed, most variables in the last decade (figure 4.24) are visibly skewed

to the right indicating that villagers perceive a degenerating state and decreasing quantity of the natural environment over time.

Figure 4.23: Change in state of natural resources, 1980 to 1990

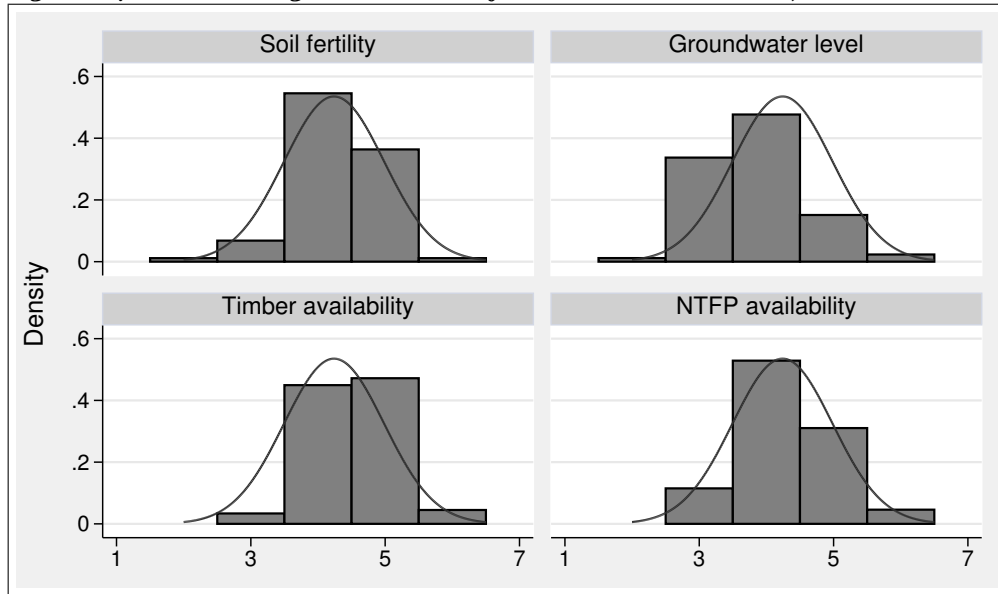
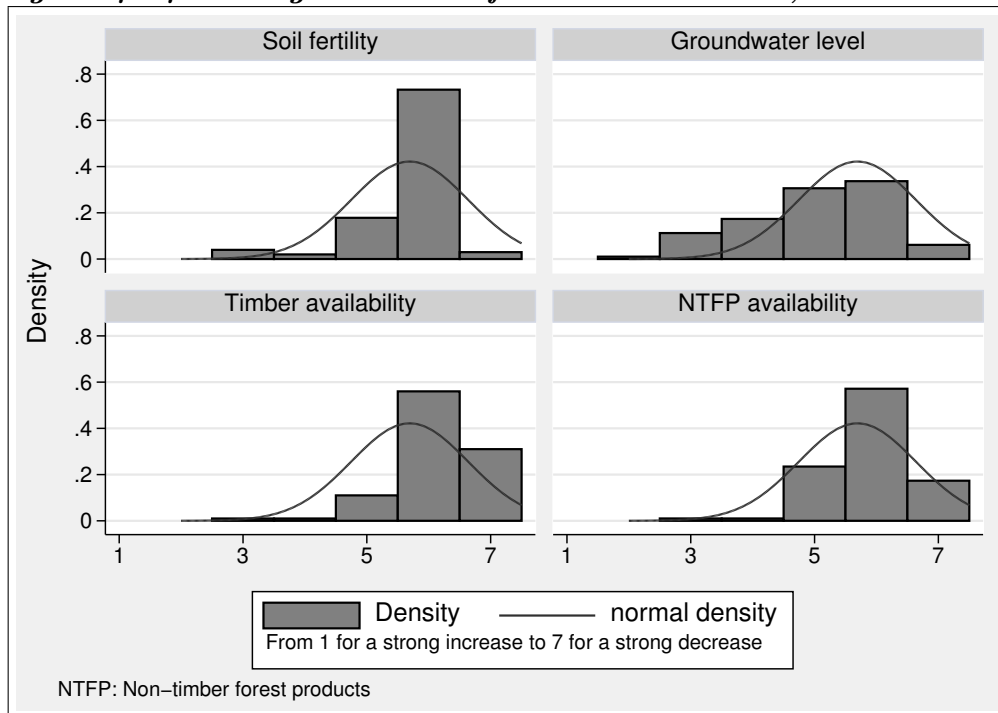


Figure 4.24: Change in state of natural resources, 1990 to 2000



Source: Village survey in Lak and Krong Bong districts, 2000/2001

For the first period, villagers already reported a decrease in timber availability over proportion, which exacerbated in the second period. Along these lines, the time necessary to reach the nearest primary forest by foot increased from less than two hours in 1980 to three and a half in 1990 and five hours in 2000. The satellite image interpretation confirms a decrease in forest cover for the first period, especially close to populated areas. Overlays of land cover maps, however, also revealed that overall forest cover increased in the last decade mainly on more remote fields (table 4.5). Yet, forests might have declined in quality over time, which is difficult to detect with by means of satellite image interpretation.

4.4 Summary

Forest cover in Dak Lak province decreased from 75% to 51% over the last 25 years. At the same time, changes in land cover in the two districts of the research area reveal only a modest decrease in forest cover of 2%. Agricultural expansion after liberation day triggered a decline in forest cover of 10% between 1975 and 1992. Forest regenerated in the last decade on areas formerly used for shifting cultivation and on mixed grassland. Landscape fragmentation changed to a more homogenic, less complex and fragmented landscape. At present, cultivated plots are closer to populated centers than a decade ago.

The distribution of socioeconomic and policy variables demonstrate the diverse characteristics of villages. Input use and productivity reached higher levels across most villages, wealth levels and food security raised, and access to markets and all-year roads improved. Yet, significant differences in above indicators are present across ethnic groups and with increasing distance to all-year roads. Most of these differences evolved since liberation day and enlarged over time, in particular in the last decade. Educational levels, measured by school enrollment rates, are high for both primary and secondary schools in 2000. School enrollment exhibits similar differences between ethnic groups with increasing differentiation over time.

Villagers reported decreasing availability and quality of the natural environment. Forest cover has increased over the last decade as testified by the satellite images, but forest quality might have declined. Average population growth rates of 5% over the last two decades might have exacerbated the demand for forest products, thereby decreasing timber and biomass volume particularly closer to village locations. Still, the decline in land-intensive agricultural activities lowered the pressure on forest

cover in the research area, especially on marginal land. The underlying reasons will be further explored in the next chapter.

Econometric Analysis of Land-cover Change

Following the conceptual framework of induced innovation (see section 2.1) and taking into account recent advances in the literature using spatial analysis in combining geophysical and socioeconomic data (see section 2.4 and chapter 3), land use is explained as a function of geophysical variables (level and variance of rainfall, soil suitability, altitude and slope of land), population growth, socioeconomic village characteristics, operational variables for measuring the spatial placement of policy-induced investments in road and market infrastructures and the introduction of new agricultural technology and crops. Endogenous variables are included as lagged values where appropriate to obtain consistent econometric estimates.

The empirical model in this study estimates land use with a multinomial logit model (MNL) as a static process. The relationships between these variables will be examined in a spatially explicit framework and their respective direction and magnitude quantified. The estimation of two models for the two time periods allows the interpretation of changes in influencing factors between the period from 1980 to 1990 and from 1990 to 2000. A comparison of coefficients can yield important insights into pathways of land-use change over the two periods as well as into the effect of important policy variables on farmer's reactions concerning their use of land.

5.1 Multinomial logit

The MNL model has five land-cover classes as categorical, unordered dependent variables (for the descriptive statistics of the land-cover classes employed in the estimation see table A.1 in the appendix). One multinomial logit model was performed for each of the two time periods. Hereafter, the first period, for which land-cover data is available for the years 1975 and 1992, will be referred to as the *model 1990*. The second period with available land-cover data of 1992 and 2000 will subsequently be labeled *model 2000* (see subsection 3.3.2). Survey recall data for Liberation Day and 1980 was used as *best estimate* for predetermined variables in 1975 and, similarly, recall data from 1990 was employed for 1992 (the descriptive statistics for the RHS variables are listed in table A.2 in the appendix). It would have been more suitable to arrive at equally long time periods, which was not possible.¹ To control for potential endogeneity problems, only lagged values for time-variant independent variables such as population growth and placement of schools are considered in the empirical applications. In addition, all variables were tested for multicollinearity. The empirical results are reported as raw coefficients in table A.3 of the appendix for the *model 2000* and in table A.4 for the model 1990. Both models are estimated with closed forest as the comparison or reference group. A MNL model yields coefficients for all possible comparisons of the outcome categories and each independent variable. Therefore, a MNL generates $k * m$ coefficients where k is the number of independent variables and m the number of outcome categories. For the estimations in this study, the MNL model results in 125 coefficients, each of them of potential interest for the interpretation. The output tables A.3 and A.4 in the appendix report only coefficients with closed forest as comparison group. In the discussion hereafter some signs of other interesting coefficients not included in the tables A.3 and A.4 of the appendix will also be mentioned.

The MNL models do not include lagged values of land cover due to relatively little variation in the overall land cover over time. About 85% of the pixels from 1992 to 2000 and more than 70% in the period 1975 to 1992 did not change over time, mainly on the remote forested areas (see also table 4.5 and figure 4.13). Therefore, the lagged values of some land-cover categories predict land use almost perfectly, making model estimation problematic. Equally, the lagged value of convertible land predicted the category *paddy* perfectly as there was no land suitable for an expansion

¹See footnote 10 in chapter 3.

of the area under paddy cultivation given the calculation of this indicator (see section 3.2.3.1). Hence, these variables were excluded from the estimations. Since the first nature reserve in the research area was only established in 1991, the protection variable and its interaction terms are only included in the *model 2000*.

5.2 Empirical results

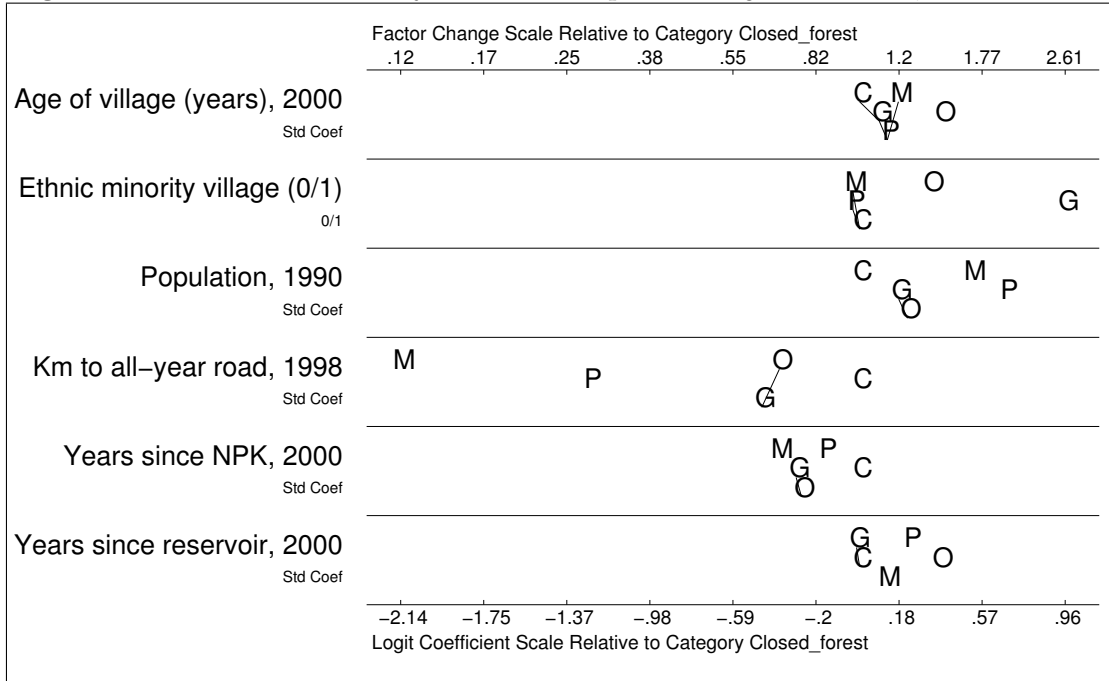
Coefficients for most topographic variables are significant at the 1% level and show the expected signs in both models. Agricultural land uses are more likely on lower altitudes, on flatter land and on more suitable soils. Higher rainfall and less variance increases the likelihood of pixels to be covered by mixed agriculture and paddy in both models. The coefficients for soil suitability for paddy has the expected influence and better suitability increases the probability of paddy against all and the probability of agriculture against the remaining soil classes. Surprisingly, better soils increase the odds of open canopy forest compared to all other land-use categories, however it also makes agricultural land uses more probable compared to grass and paddy. Soil suitability for agriculture has an insignificant effect on the probability of closed forest cover compared to agricultural land uses. Closed forest is less likely on soils suitable for mixed agriculture in the *model 1990* (table A.4). However, better soils also increase the probability of mixed agriculture compared to all other land uses.²

Odds ratios, or the ratios of the odds of two outcomes, are another method to compare model outcomes. Odds ratios can facilitate model interpretation as they are constant for changes in other categories of the dependent variable. In figure 5.1 and 5.2 the odds ratios of selected socioeconomic and policy variables are plotted for the *model 1990* and *model 2000*.

Initial letters in figure 5.1 and 5.2 indicate variable values of land-use classes, *M* denotes *mixed agriculture*, *P* *paddy*, *C* *closed forest*, *O* *open forest* and *G* stands for *grass*. The likelihood increases in magnitude from left to right. Closed forest (C) as the comparison group are stacked on top of each other. Closed forest has a factor change scale of one and variables are either evaluated for a change of one standard deviation (*Std Coeff*) or for a discrete change from zero to one (*0/1*). Values connected by lines are **not** significant at the 10% level. The multiplicative scale on the top axis in figure 5.1 measures the $\exp(\beta_{k,m|n})$ and the additive scale on

²These coefficients are not obvious from table A.4 as closed forest is the chosen comparison group.

Figure 5.1: Odds ratios of selected explanatory variables, model 2000



Source: Own calculations

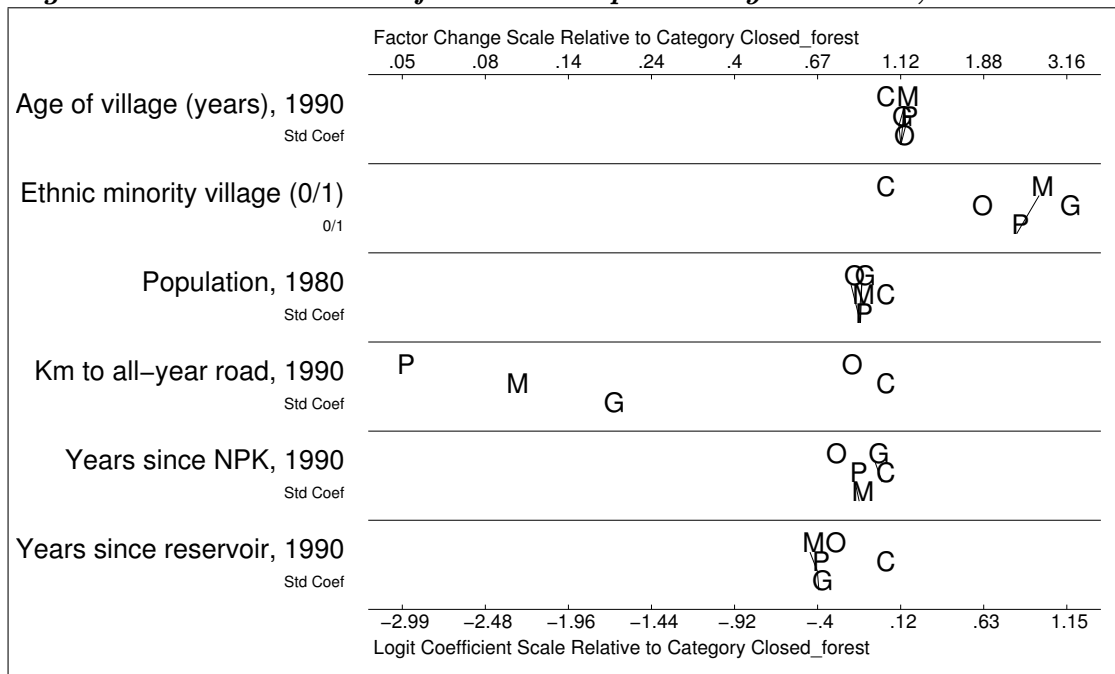
the bottom axis $\beta_{k,m|n}$ (see section 3.1.1).³ One example to interpret figure 5.1 for population in row three: A unit increase in population makes it more likely to observe paddy (P) compared to all other categories. The least likely category is closed forest (C), which corresponds to the positive coefficients of all categories relative to closed forest in table A.3 of the appendix. The distance between the letters indicate the magnitude of the effect. Paddy and mixed agriculture (M) are significantly more likely at a unit increase of population than all other land categories. Open forest (O) is **not** significantly more likely than grass (G) as the two are connected by a line. Yet, both are significantly more likely than closed forest as they are to the right of closed forest in figure 5.1.

Older villages increase the likelihood of open forest (O) in the model 2000 compared to all other categories and mixed agriculture (A) compared to closed forest (C). Effects in the model 1990 are small, but mixed agriculture and paddy (P) are more likely than closed forest (row one in 5.1 and 5.2). Ethnic villages (row two)

³The corresponding Stata[®] code is programmed by J. Scott Long and Jeremy Freese and publicly available. For thorough explanations see Long & Freese (2001).

have the lowest odds for closed forest in both models. In the *model 2000*, open forest and grass (G) are significantly more likely within catchments of ethnic villages than other land uses, which points to the abandonment of agricultural areas and to forest regeneration on areas formerly used for shifting cultivation and to the decrease in these cultivation techniques over time. Grass, mixed agriculture and paddy have higher odds than the two forest categories in the *model 1990*, an indication for the land intensive agricultural systems.

Figure 5.2: Odds ratios of selected explanatory variables, model 1990



Source: Own calculations

Higher *lagged population* increases the probability of other land-use classes relative to closed forest in the *model 2000* as expected. A comparison between the two periods reveals some interesting changes. The effect of population is significant at the 1% level in the *model 2000*, but had smaller and partly insignificant effects in the *model 1990*. This indicates that – at lower levels of population density during the first period – population induced little pressure on change in land use, but growing population density generated a demand for paddy and agricultural land. Land was converted into agricultural uses at locations with high population density, but regenerated away from populated areas. The status of a *hero village*, an indicator for political capital, decreases the likelihood for agricultural land uses in

the *models 1990* (tables A.3 and A.4 in the appendix). Hero villages have higher odds for forest cover within their catchments. In both models paddy is least likely in hero villages. The spatially explicit model has its strengths in the assessment of locational changes, and does not consider explicitly intensification effects on existing areas. Hero villages might have enjoyed greater levels of government assistance, thereby being able to increase agricultural production on existing land.

The coefficients for *distance from all-year roads* have the expected signs and higher values strongly increase the odds for closed and open canopy forest. Not surprisingly, agriculture and paddy are more probable to be located closer to roads. The years since the availability of *primary schools* increases all land-use classes relative to closed forest in the *model 2000*. In the first period, an earlier establishment of primary schools makes paddy more likely than other land uses. *Irrigation* has a two-fold effect. An earlier construction of reservoirs decrease while more irrigated area increase the likelihood of pixels to be covered by primary forest in the first period. Thus, more irrigated area might have a positive effect on forest protection through an increase in the land use intensity. The age of *protected areas* increases the likelihood of open and closed forest cover compared to all other categories and, therefore, contributed to forest regeneration. On steeper land, forest protection increases the probability of a pixel covered with closed forest relative to grass land while the interaction terms with soil quality do not affect land-use classes. Thus, the hypothesis that protected areas are established on more marginal lands is not verified by the data.

The index values for *latitude and longitude* to correct for spatial autocorrelation yield significant coefficients in X and Y direction for most classes except paddy and open forest in *model 1990*. The direction of the coefficients indicates the southeastern part of the research area (higher X- and Y-values) as an area where agricultural land use is more likely in the *model 1990* and, therefore, agricultural development has a higher likelihood to be concentrated.

5.3 Prediction accuracy

Test statistics are reported in table A.5 of the appendix. Overall, the *model 2000* performs better than the *model 1990*, possibly due to fewer missing values in the survey data and more exact data on land cover and recall information. Likelihood-ratio tests for the independent variables are reported in table A.7 of the appendix.

The likelihood ratio tests reveal that most coefficients are significant at the 1% level in both models, except the interaction coefficients of protected areas with soil suitability. Likelihood-ratio tests for combining dependent categories allow to legitimate the rejection of the hypothesis that categories are indistinguishable (table A.6 in the appendix).

Table 5.1: Correctly predicted cells at different cut-off points

Model	Cut-off point	Mean	Std.Dev.	Min.	Max.
2000	Highest P	0.783	0.412	0	1
	50%	0.767	0.412	0	1
	60%	0.701	0.458	0	1
	70%	0.606	0.489	0	1
1990	Highest P	0.696	0.460	0	1
	50%	0.630	0.483	0	1
	60%	0.520	0.500	0	1
	70%	0.394	0.489	0	1

Source: Own calculations

Once predictions have been made, overall predictive power can be measured by the locations predicted correctly. In table 5.1 the prediction accuracy is assessed for both models. Correct predictions take the value one and incorrect predictions the value zero. Table 5.1 shows that, taking the highest predicted probability as predicted land use, the *model 2000* performs better with an accuracy of 78% compared to 70% in the *model 1990*. Prediction accuracy naturally decreases with increasing prediction thresholds. In the *model 1990* only 39% of the correctly predicted cells have a predicted probability above 70% while 61% of the predictions are accurate at prediction levels above a 70% threshold in the *model 2000*.

Predictive power can also be assessed by calculating a *prediction matrix*, where observed land-cover categories are tabulated against predicted categories (Nelson & Geoghegan, 2002). In table 5.2 one prediction matrix is presented for each model. The rows in the matrix of table 5.2 show the number of locations in one of the five observed land-use categories and its columns the number of locations predicted to be in one of the five land-use categories. In the diagonals of the matrix one finds

Table 5.2: Prediction matrices

Observed land use, 2000	Predicted land use, 2000					Total
	Mixed agriculture	Paddy	Closed forest	Open forest	Grass	
Mixed agriculture	4,375	346	16	436	89	5,262
Paddy	650	1,063	0	26	13	1,752
Closed forest	66	5	11,845	2,217	15	14,148
Open forest	533	18	1,904	12,085	115	14,655
Grass	494	18	98	781	197	1,588
Total	6,118	1,450	13,863	15,545	429	37,405

Observed land use, 1992	Predicted land use, 1990					Total
	Mixed agriculture	Paddy	Closed forest	Open forest	Grass	
Mixed agriculture	3,465	506	65	1,099	307	5,442
Paddy	511	1,307	1	38	12	1,869
Closed forest	20	0	11,022	1,979	102	13,123
Open forest	602	12	2,019	9,894	390	12,917
Grass	1,371	58	274	2,021	601	4,325
Total	5,969	1,883	13,381	15,031	1,412	37,676

Source: Own calculations

the correctly predicted elements. For example, 533 cells are predicted to be mixed agriculture in 2000, however, in the land-cover maps they are observed to be covered by open forest (table 5.2).

The majority of the false predictions in both models are pixels covered with grass, which were correctly predicted only in 12% of the cases in 2000 and 14% in 1990 (table 5.3). Grassland is also the smallest category, which adds to the problems observed in its prediction accuracy. In 2000, grass is predicted to be mixed agriculture in 31% and open forest in 49% of the cases (table 5.2, e.g. $781/1,588=0.49$). Prediction mistakes are similar in the *model 1990*. Grassland is apparently most difficult to distinguish from these two categories, as the spectral characteristics of

Table 5.3: Percentage of correctly predicted cells by category

Cover class	2000	1990
Mixed agriculture	83.1	63.7
Paddy	60.7	69.9
Closed forest	83.7	84.0
Open forest	82.5	76.6
Grass	12.4	13.9
Total	79.0	69.8

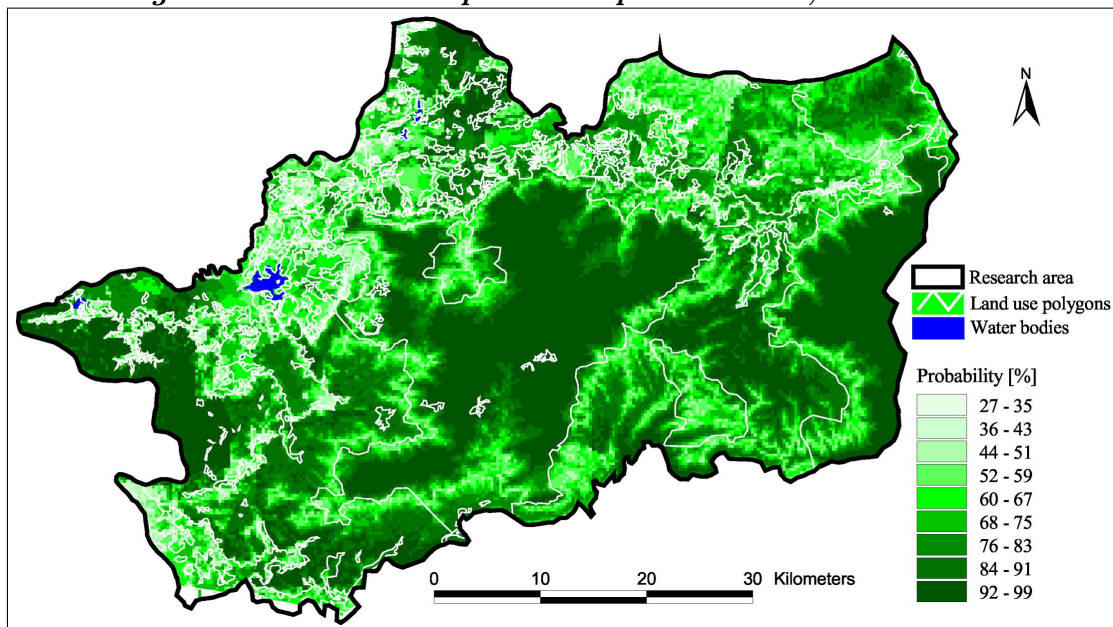
Source: Own calculations

land cover are quasi-continuous and often difficult to separate. Highest probabilities are found for the three biggest classes of mixed agriculture, closed and open canopy forest with over 80% of correct predictions in the *model 2000*.

There are substantial differences in accuracy across categories. Low prediction accuracy in a category might hint at a low degree of separability between this and another category, at differences in the quality of the actual land cover map or at differences in the predictive power of the right-hand side variables for the different categories. The prediction matrix gives little information on the spatial accuracy of the prediction (see section 3.1.1.1). Another approach is to use the probability values directly. Two related graphical measures of predictive power are presented in Nelson et al. (2001a). The first maps the maximum probability value Pr_{max} at every location. These maximal probabilities, which are derived from the econometric estimation, are mapped in figure 5.3 together with the land use polygons of the year 2000.⁴ The resulting map gives a spatial representation of the power of the prediction, but does not convey any information about prediction accuracy Nelson et al. (2001a). The relatively low prediction power along the borders of the discrete land-cover categories can be observed from figure 5.3. High predictive power can be discovered for the category closed forest in the middle of figure 5.3, in the eastern and in the western part (compare also the land-cover maps in figure A.1). These areas have a low natural potential for agricultural production, i.e. high elevations, steep slopes and unsuitable soils. Prediction accuracy for agricultural land uses is

⁴This figure is only included for the *model 2000* as results for the *model 1990* are very similar.

Figure 5.3: Maximal predicted probabilities, model 2000

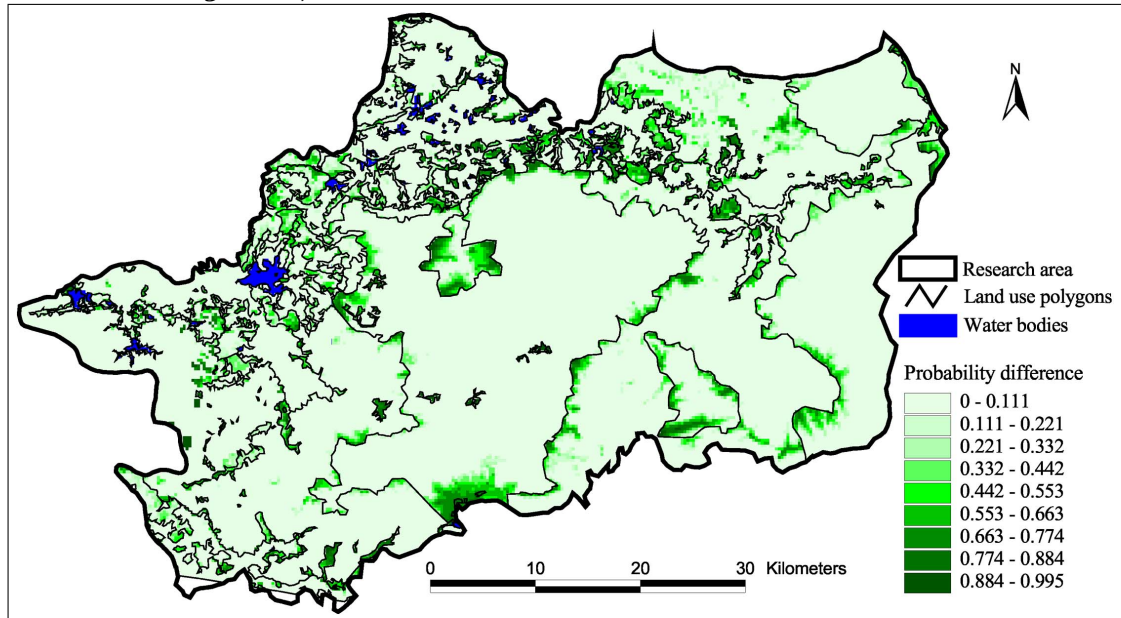


Source: Own calculations

equally high in the north of the research area, a region with suitable agricultural conditions as well as good market access.

Another measure to visually assess predictive power is to plot all locations where the predicted and observed land uses differ Nelson et al. (2001a). In that way, the difference between the maximum probability and the probability for the observed land use is measured, or $Pr_{diff} = Pr_{max} - Pr_{actual}$. If the category with the highest probability value is also the observed category, then $Pr_{diff} = 0$, otherwise $0 < Pr_{diff} < 1$. Figure 5.4 graphs the probability differences for the *model 2000* and shows that biggest differences are located along the boundaries of the land use polygons.⁴ Less accurate predictions can be observed predominantly on the borders between open and closed canopy forest, situated in the middle and south-eastern part of figure 5.4. The reason is most probably the above mentioned problem related to the difficult separability between these two classes. As the observed land-cover categories are a discrete representation of the quasi-continuous color values of the raw satellite image, this lack of prediction accuracy between closely related categories is expected.

Figure 5.4: Probability differences between observed and predicted categories, model 2000



Source: Own calculations

5.4 Independence of irrelevant alternatives

The MNL models are based on the assumption that outcomes are independent of other alternatives (see section 3.1.1.2). The H_0 hypothesis in table 5.4 is that the odds ($Outcome - J$ vs $Outcome - K$) are independent of the other alternatives. The negative χ^2 statistic in table 5.4 is common and Hausman & McFadden (1984) conclude that this result is evidence that the assumption of IIA is not violated (Long & Freese, 2001).

For the Hausman test the assumption of IIA is violated for all categories except paddy in the *model 2000* (table 5.4). This result indicates that the inclusion of the category grass will lead to mutually dependent disturbances. Grass and another land-cover class probably share unobserved characteristics, which leads to mutually correlated disturbances.

In the *model 1990* IIA is violated for the category mixed agriculture and grass (table 5.4). The Small-Hsiao test also provides mixed evidence for the IIA assump-

Table 5.4: Hausman test of IIA assumption

Omitted	χ^2	df	$P > \chi^2$	Evidence
Model 2000^a				
Mixed agriculture	226.938	43	0.000	against H_0
Paddy	20.580	53	1.000	for H_0
Open forest	118.037	49	0.000	against H_0
Grass	89.346	50	0.000	against H_0
Model 1990^a				
Mixed agriculture	562.289	39	0.000	against H_0
Paddy	13.144	41	1.000	for H_0
Open forest	1,509.097	47	0.000	against H_0
Grass	-705.206	50	1.000	for H_0

Source: Own calculations

^aIf $\chi^2 < 0$, the estimated model does not meet asymptotic assumptions of the test.

tion, depending on the subsamples of the data taken.⁵ This thesis has to rely on the suggestion of [McFadden \(1973\)](#) that MNL models can only be used when outcome categories “*can plausibly be assumed to be distinct and weighed independently in the eyes of a decision maker*”. The strong believe in the quality of the land-cover data at hand legitimates the use of the MNL model. Problematic is the quasi-continuous regeneration of abandoned shifting cultivation fields into mixed grassland.

5.5 Model shortcomings

To relax the assumptions of IIA, the data can be estimated in a multinomial probit (MNP) or nested logit (NL) formulation. The MNP requires the calculation of $m-1$ multivariate normal probabilities, with m being the number of categories of the dependent variable ([Greene, 1997](#)). For more than three outcomes this quickly gets computationally infeasible and the author did not manage to get an MNP converg-

⁵Results from the Small-Hsiao test are not presented here. Taking a “suitable” subsample allows for the acceptance of the IIA assumption. Other subsamples violate IIA.

ing despite employing different sample sizes and model specifications.⁶ NL models partially relax the IIA assumption by grouping categories into subgroups. The IIA assumption holds within groups, but not across groups. NL is useful when some alternatives are similar to other alternatives in unobserved factors. For the land use data at hand, categories could be nested as forest (closed and open), agriculture (mixed and paddy) and grass. For the data set at hand, NL did not converge. With a narrower spatial sample, which results in significantly more observations, convergence was achieved. However, problems resulting from spatial autocorrelation are assumed to have increased as closer neighbors had to be selected in the regular spatial sample. The NL model involves additional costs through the increase in calculation time and complicates post-estimation and interpretation. Nevertheless, a converged NL model with above data and every fourth cell selected in the regular spatial sample did not show different coefficient signs. Standard errors and significance levels, on the other hand, differed for some of the variables.

The model is estimated in a static fashion although land use is constantly changing, reversible and dynamic. Panel data models combine time series data with cross-sectional observations. An estimation of fixed and random effects models provide greater flexibility in modeling differences across individuals (Greene, 1997). Panel models also require to deal with issues of serial autocorrelation and heterogeneity. Because of this, and little potential efficiency gains with the short panel at hand ($T = 3$), these models were not considered.

An explicit incorporation of the spatial variation of prices in the model framework can allow to examine the environmental impacts stemming from changes in agricultural, trade or macroeconomic policies through their effect on input and output prices (Chomitz & Gray, 1996). The collection of data for a continuous *price surface* was not considered to justify the additional efforts involved in collecting the relevant data given the time-span of this research project. The approximation of the cost-surface using estimated transport costs as one major factor determining input and output prices is believed to sufficiently capture the spatial variation in prices.

⁶The MNP was estimated using the software package LIMDEP[®] (version 7.0), which provides code to estimate MNP models while Stata 8 lacks built-in code to estimate these models.

5.6 Summary

Over time, agricultural land use became more locationally concentrated (fewer patches) with potential environmental benefits for preserving the integrity of ecosystems and endangered species populations. The changes in land use show that shifting cultivation as the traditional farming system practiced by the indigenous population in the research area almost entirely disappeared in its traditional form during the last decade (Müller & Zeller, 2002).

Population increases the likelihood of agricultural land uses at locations with high population density in the *model 2000* while the effects of population are small in the *model 1990*. Technological progress has a land-saving effect, manifested by less land under agricultural production at higher population densities. Better access to roads yields the expected significant coefficients rendering forests more probable further away from all-year roads.

Both models inhibit acceptable levels of prediction accuracy with most prediction mistakes at the borders of the discrete land-use categories and with most incorrectly predicted cells in the land category grass. The assumption of independent outcome categories can not be completely rejected with some evidence of a violation of this assumption in both models. Therefore, some coefficients of the models might be inefficient, but not inconsistent.

CHAPTER 6

Policy simulations

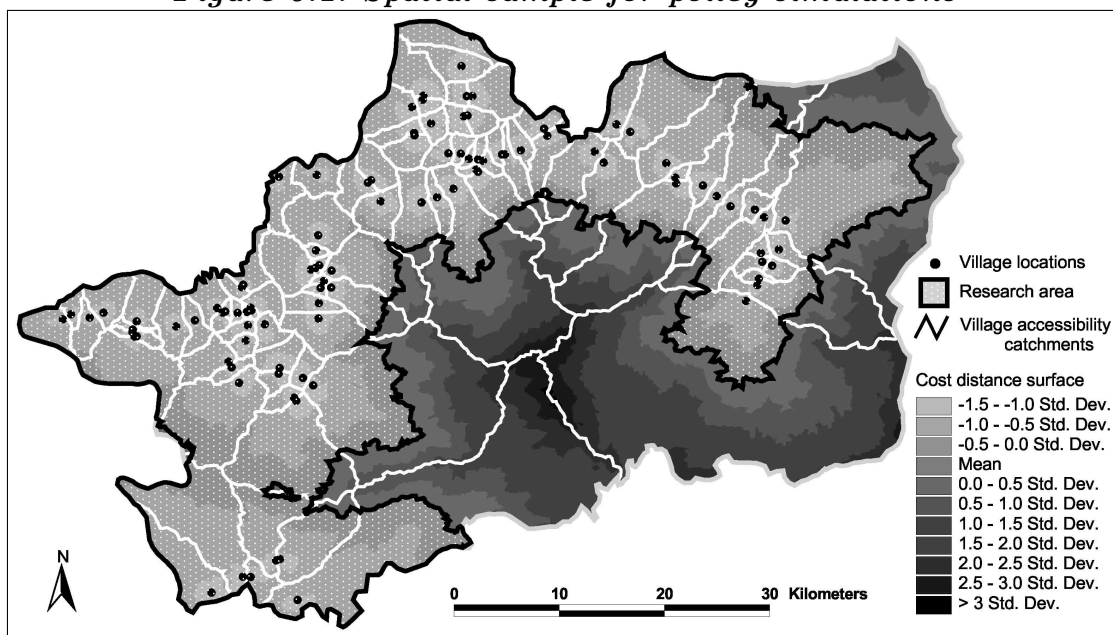
The objective of this chapter is to simulate influences on land use stemming from potential policy interventions. The underlying question of the simulations is “what would have happened if...” as the prediction of future land use – “what will happen if...” – would necessitate the extrapolation of the dependent and independent variables. The focus of the simulations is on areas already cultivated and the adjoining forest margins where most land-use changes relevant for policy makers took place according to chapter 5. A different spatial sample is drawn to focus on changes at these locations, which were demonstrated to be influenced by human interventions. To reduce model complexity and to ease the interpretation of results the land-cover classes are aggregated into the three classes *agricultural land*, *paddy* and *non-agricultural land*. A third difference to the analysis presented in chapter 5 is that the simulations are limited to the second period from 1990 to 2000 as this is the relevant period for potential policy measures. In addition, the previous chapter revealed that model and prediction accuracy is of higher quality in the second model. Similar measures are presented to evaluate model fit. The consequences of the selected policy interventions are, in addition to the prediction matrices already presented in chapter 5, also demonstrated with prediction maps, which assess the predicted locational changes in land use as a result of the policy measures.

6.1 Model specifications

The MNL model in this chapter is estimated for three land-cover classes in contrast to the preceding chapter. To focus on agricultural areas, the land-cover maps are aggregated into *mixed agricultural land*, *wetrice (paddy) areas* and all other land, hereafter referred to as *non-agricultural land*.

A different spatial sample is drawn to limit the analysis to pixels, which are below the mean of the transportation cost surface in the two districts (figure 6.1, for the description of the cost surface see section 3.3.1). In that way, virtually all the agricultural area in 2000 and most of the agricultural land-use changes between 1992 and 2000 are included (compare to figure 3.2). Remote and high mountainous areas covered mostly with thick primary forest are eliminated from the analysis. Applying this sample captures most areas with anthropogenic influences in the last decade and exclude areas where hardly any land-use changes took place due to human interventions. The regular spatial sample is drawn in the same way as in chapter 5 using Besag's coding scheme and every fifth cell in X and Y direction is selected. These spatial selection procedures results in a final data set of 22,255 grid cells used for the econometric estimations in this chapter.

Figure 6.1: Spatial sample for policy simulations



Source: Author

The model is – as a consequence of the estimation results obtained in section 5.2 and due to the slightly different objectives stated above – estimated with a different set of independent variables. Excluded are the age of villages, the indicator for political capital (*hero villages*), the years since the introduction of primary schools, the interaction terms of forest protection and the X and Y coordinates. Additional variables included in the MNL model of this chapter are the years since the establishment of a New Economic Zone (NEZ) within a village area. The protection term is calculated only for the protected area *Nam Ca*, which is almost completely located within the spatial sample for the simulation models. Another variable is computed as the product of the amount of irrigated area per village and the age of a reservoir as an indicator for the number of irrigated hectares times the duration of their irrigation, again for each village area. Landscape fragmentation is measured as the perimeter-area ratio of land cover patches (see section 4.3.1). Descriptive statistics of the dependent and independent variables of the new sample are presented in table A.8 in the appendix.

Since observations within the villages are likely not independent, the estimation might underestimate standard errors. To account for this, the MNL model is estimated using robust standard errors and adjusted for clustering on village level.¹ Robust standard errors, also called Huber-White or sandwich estimator of variance, provide correct standard errors in the presence of violations of model assumptions like an incorrectly specified probability density function (Gould & Sribney, 1999; StataCorp, 2003). Village clustering relaxes the assumption of independence of observations and produces correct measurements of the standard errors even if observations are correlated (StataCorp, 2003). Robust standard errors clustered for village areas still require pixels to be independent across the clusters.

6.2 Empirical results from base estimation

Estimation results of the base estimation are presented in table A.9 of the appendix. Coefficients for the geophysical variables are significant at the one percent level and show the expected signs with high predictive power. Agriculture is more likely at lower altitudes, flatter land, and more suitable soils, the same result as in the two models of chapter 5. A higher amount and less variance of rainfall decreases the likelihood of non-agricultural land uses. Surprisingly, access to all-year roads did not

¹For this hint the author is grateful to John L. Pender, personal communication, 2002.

have a significant effect on the probability of a certain land-use class. This might be due to the relatively large areas along parts of the all-year road network, which are occupied by non-agricultural land uses, because of low population densities and a low potential for agricultural production due to adverse geophysical characteristics. Earlier introduction of mineral fertilizer strongly increases the likelihood that pixels are under agricultural production and more irrigated area increases the likelihood of paddy.

Higher population increases the probability of the amount of area under agricultural land uses relative to non-agricultural land, consistent with the results from the previous chapter. The years since the establishment of a NEZ increase the likelihood of mixed agriculture relative to other uses and decrease the likelihood of paddy, however, the effects are small. Earlier introduction of NPK increases the likelihood of paddy against all and mixed agriculture against non-agricultural land. This seems to contradict the results from the estimations in chapter 5, due to the different sample with substantially less areas of non-agricultural land uses. The interaction term between the age of reservoirs and the amount of irrigated area increases the likelihood of paddy. The indicator on landscape fragmentation renders non-agricultural land uses more probable. As expected, more fragmented patches of land were more likely to convert into non-agricultural land uses in the last decade.

Table 6.1: Prediction matrix for base estimation

Observed land use	Predicted land use			Total	Correctly predicted cells
	Mixed agriculture	Paddy	Non-agricultural land		
Mixed agriculture	4,270	333	815	5,418	79%
Paddy	704	1,010	48	1,762	57%
Non-agricultural land	875	41	14,225	15,141	94%
Total	5,849	1,384	15,088	22,321	87%

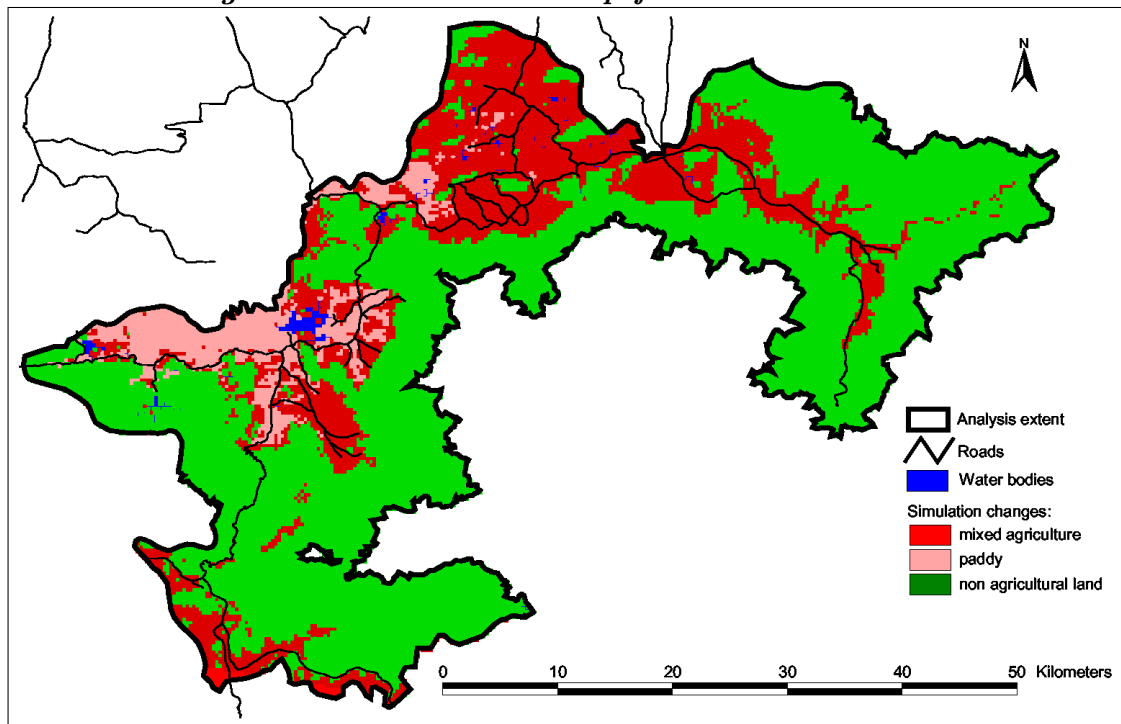
Source: Own calculations

The prediction matrix in table 6.1 tabulates actual land use against predicted land use, where correctly predicted cells are assumed to be the cells with the highest probability. The matrix reveals that 79% of the cells are correctly predicted for

the category mixed agriculture. Prediction accuracy is 57% for paddy and 94% for non-agricultural land. 87% of all analyzed cells are accurately predicted to be in the observed land use category. 60% (equivalent to 875 + 815 or 1,690 cells) of incorrect predictions belong to the classes of mixed agriculture and non-agricultural land and a further 37% (704 + 333) between mixed agriculture and paddy.

Calculating predictions for every cell and mapping these predictions with GIS software yields *prediction maps* of land use. Figure 6.2 is the visual equivalent of the bottom row in table 6.1 and maps the predictions for the base estimation. This is a graphical way to illustrate predicted land use in a spatial manner and also enables a visual comparison with the observed land use maps in figure A.1 of the appendix.

Figure 6.2: Prediction map for base estimation



Source: Own calculations

With the spatial sample employed in this chapter the Hausman test for IIA suggest that categories are independent of other alternatives (table 6.2). Therefore, the H_0 hypothesis that the odds ($Outcome-J$ vs $Outcome-K$) are independent of other categories can be accepted and, according to the Hausman test, adding another category to the model will not change the odds between alternatives.

Table 6.2: Hausman test of IIA assumption, base prediction

Omitted	χ^2 ^a	df	$P > \chi^2$	Evidence
Mixed agriculture	-48.686	15	1.000	for H_0
Paddy	-1.084	15	1.000	for H_0

Source: Own calculations

^aIf $\chi^2 < 0$, the estimated model does not meet asymptotic assumptions of the test.

6.3 Description of scenarios

The impact of likely changes in socioeconomic variables on land use can be simulated by changing values of selected explanatory variables. This implies that the estimations in this chapter do not predict future land use, but rather predict land use dependent on the past variations in policy indicators. Thus, the question here is rather “what would have happened, if...” than “what will happen if...”. To predict future land use in the econometric framework employed here, both LHS and RHS variables would have to be extrapolated into the future.² New predictions and probabilities for land use can be generated by using the estimated coefficients from the base estimation in section 6.2. A comparison of predicted land use from the simulations with the predicted land use from the base estimation yields an approximation for the effects of varying levels of explanatory variables on land use (Nelson et al., 1999). In spatially explicit land-use models the results are spatially explicit, which allows for the assessment of locational changes in land use as a result of policy interventions.

Four policy scenarios are distinguished as potential rural development interventions:³

1. The **first scenario** assumes *low levels of government investments* in agricultural intensification, which is proxied by a 50% reduction of the area under irrigation and a five year delay on the introduction of NPK fertilizers;

²The prediction of future land use requires different modeling approaches than the one employed here. One potential class of modeling approaches are agent-based simulation models, which allow to model the agents’ interactions in a spatially explicit dynamic setting (Berger, 2001).

³An upgrading of seasonal roads to year-round access is not simulated as varying road access would not result in measurable land-use changes due to the insignificant coefficients of distance to road in the base estimation (see the estimation results in table A.9 of the appendix).

2. The **second scenario** involves *high levels of government investments*, proxied by a 50% increase of the area under irrigation and a five year earlier introduction of NPK fertilizers;
3. The **third scenario**, subsequently called *protection scenario*, follows a government guideline that discourages agricultural production above slopes of 15 degrees. It is assumed here that sloping land above 15 degrees and, in addition, all forest with a closed canopy cover are added to already existing protected areas;
4. **Scenario four** combines *increasing agricultural intensification* (scenario 2) with additional *forest protection* (scenario 3) to assess the effect of this policy strategy, often mentioned as a “first-best” policy scenario in Vietnam.

6.4 Scenario results

A comparison of simulated changes with the baseline predictions (figure 6.2 and table A.9 in the appendix) is expected to shed light on the impact of the four rural development policy scenarios on land use. The spatially explicit framework allows for an assessment of both the magnitude as well as the location and spatial arrangement of the simulated changes. This will enable the identification of areas, which are likely to change into other forms of land use under a certain policy setting. Results of the four policy scenarios are discussed in more detail in the following sections and illustrated as *simulation matrices* (defined in analogy to the prediction matrices in the previous chapter and in table 6.1) and *simulation maps*, which visualize the modifications between the estimated land use from the base prediction and the simulated land use.

Simulated land use in all four scenarios is assumed to be the land use, which has the highest predicted probability. Grey pixels in the figures A.4 to A.7 in the appendix indicate no change between the base prediction and simulated predictions. The maps resulting from the econometric estimations contain one prediction of simulated land use in non-contiguous cells every 200 meters due to the spatial sample, in which every fifth cell in X and Y direction is selected. The pixels containing the predictions are then expanded by 100 meters (or two cells) in X and Y direction to arrive at a continuous surface, which enables a better visual comparability. The

expansion of pixels is solely a tool to facilitate a graphical description of the changes resulting from the four simulations.

6.4.1 Low investments in irrigation

The *simulation matrix* in table 6.3 illustrates the results from the first simulation, which approximates a decrease of government investments in agricultural intensification. The predicted land use from the base estimation is depicted in the rows and the simulated land use in the columns.

Table 6.3: Scenario 1 – Low levels of government investments

Base prediction	Simulated prediction			Total
	Mixed agriculture	Paddy	Non-agricultural land	
Mixed agriculture	5,740	0	240	5,980
Paddy	190	1,214	3	1,407
Non-agricultural land	0	0	15,274	15,274
Total	5,930	1,214	15,517	22,661

Source: Own calculations

The scenario of *low levels of government investments* resulted in a modest increase in non-agricultural land by 240 pixels and a conversion of 190 pixels under paddy into other agricultural land uses (table 6.3).

The simulation map in figure A.4 in the appendix shows the resulting locational changes of the first scenario. The *low investment scenario* results in a regeneration of forested land on marginal lands further away from the major roads as can be seen in the eastern part of figure A.4. Some of the paddy area converts into mixed agricultural land, mirroring the *lower intensity level* of a reduced area under irrigation and later fertilizer introduction.

6.4.2 High levels of government investments

In the second scenario of *high levels of government investments*, the effects of agricultural intensification encourage farmers to bring more land under paddy cultivation and to expand mixed agricultural production at the expense of non-agricultural land

uses (see table 6.4). The effect seems to be small in absolute terms as the area under paddy “only” increases by 320 pixels (1,727 – 1,407 in table 6.4). Still, this is equivalent to an increase of almost 19%, a significant augmentation considering the high land intensity and productivity of paddy cultivation. In addition, the higher investments in irrigation and NPK stimulate the conversion 293 previously forested pixels into mixed agricultural land.

Table 6.4: Scenario 2 – High investments in irrigation

Base prediction	Simulated prediction			Total
	Mixed agriculture	Paddy	Non-agricultural land	
Mixed agriculture	5,664	316	0	5,980
Paddy	0	1,407	0	1,407
Non-agricultural land	293	4	14,977	15,274
Total	5,957	1,727	14,977	22,661

Source: Own calculations

Land use also shows more homogeneous patterns (least patches) in this scenario (figure A.5 in the appendix).⁴ Non-agricultural land is converted to agricultural uses at the agricultural-forest frontier and mixed agricultural land use is converted into paddy land on areas with suitable topographic and soil conditions.

Obviously, the first and second scenario move in opposite directions as the simulated changes are the opposite. The resulting spatial arrangement and the locational changes are, however, of interest in these two scenarios as they do *not* move in opposite direction, but show varying magnitude and distinct spatial patterns (figure A.4 and A.5 in the appendix).

⁴As only non-contiguous observations are selected, the analysis of landscape metrics is not entirely valid as it is unknown what happens between the selected observations. It is mentioned here, when some of the simulation maps clearly show distinguishable spatial patterns. Combined with the results from the regressions the author assumes that spatial fragmentation in this scenario is decreasing.

6.4.3 Effect of increased forest protection

The scenario of *forest protection* implies a drastic increase in protected forest areas from 8% in the base scenario to 51% in the simulated scenario. Yet, most of this protection is on steep slopes unsuitable for agricultural production. Control and enforcement of forest protection are not considered as the analysis in the simulations only takes the coefficients from the base estimation. Agricultural land uses decrease by modest 124 pixels or 2.2% (table 6.5) at the benefit of non-agricultural vegetation. Protected areas – if assumed to be located according to slope and existing forest cover – increase forest cover on more marginal lands along the agricultural-forest frontier (figure A.6 in the appendix).

Table 6.5: Scenario 3 – Increased forest protection

	Simulated prediction			Total
	Mixed agriculture	Paddy	Non-agricultural land	
Base prediction				
Mixed agriculture	5,856	0	124	5,980
Paddy	0	1,404	3	1,407
Non-agricultural land	0	0	15,274	15,274
Total	5,856	1,404	15,401	22,661

Source: Own calculations

The apparently increasing number of patches (see footnote 4) leads to a more scattered spatial arrangement, which is less desirable from an ecological viewpoint. The result suggests that forest protection strategies ought to be combined with ecological valuations that explicitly take into account the value of contiguous protected areas for conserving precious biodiversity as opposed to spatially scattered protected areas.

6.4.4 Effect of high levels of government investments combined with forest protection

Significant changes in the spatial patterns of land use result from *increasing investments in agricultural intensification combined with increased forest protection* (see table 6.6 and figure A.7 in the appendix). This scenario induces a reduction in the

area under mixed agriculture at the extent of rice and, simultaneously, increases the area under non-agricultural land uses, most of it previously covered with forests.

Table 6.6: Scenario 4 – High investments in irrigation combined with forest protection

	Simulated prediction			Total
	Mixed agriculture	Paddy	Non-agricultural land	
Base prediction				
Mixed agriculture	5,191	764	25	5,980
Paddy	0	1,407	0	1,407
Non-agricultural land	523	23	14,728	15,274
Total	5,714	2,194	14,753	22,661

Source: Own calculations

Overall, agricultural land uses decrease by 25 pixels compared to the base situation. 564 pixels shift from non-agricultural land uses to agriculture and 764 pixels from mixed agricultural land use to paddy. Both scenarios, which involve investments in agricultural development, and here specifically investments in land-saving technologies, might in addition have a considerable effect on the intensity of land use on a plot of land, and less on the expansion of the area under agriculture. The spatial approach for the analysis of locational changes taken in this study will not reveal these intensification effects on already cultivated land, which will result, for example, in increasing crop productivity.

6.5 Summary

The simulations mostly resulted in the expected changes compared to the baseline prediction. However, considerable alterations in the spatial patterns are revealed from the maps depicting the simulated land use changes. A thorough observation of these changes in space together with field verifications could potentially yield important insights for the formulation of rural development policies. Both scenarios of high and low investments in irrigation had a strong influence on the area under paddy rice, by far the most important food crop for farmers in the research area. The forest protection scenario increases the patchiness of the landscape and induced

higher pressure on agricultural land around the simulated additional areas under forest protection.

This kind of spatial simulations can help to generate an understanding of the geographical consequences of policy actions. For that reason, this analysis is potentially useful when combined with field verifications to identify the validity of the results and to investigate the identified hot spots in more detail. The various socioeconomic indicators included in the spatially explicit data set at hand allows for the simulation of these variables, which are of crucial importance to the interests of decision makers. In that way, this methodology can assist planners to assess the severity of tradeoffs in the environment-development nexus. More disaggregated spatially explicit data on relevant socioeconomic indicators could further improve the viability and strengths of the predictions.

Conclusions and Policy Implications

This dissertation examines exogenous driving forces of land-cover and land-use change in Dak Lak province in the Central Highlands of Vietnam. The objective is to reveal the influence of socioeconomic, agroecological, geophysical, and policy variables on land use. Specifically, several hypotheses of land-use change and its related driving forces are tested to assess the options of agricultural expansion and intensification as land-use responses of local agents. These responses are considered an agent's reaction to increasing population density, technological progress and improved market access. To achieve these aims, novel approaches are developed for combining data from a survey in randomly selected villages with remote sensing data derived from Landsat satellite images and spatially explicit Geographic Information System (GIS) data. The application of GIS techniques to integrate location and the spatial characteristics of objects provides additional insights on land-use aspects and yields spatially explicit results.

Econometric analysis is carried out using a reduced-form, spatially explicit multinomial logit model to estimate the direction and strength of various exogenous and predetermined driving forces on changes in land use. The effects of potential policy interventions and the quantity and location of resulting land-use changes are simulated to facilitate better geographical targeting of rural development policies. Improved information about processes and patterns of land-use changes should pro-

vide decision makers with a better planning basis for the implementation of various rural development and resource management policies.

The results from satellite image interpretation show a decrease in open canopy forest in the first period due to conversion into rain-fed and irrigated agricultural land. Forest cover increased during the 1990s by 8%, mainly due to natural regeneration of mixed grassland. Fallowed fields formerly used for shifting cultivation were largely abandoned in the last decade and regenerated into bamboo and open canopy secondary forest. Overall, rainier mixed agriculture remained constant in the 90s, as the reduction in shifting cultivation was compensated by an expansion of permanent cropping area. Over time, agricultural production became more locationally concentrated (fewer patches) with potential environmental benefits for preserving the integrity of ecosystems and endangered species populations. These changes in land use show that shifting cultivation as the traditional farming system practiced by the indigenous population in the study area almost entirely disappeared in its traditional form during the last decade. According to village interviews degradation in the resource base sustaining rural livelihoods is occurring closer to populated areas. Evidence is found for a decrease in forest quality and timber volume, presumably due to increasing demand for forest products.

The distribution of socioeconomic and policy variables demonstrate the diverse characteristics of villages. Over the last two decades, input use and agricultural productivity increased across most villages. Wealth levels, food security and school enrollment rates as well as market access improved. Yet, significant differences in above indicators prevail across ethnic groups and with increasing distance to all-year roads. Most of these differences evolved since liberation day and enlarged over time, in particular during the last decade. To prevent future social unrest more government efforts are required to narrow the welfare gap between indigenous ethnic groups and ethnic Vietnamese. This could be achieved through enhanced extension in agriculture and forestry, the devolution of secure long-term property rights for forest land and better supply and access to external capital sources with less demanding formalities. The shift to market-oriented production, which in particular many ethnic villages are currently undergoing, requires intensified investments in advisory services to adapt to the needs of commercialized agricultural production and marketing including the provision of comprehensible and accessible market information.

The effect of ethnic minority villages on land use is highly significant in both models. Ethnic villages are on average closer to forested and grass areas and further away from all-year roads and political centers. Forest regeneration at the expense of agricultural area predominantly occurs closer to ethnic villages. The increase in forest cover over the last decade despite the observed population growth does not correspond to the widely stated positive correlation between higher population density and lower forest cover. It instead supports the findings from [Pender et al. \(2001b\)](#) that agricultural development was induced by market and technological forces rather than by population pressure. Permanent conversion of primary forest into agricultural land was hardly observed. When it was, it took place between liberation day and 1990, the key period of government resettlement programs.

Access to all-year roads improved substantially in the last decade thereby facilitating market integration, access to infrastructure, agricultural inputs and public services. The investments in irrigation and infrastructure, combined with improved access to markets and services, were successful in intensifying agricultural production. Higher agricultural productivity on existing land reduced the need for shifting cultivation, thus preserving forest cover while sustaining a much greater population on virtually the same agricultural land area. This was possible through increases in the productivity of food and cash crops (mainly rice), as well as an increase in cash crop production (mainly maize and, to a lesser extent, coffee). Because of this, upland arable area reduced slightly, whereas much of the grassland of the earlier period (due to shifting cultivation) regenerated into open canopy and closed forest. The establishment of protected areas in the last decade contributed to this development by decreasing the likelihood of non-forest land uses. Thus, the policy measures have contributed to a reduction in shifting cultivation and facilitated agricultural intensification. [Zeller et al. \(2000\)](#) find similar linkages between agricultural intensification and rural income diversification on the one hand and reduced pressure on forests and soils on the other hand in one of their research areas in Madagascar.

In the first decade a pathway of land expansion into previously uncultivated areas can be observed. Low population density facilitated extensive land cultivation based on forest fallows. Few farmers applied intensive cultivation techniques and production was subsistence-oriented. The last decade brought about significant changes in the policy environment like investments in road improvements, irrigation facilities and introduction of new technologies. This period is characterized by inten-

sification without further land expansion and allowed forests to regenerate. Today, forests cover nearly the same area as in 1975. The land-saving effect of technological progress prevailed over the labor-saving effect. In other words, labor-intensive technical progress diverted resources away from land intensive farming systems like the various forms of shifting cultivation by augmenting the return on labor. The growth in food production outpaced population growth. Today, more people produce more per capita than ten years ago on virtually the same amount of land under cultivation.

Improvements and extension of the road network did not significantly reduce forest cover as it was outweighed by intensification on existing land. Road development seems to contribute significantly to improvements in agricultural productivity and welfare. In this manner, better roads can reduce poverty and food insecurity while at the same time curbing land expansion and relaxing the pressure from forests. At the same time, the shift to a market-driven economy in 1986 (*doi moi*), state policies on sedentarization and forest protection encouraged production and made traditional farming techniques more difficult to pursue, thus further contributing to the transformation from traditional subsistence-based into market-driven production. The construction of reservoirs and improvement of irrigation channels enabled and facilitated agricultural development. Rice yields more than doubled in the study region in the last decade and land intensity of paddy rice production rose. Intensification from improved irrigation systems may have greatly reduced the pressure to expand cultivated area at the expense of forest.

These results show that intensification of agriculture – combined with the protection of forested areas – can reduce the pressure on forested land, and slow down or even halt the expansion of agricultural land, if coupled with the right policy instruments. The analysis reveals that investments in infrastructure can facilitate this much-needed intensification of agriculture without detrimental consequences on forest resources.

The two districts where the research project took place were purposively selected and are not representative of the entire province. The secondary data and satellite images of other districts in Dak Lak Province not presented in this thesis clearly show a less positive picture with respect to several environmental indicators since forest cover there continued to decline during the last decade. This is presumably due to less mountainous topography and the widespread existence of the basaltic soil

in many other districts, which are suitable for the production of various cash crops. For that reason, the area allocated to the plantation of coffee and pepper enlarged rapidly in other districts in the 1990s at the expense of forest cover. Therefore, the findings relate to one part of Dak Lak province and cannot be generalized to all of the Central Highlands.

Large inequalities in geophysical and agroecological conditions, in socioeconomic indicators and the state of natural resources are also observed within the research area. This spatial heterogeneity both in Dak Lak province and the research area calls for increased efforts of spatial planning in rural development activities and the management of natural resources. Spatial planning can increase the productivity of public policy efforts and provide the possibility for spatial decision-making, which in turn can lead to better geographical targeting of the allocation of limited resources. As in other parts of the world, comprehensive, consistent and up-to-date geographic data is increasingly available across Vietnam. To institutionalize adequate spatial planning and decision-making several constraints have to be addressed: First, more investments in human capital formation are required to relieve the lack of adequate human resources, particularly on provincial and district level. Second, practices in data collection, input, management, updating, validation and dissemination have to become more transparent to acquire data in desired quality and sufficient quantity. Third, organizational development, and here especially inter-departmental cooperation, has to be bolstered to achieve efficiency gains by means of better administration, information flow, cooperation, coordination and, in the end, decision-making.

Econometric modeling with larger samples no longer poses serious computational problems because of advances in disk space, processing speed and software capabilities. Remote sensing scanners of resolutions as used in this study are available at low costs for the public and continue to increase in quality. The potential drawbacks of spatially explicit modeling and spatial statistics have more to do with the combination of data from natural sciences with socioeconomic information at an acceptable scale of analysis that best represents the underlying processes with minimal error.

The integration of socioeconomic data with continuous spatial layers poses several challenges. The approach employed to combine interpolation techniques with accessibility catchments on village level addresses this issue. The approximation of village borders based on estimated transport costs defines village areas according

to the purpose of this study and is an attempt to capture the influence of spatially varying accessibility on land use. It might be a reasonable approach considering the size of the research area. However, this aggregation masks decision-making processes at farm and plot level. A combination of household level data with more GPS data for field boundaries would improve the strength of the overall model. Census data with observations for all entities allow for the generation of continuous surfaces without “holes” at locations of non-sampled observations. In that way, censuses narrow the gap between the data resolution and the level of decision-making. Continuous surfaces have the additional advantage of superior possibilities for geovisualization of data and results, one of the many strengths inherent in GIS software packages.

The present data sets allow for a variety of possible extensions. Dynamics of forest-cover change could be more deeply examined with binary outcome models in a panel formulation. With regard to the assumptions of independence of irrelevant alternatives, a multinomial probit or nested logit approach will yield more consistent estimates, but increase computation time and add complexity to post-estimation and interpretation. Inferences about underlying causalities are still difficult to make given the model used in the analysis. A combination with structural models including endogenous variables will yield further insights to improve policy analysis in the future. An additional consideration might be a thorough examination of the placement of government programs to improve investment decisions. However the analysis presented here highlights the usefulness of spatially differentiated models of land-use change for agricultural and rural development analysis.

Bibliography

- ANGELSEN, A. (2001): Playing games in the forest: State-local conflicts of land appropriation. In: *Land Economics*, Vol. 77 (2), pp. 285–299.
- ANGELSEN, A. and D. KAIMOWITZ (2001): Agricultural Technology and Forests: A Recapitulation. In: ANGELSEN, A. and D. KAIMOWITZ (Eds.): *Agricultural Technologies and Tropical Deforestation*, pp. 383–402. CAB International, Oxon.
- ANSELIN, L. (1988): *Spatial Econometrics: Methods and Models*. Studies in Operational Regional Science. Kluwer Academic Publishers, Dordrecht.
- ANSELIN, L. (2000): Spatial Econometrics. In: BALTAGI, B.H. (Ed.): *Companion to Econometrics*, pp. 310–330. Basil Blackwell, Oxford.
- ANSELIN, L. (2001a): Issues in Spatial Probit Models. In: *Paper presented at Informal Workshop on Qualitative Dependent Variable Estimation and Spatial Effects*. Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign.
- ANSELIN, L. (2001b): Spatial Effects in Econometric Practice in Environmental and Resource Economics. In: *American Journal of Agricultural Economics*, Vol. 83 (3), pp. 705–710.

- ANSELIN, L. (2002): Under the hood: Issues in the specification and interpretation of spatial regression models. In: *Agricultural Economics*, Vol. 27 (3), pp. 247–267.
- ARONOFF, S. (1995): *Geographic Information Systems: A Management Perspective*. WDL Publications, Ottawa.
- AYLWARD, B. (2000): Economic Analysis of Land-use Change in a Watershed Context. Tech. Rep., World Commission on Dams, Kuala Lumpur, Malaysia.
- BAKKER, W.H., L. F. JANSSEN, M. J. WEIR, B. G. GORTE, C. POHL, T. WOLDAI, J. A. HORN and C. V. REEVES (2000): *Principles of Remote Sensing*. The International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede.
- BALAND, J.M. and J.-P. PLATTEAU (1998): Division of the Commons: A Partial Assessment of the New Institutional Economics of Land Rights. In: *American Journal of Agricultural Economics*, Vol. 80, pp. 644–650.
- BERGER, T. (2001): Agent-based spatial models applied to agriculture: a simulation tool for technology diffusion, resource use changes and policy analysis. In: *Agricultural Economics*, Vol. 25, pp. 245–260.
- BERGERON, G. and J. L. PENDER (1999): *Determinants of Land Use Change: Evidence from a Community Study in Honduras*. Discussion Paper No. 46. Environment and Production Technology Division, International Food Policy Research Institute, Washington, D.C.
- BESAG, J. (1974): Spatial interaction and the statistical analysis of lattice systems. In: *Journal of the Royal Statistical Society*, Vol. Series B 36, pp. 192–236.
- BIGMAN, D. and U. DEICHMANN (2000): Spatial Indicators of Access and Fairness for the Location of Public Facilities. In: BIGMAN, D. and H. FOFACK (Eds.): *Geographical Targeting for Poverty Alleviation: Methodology and Applications*, pp. 181–206. The World Bank, Washington, D.C.
- BIGMAN, D. and H. FOFACK (2000): *Geographical targeting for poverty alleviation: Methodology and applications*. World Bank Regional and Sectoral Studies Series. World Bank, Washington, D.C.

- BILSBORROW, R.E. and D. L. CARR (2001): Population, Agricultural Land Use and the Environment in Developing Countries. In: LEE, D. and C. BARRET (Eds.): *Tradeoffs or Synergies? Agricultural Intensification, Economic Development and the Environment*, pp. 35–55. CAB International, Wallingford (U.K.).
- BIRDLIFE INTERNATIONAL (2001): *Sourcebook of Existing and Proposed Protected Areas in Vietnam*. Birdlife International – Vietnam Programme and Forest Inventory and Planning Institute, Hanoi, Vietnam.
- BIRNER, R. and H. WITTMER (2003): Using Social Capital to Create Political Capital: How Do Local Communities Gain Political Influence? A Theoretical Approach and Empirical Evidence from Thailand. In: DOLŠAK, N. and E. OSTROM (Eds.): *The Commons in the New Millennium – Challenges and Adaptation*, pp. 291–334. MIT Press, Cambridge, Massachusetts.
- BOSERUP, E. (1965): *The Conditions of Agricultural Growth*. George Allen & Unwin Ltd, London.
- BRACKEN, I. (1994): A surface model approach to the representation of population-related social indicators. In: FOTHERINGHAM, S.A. and P. ROGERSON (Eds.): *Spatial Analysis and GIS*, pp. 247–259. Taylor & Francis, London.
- BRIASSOULIS, H. (2000): *Analysis of Land Use Change: Theoretical and Modeling Approaches*. The Web Book of Regional Science. Regional Research Institute, West Virginia University, www.rrri.wvu.edu/regscbooks.htm, accessed: Nov. 2001.
- BURMEISTER, L.L. (1987): The South Korean Green Revolution: Induced or Directed Innovation?. In: *Economic Development and Cultural Change*, Vol. 35 (4), pp. 767–790.
- CHOMITZ, K.M. and D. GRAY (1996): Roads, Lands Use, and Deforestation: A Spatial Model Applied to Belize. In: *World Bank Economic Review*, Vol. 10(3), pp. 487–512.
- CHOMITZ, K.M. and T. S. THOMAS (2001): *Geographic Patterns of Land Use and Land Intensity in the Brazilian Amazon*. Development Research Group. World Bank, Washington, D.C.

- CHRIST, H. and D. KLOSS (1998): *Land Use Planning and Land Allocation in Vietnam with Particular Reference to Improvement of its Processes in the Social Forestry Development Project Song Da (SFPD)*. Consultancy Report on behalf of the Social Forestry Development Project Song Da (SFDP), Hanoi, Vietnam.
- DAK LAK STATISTICAL DEPARTMENT (1981): *Statistical Yearbook 1980*. Buon Ma Thuot, Vietnam.
- DAK LAK STATISTICAL DEPARTMENT (1991): *Statistical Yearbook 1990*. Buon Ma Thuot, Vietnam.
- DAK LAK STATISTICAL DEPARTMENT (2002): *Statistical Yearbook 2001*. Buon Ma Thuot, Vietnam.
- DE PINTO, A. and G. C. NELSON (2002): *Correcting for Spatial Effects in Limited Dependent Variable Regression: Assessing the Value of "Ad-Hoc" Techniques*. Paper prepared for the American Agricultural Economics Association annual meeting, Long Beach, California.
- DEICHMANN, U. (1997): *Accessibility Indicators in GIS*. United Nations Statistics Division, Department for Economic and Policy Analysis, United Nations, New York.
- DEININGER, K. and B. MINTEN (2002): Determinants of Deforestation and the Economics of Protection: An Application to Mexico. In: *American Journal of Agricultural Economics*, Vol. 84 (4), pp. 943–960.
- DEMERS, M.N. (2002): *GIS Modeling in Raster*. John Wiley & Sons, New York.
- DO DINH SAM (1994): *Shifting Cultivation in Vietnam: its social, economic and environmental values relative to alternative land use*. Land Use Working Group, Ministry of Forestry and Forest Science Institute, Hanoi.
- ELNAGHEEB, A.H. and D. W. BROMLEY (1994): Extensification of Agriculture and Deforestation: Empirical Evidence from Sudan. In: *Agricultural Economics*, Vol. 10, pp. 193–200.
- ENTWISLE, B., S. J. WALSH, R. R. RINDFUSS and A. CHAMRATRITHIRONG (1998): *Land-Use/Land-Cover and Population Dynamics*, Nang Rong, Thailand.

- In: LIVERMAN, D., E. F. MORAN, R. R. RINDFUSS and P. C. STERN (Eds.): *People and Pixels: Linking Remote Sensing and Social Science*, pp. 121–144. National Research Council, Washington, D.C.
- ERDAS (1999): *ERDAS Field GuideTM*. ERDAS, Inc., Atlanta, Georgia, 5th Edn.
- ESRI (1999): *ArcInfo 8.0.1 On-line Help*. Environmental Systems Research Institute, Inc., Redlands, CA.
- ESRI (2000): *ArcView 3.2 On-line Help*. Environmental Systems Research Institute, Inc., Redlands, CA.
- FELLMANN, G.D., A. GETIS and J. GETIS (2002): *Human Geography*. McGraw-Hill, Boston.
- FOX, J., R. KANTER, S. YARNASARN, M. EKASINGH and R. JONES (1994): Farmer Decision Making and Spatial Variables in Northern Thailand. In: *Environmental Management*, Vol. 18 (3), pp. 391–399.
- GELMAN, A., C. PASARICA and R. DODHIA (2002): Let's Practice What We Preach: Turning Tables into Graphs. In: *The American Statistician*, Vol. 56 (2), pp. 121–130.
- GEOGHEGAN, J., S. C. VILLAR, P. PEIS, B. T. II and C. VANCE (2001): Modeling tropical deforestation in the southern Yucatan peninsular region: comparing survey and satellite data. In: *Agriculture, Ecosystems & Environment*, Vol. 85, pp. 25–46.
- GODOY, R., K. O'NEILL, S. GROFF, P. KOSTISHACK, A. CUBAS, J. DEMMER, K. MCSWEENEY, J. OVERMAN, D. WILKIE, N. BROKAW and M. MARTÍNEZ (1997): Household Determinants of Deforestation by Amerindians in Honduras. In: *World Development*, Vol. 25, pp. 977–987.
- GOULD, W. and W. SRIBNEY (1999): *Maximum Likelihood Estimation with Stata*. Stata Press, College Station, Texas.
- GREENE, W.H. (1997): *Econometric Analysis*. Prentice-Hall International, Inc., Upper Saddle River, New Jersey.

- HAINING, R. (1990): *Spatial data analysis in the social and environmental sciences*. Cambridge University Press, Cambridge.
- HAUSMAN, J.A. and D. MCFADDEN (1984): Specification Tests for the Multinomial Logit Model. In: *Econometrica*, Vol. 52 (5), pp. 1219–1240.
- HINTZE, J.L. and R. D. NELSON (1998): Violin Plots: A Box Plot-Density Trace Synergism. In: *The American Statistician*, Vol. 52, pp. 181–184.
- HOSMER, D.W. and S. LEMESHOW (2000): *Applied Logistic Regression*. John Wiley & Sons, New York, 2 Edn.
- JOHNSON, K. and E. BRIENG (2001): Brewing Discord. In: *TIME Asia*, Vol. 157 (13).
- KAIMOWITZ, D. and A. ANGELSEN (1998): *Economic Models of Tropical Deforestation – A Review*. Center for International Forestry Research, Bogor, Indonesia.
- KELEJIAN, H.H. and I. R. PRUCHA (2001): On the asymptotic distribution of the Moran I test statistic with applications. In: *Journal of Econometrics*, Vol. 104 (2), p. 40.
- KRONG BONG DISTRICT STATISTICAL OFFICE (1999): *Statistical Yearbook of Krong Bong District 1999*. Buon Ma Thuot, Vietnam.
- KRUGMAN, P. (1995): *Development, Geography and Economic Theory*. The MIT Press, Cambridge.
- LAK DISTRICT STATISTICAL OFFICE (1999): *Statistical Yearbook of Lak District 1999*. Buon Ma Thuot, Vietnam.
- LAMBIN, E., X. BAULIES, N. BOCKSTAEL, G. FISCHER, T. KRUG, R. LEEMANS, E. MORAN, R. RINDFUSS, Y. SATO, D. SKOLE, B. TURNER and C. VOGEL (1999): *Land-Use and Land-Cover Change (LUCC) Implementation Strategy*. IHDP publication programme, Bonn.
- LAUDON, K.C. and J. P. LAUDON (1996): *Management Information Systems*. Prentice-Hall, Inc., Upper Saddle River, New Jersey.

- LEE, D.R., C. B. BARRET, P. HAZELL and D. SOUTHGATE (2001): Assessing Tradeoffs and Synergies among Agricultural Intensification, Economic Development and Environmental Goals: Conclusions and Implications for Policy. In: LEE, D.R. and C. B. BARRET (Eds.): *Tradeoffs or Synergies? Agricultural Intensification, Economic Development and the Environment*, pp. 451–464. CAB International, Wallingford (U.K.).
- LONG, J. (1997): *Regression Models for Categorical Dependent Variables*. Sage Publications, Thousand Oaks, CA.
- LONG, J. and S. CHENG (forthcoming): Regression Models for Categorical Outcomes. In: HARDY, M. and A. BRYMAN (Eds.): *Handbook of Data Analysis*, Sage Publications.
- LONG, J. and J. FREESE (2001): *Regression Models for Categorical Dependent Variables Using Stata*. Stata Press, College Station.
- MADDALA, G. (1983): *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press, New York.
- MARTIN, D. (2000): *Geographic Information Systems*. Routledge, London.
- MCFADDEN, D. (1973): Conditional Logit Analysis of Qualitative Choice Behaviour. In: ZAREMBKA, P. (Ed.): *Frontiers in Econometrics*, pp. 105–142. Academic Press, New York.
- MCFADDEN, D., W. TYE and K. TRAIN (1976): An application of diagnostic tests for the irrelevant alternatives property of the multinomial logit model. In: *Transportation Research Record*, Vol. 637, pp. 39–46.
- MCGARIGAL, K. and B. J. MARKS (1995): *FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structures*. Forest Science Department, Oregon State University, Corvallis, Oregon.
- MERTENS, B. and E. F. LAMBIN (2000): Land-cover-change trajectories in southern Cameroon. In: *Annals of the Association of American Geographers*, Vol. 90 (3), pp. 467–494.

- MERTENS, B., W. D. SUNDERLIN, O. UNDOYE and E. F. LAMBIN (2000): Impact of Macroeconomic Change on Deforestation in South Cameroon: Integration of Household Survey and Remotely-Sensed Data. In: *World Development*, Vol. 28, pp. 983–999.
- MITAS, L. and H. MITASOVA (1999): Spatial interpolation. In: LONGLEY, P.A., M. F. GOODCHILD, D. J. MAGUIRE and D. W. RHIND (Eds.): *Geographical Informations Systems*, Vol. 1, pp. 481–492. John Wiley & Sons, New York, 2 Edn.
- MÜLLER, D. and M. ZELLER (2002): Land Use Dynamics in the Central Highlands of Vietnam: A spatial model combining village survey data and satellite imagery interpretation. In: *Agricultural Economics*, Vol. 27 (3), pp. 333–354.
- MÜLLER, D. and M. ZELLER (forthcoming): Agricultural intensification, population growth and forest cover change: Evidence from spatially explicit land use modelling in the Central Highlands of Vietnam. Paper presented at the International Symposium on “Land use, nature conservation, and stability of rainforest margins in Southeast Asia”. Georg-August University of Göttingen, Göttingen.
- MUNROE, D., J. SOUTHWORTH and C. M. TUCKER (2001): *The Dynamics of Land-Cover Change in Western Honduras: Spatial Autocorrelation and Temporal Variation*. www.cipec.org/index.html, accessed: Jan. 2002.
- MUNROE, D., J. SOUTHWORTH and C. M. TUCKER (2002): The Dynamics of Land-Cover Change in Western Honduras: Exploring Spatial and Temporal Complexity. In: *Agricultural Economics*, Vol. 27 (3), pp. 355–369.
- NATIONAL INSTITUTE FOR AGRICULTURAL PLANNING AND PROJECTION (2000): *Digital Soil Map*. Hanoi, Vietnam.
- NELSON, G.C. and J. GEOGHEGAN (2002): Deforestation and land use change: sparse data environments. In: *Agricultural Economics*, Vol. 27 (3), pp. 201–216.
- NELSON, G.C., V. HARRIS and S. STONE (1999): Spatial Econometric Analysis and Project Evaluation: Modeling Land Use Change in the Darien.
- NELSON, G.C., V. HARRIS and S. STONE (2001a): Deforestation, Land Use, and Property Rights: Empirical Evidence from Darien, Panama. In: *Land Economics*, Vol. 77 (2), pp. 187–205.

- NELSON, G.C., V. HARRIS, S. STONE and A. D. PINTO (2001b): *Land Use and Road Improvements: A Spatial Econometric Analysis*. www.ace.uiuc.edu/faculty/gnelson/, accessed: Nov. 2001.
- NELSON, G.C. and D. HELLERSTEIN (1997): Do Roads Cause Deforestation? Using Satellite Images in Econometric Analysis of Land Use. In: *American Journal of Agricultural Economics*, Vol. 79, pp. 80–88.
- NELSON, G.C., S. W. STONE and A. D. PINTO (forthcoming): Land Use and Road Improvements: A Spatial Econometric Analysis. In: *International Regional Science Review*.
- NGUYEN MANH CUONG (1999): Information Technologies for Forest Management in Vietnam. In: SUSTAINABLE MANAGEMENT OF RESOURCES IN THE LOWER MEKONG BASIN PROJECT (SMRP) (Ed.): *Application of Resource Information Technologies (GIS/GPS/RS) in Forest Land & Resources Management*, pp. 79–84. German Agency for Technical Cooperation (GTZ), Hanoi.
- PATTERSON, M. (1998): Glossary. In: LIVERMAN, D., E. F. MORAN, R. R. RINDFUSS and P. C. STERN (Eds.): *People and Pixels: Linking Remote Sensing and Social Science*, pp. 229–236. National Academy Press, Washington, D.C.
- PENDER, J.L., P. JAGGER, E. NKONYA and D. SSERUNKUUMA (2001a): Development Pathways and Land Management in Uganda: Causes and Implications. Paper presented at a Workshop on Policies for Improved Land Management in Uganda. Environment and Production Technology Division, International Food Policy Research Institute.
- PENDER, J.L., S. J. SCHERR and G. DURÓN (1999): *Pathways of Development in the Hillsides of Honduras: Causes and Implications for Agricultural Production, Poverty and Sustainable Resource Use*. Discussion Paper No. 45. Environment and Production Technology Division, International Food Policy Research Institute, Washington, D.C.
- PENDER, J.L., S. J. SCHERR and G. DURÓN (2001b): Pathways of development in the hillsides of Honduras: Causes and implications for agricultural production, poverty, and sustainable resource use. In: LEE, D.R. and C. B. BARRETT (Eds.):

- Tradeoffs or Synergies? Agricultural Intensification, Economic Development and the Environment*. CAB International, Wallingford (U.K.).
- PEOPLE'S COMMITTEE OF DAK LAK PROVINCE (1996): *Project for Hunger Eradication and Poverty Alleviation, Period 1996 - 2000*. People's Committee of Dak Lak Province, Buon Ma Thuot, Vietnam.
- PFAFF, A.S. (1997): *What Drives Deforestation in the Brazilian Amazon? Evidence from Satellite and Socioeconomic Data*. Policy Research Group Working Paper. The World Bank; Policy Research Department; Environment, Infrastructure, and Agriculture Division, Washington, D.C.
- PICHON, F.J. (1997): Settler Households and Land-Use Patterns in the Amazon Frontier: Farm-Level Evidence from Ecuador. In: *World Development*, Vol. 25, pp. 67–91.
- PITT, M.M., M. R. ROSENZWEIG and D. M. GIBBONS (1993): The Determinants and Consequences of the Placement of Government Programs in Indonesia. In: *The World Bank Economic Review*, Vol. 7 (3), p. 1993.
- POATE, C. and P. DAPLYN (1993): *Data for Agrarian Development*. Cambridge University Press, Cambridge.
- REARDON, T., C. B. BARRET, V. KELLY and K. SAVADOGO (2001): Sustainable versus Unsustainable Agricultural Intensification in Africa: Focus on Policy Reforms and Market Conditions. In: LEE, D.R. and C. B. BARRET (Eds.): *Tradeoffs or Synergies? Agricultural Intensification, Economic Development and the Environment*, pp. 365–382. CAB International, Wallingford (U.K.).
- RICARDO, D. (2002): *The Principles of Political Economy and Taxation*. Janus Publishing, London. Date of original publication: 1817.
- RINDFUSS, R.R., B. ENTWISLE, S. J. WALSH, P. PRASARTKUL, Y. SAWANGDEE, T. W. CRAWFORD and J. READE (2002): Continuous and Discrete: Where they have met in Nang Rong, Thailand. In: WALSH, S.J. and K. A. CREWS-MEYER (Eds.): *Linking People, Place, and Policy: A GIScience Approach*. Kluwer Academic Publishers, Boston.

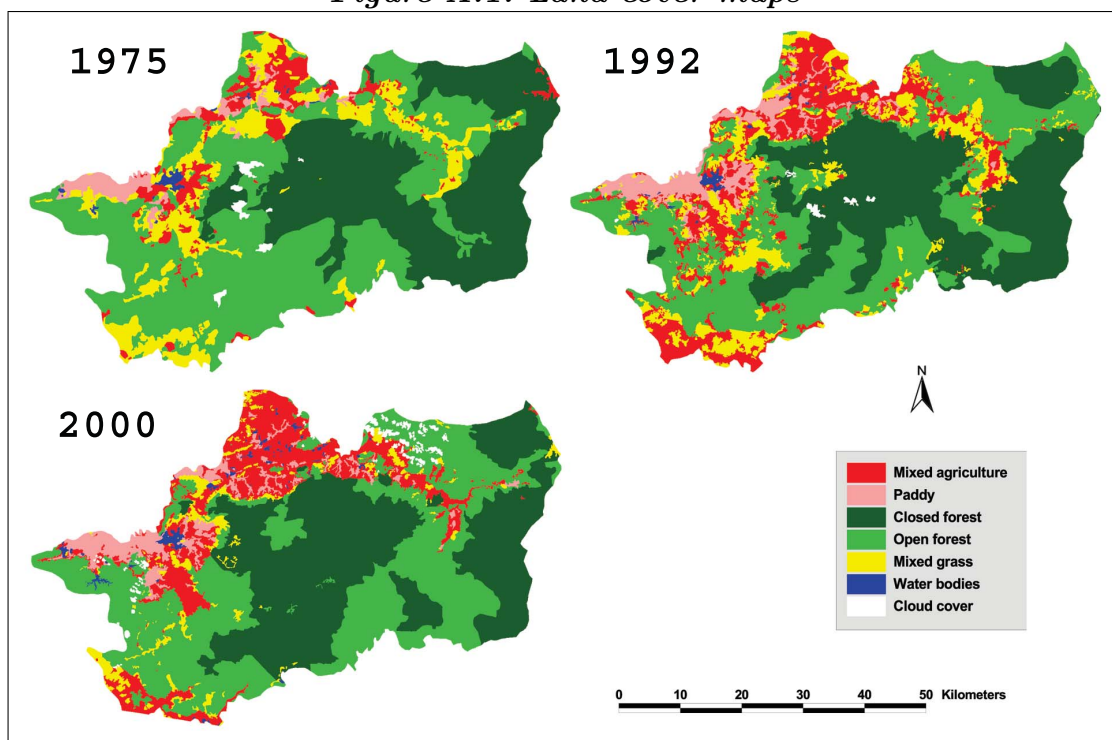
- RUTHENBERG, H. (1980): *Farming Systems in the Tropics*. Clarendon Press, Oxford, 3 Edn.
- RUTTAN, V.W. and Y. HAYAMI (1984): Towards a Theory of Induced Institutional Innovation. In: *Journal of Development Studies*, Vol. 20 (4), pp. 203–223.
- SAMUELSON, P.A. (1983): Thünen at Two Hundred. In: *Journal of Economic Literature*, Vol. 21 (4), pp. 1468–1488.
- SCHERR, S.J., G. BERGERON, J. L. PENDER and B. BARBIER (1996): *Policies for Sustainable Development in Fragile Lands: Methodology Overview*. Environment and Production Technology Division, International Food Policy Research Institute, Washington, D.C.
- SCHERR, S.J. and P. HAZELL (1994): *Sustainable Agricultural Development Strategies in Fragile Lands*. Discussion Paper No. 1. Environment and Production Technology Division, International Food Policy Research Institute, Washington, D.C.
- SERNEELS, S. and E. F. LAMBIN (2001): Proximate causes of land-use change in Narok District, Kenya: a spatial statistical model. In: *Agriculture, Ecosystems & Environment*, Vol. 85 (1-3), pp. 65–81.
- SHIVELY, G.E. and E. MARTINEZ (2001): Deforestation, Irrigation, Employment and Cautious Optimism in Southern Palawan, the Philippines. In: ANGELSEN, A. and D. KAIMOWITZ (Eds.): *Agricultural Technologies and Tropical Deforestation*, pp. 335–346. CAB International, Oxon.
- SIKOR, T. (1999): *The Political Economy of Decollectivization: A study of differentiation in and among Black Thai villages of northern Vietnam*. Ph.D. thesis, University of California, Berkeley.
- SMALL, K. and C. HSIAO (1985): Multinomial logit specifications tests. In: *International Economic Review*, Vol. 26, pp. 619–627.
- SMITH, J.P. and D. THOMAS (2003): Remembrances of things past: test-retest reliability of retrospective migration histories. In: *Journal of the Royal Statistical Society. Series A*, Vol. 166 (1), pp. 23–50.

- SOUTHGATE, D., R. SIERRA and L. BROWN (1991): The Causes of Tropical Deforestation in Ecuador: A Statistical Analysis. In: *World Development*, Vol. 19 (9), pp. 1145–1151.
- STATA CORP (2003): *Stata Statistical Software: Release 8.0*. Stata Corporation, College Station, Texas.
- TACHIBANA, T., T. M. NGUYEN and K. OTSUKA (2001): From Deforestation to Reforestation through Tenure Reforms: The Case of the Northern Hill Region of Vietnam. In: *Journal of Environmental Economics and Management*, Vol. 41, pp. 44–69.
- TEMPLETON, S.R. and S. J. SCHERR (1997): *Population Pressure and the Microeconomy of Land Management in Hills and Mountains of Developing Countries*. Discussion Paper No. 26. Environment and Production Technology Division, International Food Policy Research Institute, Washington, D.C.
- THE SOCIALIST REPUBLIC OF VIETNAM (1992): *Constitution of the Socialist Republic of Vietnam*. www.vietnamembassy-usa.org/learn/gov-constitution.php3, accessed Feb. 2003.
- THE SOCIALIST REPUBLIC OF VIETNAM (2002): *The Comprehensive Poverty Reduction And Growth Strategy (CPRGS)*. Socialist Republic of Vietnam (Approved by the Prime Minister), Hanoi.
- TIFFEN, M., M. MORTIMORE and R. KATES (1994): *More people, less erosion: Environmental recovery in Kenya*. John Wiley & Sons, Chichester.
- TOBLER, W. (1979): Cellular Geography. In: GALE, S. and G. OLSSON (Eds.): *Philosophy in Geography*, pp. 379–389. Dordrecht, Reidel.
- TURNER II, B. and W. MEYER (1994): Global Land-Use and Land-Cover Change: An Overview. In: MEYER, W. and B. TURNER (Eds.): *Changes in Land Use and Land Cover: A Global Perspective*, pp. 3–10. Cambridge University Press, Cambridge.
- TURNER II, B., D. SKOLE, S. SANDERSON, G. FISCHER, L. FRESCO and R. LEEMANS (1995): *Land-Use and Land-Cover Change Science/Research Plan*.

- IGBP Report No. 35 and HDP Report No. 7. The International Geosphere-Biosphere Programme: A Study of Global Change (IGBP) of the International Council of Scientific Unions (ICSU) and The Human Dimensions of Global Environmental Change Programme (HDP) of the International Social Science Council (ISSC), Stockholm and Geneva.
- VAN DER VEEN, A. and H. OTTER (2001): Land use changes in regional economic theory. In: *Environmental Modeling & Assessment*, Vol. 6 (2), pp. 145–150.
- VANCE, C. and J. GEOGHEGAN (2002): Temporal and spatial modelling of tropical deforestation: A survival analysis linking satellite and household survey data. In: *Agricultural Economics*, Vol. 27 (3), pp. 317–332.
- VON THÜNEN, J.H. (1960): *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie – Vol II: Der naturgemässe Arbeitslohn und dessen Verhältnis zum Zinsfuß und zur Landrente*. Loyola University Press, Chicago. Date of original publication: 1850.
- VON THÜNEN, J.H. (1990): *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*. Scientia Verlag, Aalen. Date of original publication: 1826.
- VUONG XUAN TINH AND BUI MINH DAO (2000): Traditional Land Ownership and Land Use of Ethnic Minorities in Vietnam. Tech. Rep., Vietnam - Sweden Cooperation Program on Land Administration Reform.
- WISE, S. (2002): *GIS Basics*. Taylor & Francis, New York.
- WOOLDRIDGE, J.M. (2001): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge.
- WORLD BANK (2002a): *Vietnam at a Glance*. www.worldbank.org/data/countrydata/aag/vnm-aag.pdf, accessed Jan. 2003.
- WORLD BANK (2002b): *Vietnam Development Report*. www.worldbank.org.vn, Vietnam Consultative Group Meeting, Hanoi, accessed Jan. 2003.
- WORLD BANK (2002c): *World Development Indicators database*. www.worldbank.org/data/, accessed Feb. 2003.

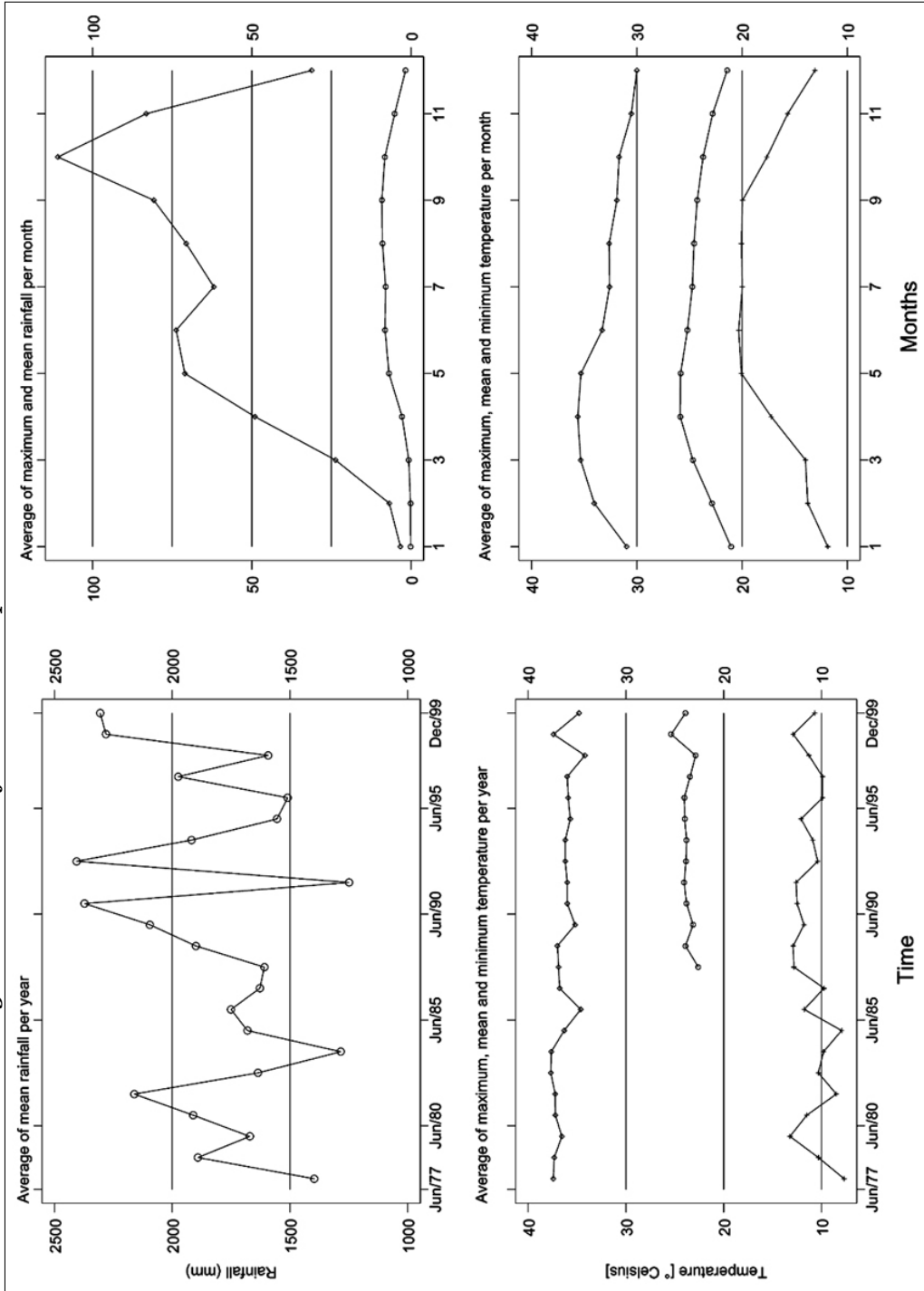
ZELLER, M., C. LAPENU, B. MINTEN, E. RALISON, D. RANDRIANAIVO and C. RANDRIANARISAO (2000): The critical triangle between environmental sustainability, economic growth, and poverty alleviation. In: MINTEN, B. and M. ZELLER (Eds.): *Beyond market liberalization: Income generation, welfare and environmental sustainability in Madagascar*, pp. 209–236. Ashgate Publishing Company, Aldershot, England.

Figure A.1: Land cover maps



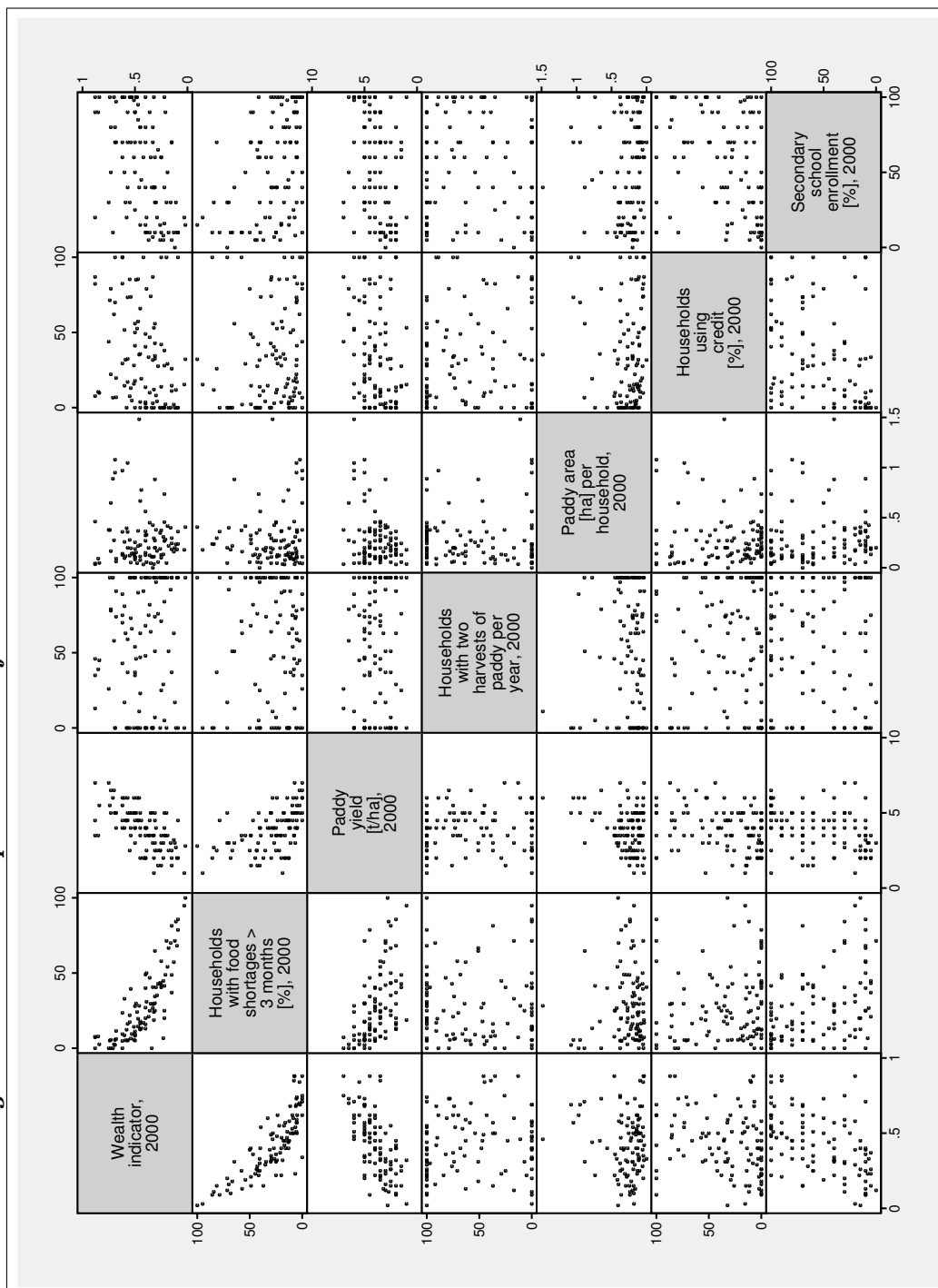
Source: Interpretation of Landsat satellite images by Nguyen Thi Thanh Huong and author.

Figure A.2: Rainfall and temperature in research area



Source: Data from DARD, Dak Lak, and Danish Agency for Development Assistance (DANIDA)

Figure A.3: Scatterplot matrix of selected socioeconomic indicators



Source: Village survey in Lak and Krong Bong districts, 2000/2001

Table A.1: Descriptive statistics for land use

Variable	Mean	Std.Dev.	N
Mixed agriculture, 2000	0.14	0.35	37,405
Paddy, 2000	0.05	0.21	37,405
Closed forest, 2000	0.38	0.48	37,405
Open forest, 2000	0.39	0.49	37,405
Grass, 2000	0.04	0.2	37,405
Mixed agriculture, 1992	0.14	0.35	37,676
Paddy, 1992	0.05	0.22	37,676
Closed forest, 1992	0.35	0.48	37,676
Open forest, 1992	0.34	0.47	37,676
Grass, 1992	0.11	0.32	37,676
Mixed agriculture, 1975	0.06	0.24	37,755
Paddy, 1975	0.03	0.16	37,755
Closed forest, 1975	0.35	0.48	37,755
Open forest, 1975	0.43	0.5	37,755
Grass, 1975	0.12	0.33	37,755

Source: Own calculations

Table A.2: Descriptive statistics for independent variables

Variable	Mean	Std.Dev.	Min.	Max.	N
Slope (degrees)	13.24	9.5	0	50	37,731
Elevation (100m)	8.14	3.65	4.19	23.88	37,731
Soil suitability for agriculture	6.84	1.31	1	9	37,755
Soil suitability for paddy	6.57	1.34	1	9	37,755
Rainfall sum (100mm), 1992-99	20.6	0.99	18.91	23.65	37,731
Rainfall variance ("), 1992-99	29.96	6.82	16.73	42.05	37,731
Rainfall sum ("), 1976-91	15.8	2.57	11.59	20.08	37,731
Rainfall variance ("), 1976-91	20.54	2.4	14.2	25.3	37,731
Convertible land (0/1), 1992	0.01	0.11	0	1	37,652
Convertible land (0/1), 1975	0.05	0.23	0	1	37,731
Age of village (years), 2000	28.82	17.44	1	55	37,651
Ethnic minority village (0/1)	0.69	0.46	0	1	37,651
Population, 1990	52.7	31.53	0	292.89	37,737
Population, 1980	29.24	21.83	0	338.9	37,737
Hero village (0/1)	0.32	0.47	0	1	37,651
Km to all-year road, 1998	5.87	4.94	0	20.39	37,755
Km to all-year road, 1990	17.81	11.04	0	43.25	37,755
Years since primary school, 1990	11.01	8.98	0	55	37,651
Years since reservoir, 2000	8.95	9.79	0	68	37,651
Increase in irrigated area (ha), 2000	8	11.15	0	62.5	37,651
Increase in irrigated area (ha), 1990	11.56	20.9	0	62.5	37,755
Years since NPK, 2000	3.17	3.23	0	20	37,651
Years since protected area	1	2.51	0	9	37,755
Coordinate in x-direction	867.89	385.35	4	1,619	37,755
Coordinate in y-direction	512.04	228.89	4	1,039	37,755

Source: Own calculations

**Table A.3: Multinomial logit estimation of land use, model 2000
(closed forest as comparison group)**

Variable	Mixed agriculture	Paddy	Open forest	Grass
Slope (degrees)	-0.1 (12.46)**	-0.176 (7.57)**	0.006 (2.01)*	-0.025 (3.95)**
Spatial lag slope (degrees)	-0.05 (11.39)**	-0.066 (7.18)**	0.011 (5.66)**	-0.013 (3.38)**
Elevation (100m)	-3.405 (28.65)**	-7.75 (18.51)**	-0.866 (65.42)**	-1.393 (29.23)**
Soil suitability for agriculture	0.031 -0.42	0.092 -1.2	0.33 (4.91)**	0.153 -1.93
Soil suitability for paddy	-0.06 -0.62	-0.431 (4.30)**	-0.087 -0.93	0 0
Rainfall sum (100mm), 1992-99	0.818 (7.12)**	2.099 (9.92)**	-0.028 -0.41	1.054 (10.00)**
Rainfall variance (°), 1992-99	-0.322 (6.43)**	-0.42 (5.62)**	0.198 (8.42)**	0.255 (5.69)**
Age of village (years), 2000	0.01 (3.10)**	0.007 -1.71	0.022 (10.68)**	0.005 -1.59
Ethnic minority village (0/1)	-0.032 -0.26	-0.031 -0.2	0.331 (3.81)**	0.959 (7.76)**
Population, 1990	0.017 (7.29)**	0.022 (7.60)**	0.007 (4.98)**	0.006 (2.52)*
Hero village (0/1)	0.356 (2.84)**	-1.23 (6.96)**	-0.657 (9.11)**	-0.412 (3.33)**
Km to all-year road, 1998	-0.433 (18.82)**	-0.255 (5.97)**	-0.076 (10.17)**	-0.092 (5.43)**
Years since primary school, 1990	0.034 (6.59)**	0.03 (4.68)**	0.025 (6.83)**	0.016 (3.04)**
Years since NPK, 2000	-0.117 (8.36)**	-0.051 (3.02)**	-0.084 (8.24)**	-0.092 (6.13)**
Years since reservoir, 2000	0.013 (2.05)*	0.024 (3.23)**	0.038 (7.45)**	-0.001 -0.23
Increase in irrigated area (ha), 2000	-0.027 (7.06)**	-0.014 (3.05)**	-0.023 (8.57)**	-0.012 (2.98)**
Years since protected area	-0.546 (6.58)**	-0.6 (6.82)**	-0.188 (2.63)**	-0.463 (4.63)**
Interaction protection - slope	-0.033 -0.94	-0.117 -1.11	0.004 -0.55	-0.133 (4.19)**
Interaction protection - maize soil	0.303 (3.44)**	0.279 (2.57)*	0.244 (3.61)**	0.251 (2.29)*
Coordinate in x-direction	0.002 (2.44)*	0 -0.09	-0.004 (8.67)**	-0.01 (12.85)**
Coordinate in y-direction	0 -0.31	0.001 (2.19)*	0.006 (29.78)**	0.001 -1.36
Constant	11.486 (5.18)**	8.256 (2.16)*	-0.548 -0.39	-13.189 (6.25)**
Observations	37,351	37,351	37,351	37,351

Note: Robust z statistics in parentheses; * significant at 5%; ** significant at 1%; bold coefficient indicate significance at the 1% level.

Source: Own calculations

*Table A.4: Multinomial logit estimation of land use, model 1990
(closed forest as comparison group)*

Variable	Mixed agriculture	Paddy	Open forest	Grass
Slope (degrees)	-0.036 (6.78)**	-0.173 (7.18)**	0.011 (3.58)**	0.015 (3.54)**
Spatial lag slope (degrees)	-0.04 (12.75)**	-0.029 (3.49)**	0.005 (2.76)**	-0.01 (3.84)**
Elevation (100m)	-1.309 (47.62)**	-7.48 (13.68)**	-0.674 (62.42)**	-1.227 (59.94)**
Soil suitability for agriculture	0.723 (14.82)**	0.888 (16.45)**	0.995 (19.07)**	0.847 (16.63)**
Soil suitability for paddy	-4.043 (8.39)**	-4.411 (9.14)**	-3.668 (7.61)**	-3.976 (8.25)**
Rainfall sum (100mm), 1976-91	1.708 (18.13)**	1.633 (9.52)**	0.863 (11.36)**	1.663 (18.31)**
Rainfall variance ("), 1976-91	-0.817 (15.07)**	-0.735 (6.40)**	-0.008 -0.17	-0.641 (12.07)**
Age of village (years), 1990	0.013 (3.39)**	0.014 (2.96)**	0.011 (3.21)**	0.01 (2.64)**
Ethnic minority village (0/1)	0.979 (10.35)**	0.839 (6.37)**	0.603 (8.30)**	1.152 (12.94)**
Population, 1980	-0.006 (2.92)**	-0.006 (2.37)*	-0.009 (4.83)**	-0.006 (2.76)**
Hero village (0/1)	-0.537 (5.78)**	-1.526 (9.56)**	-0.447 (6.81)**	-0.625 (7.26)**
Km to all-year road, 1990	-0.208 (23.00)**	-0.271 (14.21)**	-0.019 (2.67)**	-0.153 (18.62)**
Years since primary school 1990	0.014 (2.27)*	0.041 (5.08)**	0.024 (4.68)**	0.01 -1.69
Years since NPK, 1990	-0.041 (3.93)**	-0.048 (2.71)**	-0.087 (12.83)**	-0.012 -1.36
Years since reservoir, 1990	-0.109 (11.43)**	-0.098 (8.22)**	-0.075 (10.32)**	-0.095 (10.64)**
Increase in irrigated area (ha), 1990	0.016 (7.01)**	0.013 (2.59)**	-0.008 (5.56)**	0.011 (5.18)**
Coordinate in x-direction	0.008 (14.77)**	0.004 (4.30)**	0.001 -1.67	0.007 (13.48)**
Coordinate in y-direction	0.003 (8.78)**	0.002 -1.55	-0.002 (7.52)**	0.001 (3.39)**
Constant	17.717 (5.26)**	47.799 (11.50)**	11.845 (3.55)**	13.277 (3.96)**
Observations	37,516	37,516	37,516	37,516

Note: Robust z statistics in parentheses; * significant at 5%; ** significant at 1%; bold coefficient indicate significance at the 1% level.

Source: Own calculations

Table A.5: Measures of fit for multinomial logit models

Statistic	Model 2000	Model 1990
Log-Likelihood Intercept Only	-48,108.8	-52,877.9
Log-Likelihood Full Model	-19,642.9	-27,625.2
Likelihood Ratio (df)	56,931.8 (84)	50,505.3 (74)
Probability > LR	0.000	0.000
McFadden's R^2	0.592	0.478
McFadden's Adjusted R^2	0.590	0.476
Maximum Likelihood R^2	0.782	0.740
Cragg & Uhler's R^2	0.847	0.787
Count R^2	0.791	0.700
Adj. Count R^2	0.657	0.539

Source: Own calculations

Table A.6: Likelihood ratio tests for combining outcome categories

Categories tested	Model 2000		Model 1990	
	χ^2	$P > \chi^2$	χ^2	$P > \chi^2$
Mixed agriculture – paddy	2,794.7	0	3,837.3	0
Mixed agriculture – open forest	18,456.2	0	12,434.8	0
Mixed agriculture – grass	3,271.8	0	2,292.6	0
Mixed agriculture – closed forest	26,784.0	0	24,455.7	0
Paddy – open forest	11,514.4	0	11,692.1	0
Paddy – grass	3,955.4	0	6,016.6	0
Paddy – closed forest	13,354.0	0	15,745.5	0
Open forest – grass	3,793.9	0	4,616.9	0
Open forest- closed forest	20,762.7	0	17,589.0	0
Grass – closed forest	9,436.7	0	17,512.5	0

Source: Own calculations

Table A.7: Likelihood ratio tests for independent variables

Variable	Model 2000		Model 1990	
	χ^2	$P > \chi^2$	χ^2	$P > \chi^2$
Slope (degrees)	235.4	0	185.2	0
Spatial lag slope (degrees)	281.7	0	294.5	0
Elevation (100m)	8,843.3	0	11,225.4	0
Soil suitability for agriculture	69.0	0	394.7	0
Soil suitability for paddy	243.9	0	682.3	0
Rainfall sum (100mm), 1992-99	243.2	0	429.8	0
Rainfall variance (100mm), 1992-99	249.6	0	726.4	0
Age of village (years), 2000	158.1	0	13.3	0.01
Ethnic minority village (0/1)	113.0	0	176.8	0
Population, 1990	75.3	0	25.5	0
Hero village (0/1)	339.5	0	117.8	0
Km to all-year road, 1998	473.9	0	1,422.8	0
Years since primary school, 1990	64.7	0	51.8	0
Years since NPK, 2000	121.1	0	231.9	0
Years since reservoir, 2000	162.6	0	145.9	0
Increase in irrigated area (ha), 2000	103.5	0	213.4	0
Years since protected area	47.7	0		
Interaction protection - slope	25.4	0		
Interaction protection - maize soil	12.6	0.013		
Coordinate in x-direction	317.6	0	728.8	0
Coordinate in y-direction	1,169.3	0	272.6	0

Source: Own calculations

Table A.8: Descriptive statistics for variables in simulation

Variable	Mean	Std.Dev.	Min.	Max.	N
Mixed agriculture, 2000	0.24	0.43	0	1	22,291
Paddy, 2000	0.08	0.27	0	1	22,291
Non-agricultural land, 2000	0.68	0.47	0	1	22,291
Slope (degrees)	10.18	9.04	0	50	22,291
Spatial lag slope (degrees)	21.06	16.75	0	77.9	22,291
Elevation (100m)	5.88	1.67	4.2	13	22,291
Soil suitability for paddy	0.14	0.35	0	1	22,291
Soil suitability for agriculture	0.04	0.19	0	1	22,291
Rainfall amount (100mm), 1992-99	20.3	0.86	18.9	23.4	22,280
Rainfall variance ("), 1992-99	27.62	7.07	16.7	41.5	22,280
Population (households), 1990	65.08	26.66	1.8	292.9	22,272
Ethnic minority village (0/1)	0.62	0.49	0	1	22,291
Km to all-year road, 1998	5.85	5.36	0	23	22,291
Years since establishment of NEZ	9.26	8.54	0	24	22,291
Years since introduction of NPK	4.26	3.7	0	20	22,291
Irrigated area * age of reservoir	123.34	208.73	0	1,020	22,291
Protection dummy Nam Ca (0/1)	0.08	0.27	0	1	22,291
Landscape fragmentation, 1992	84.45	28	1.7	110.3	22,279

Source: Own calculations

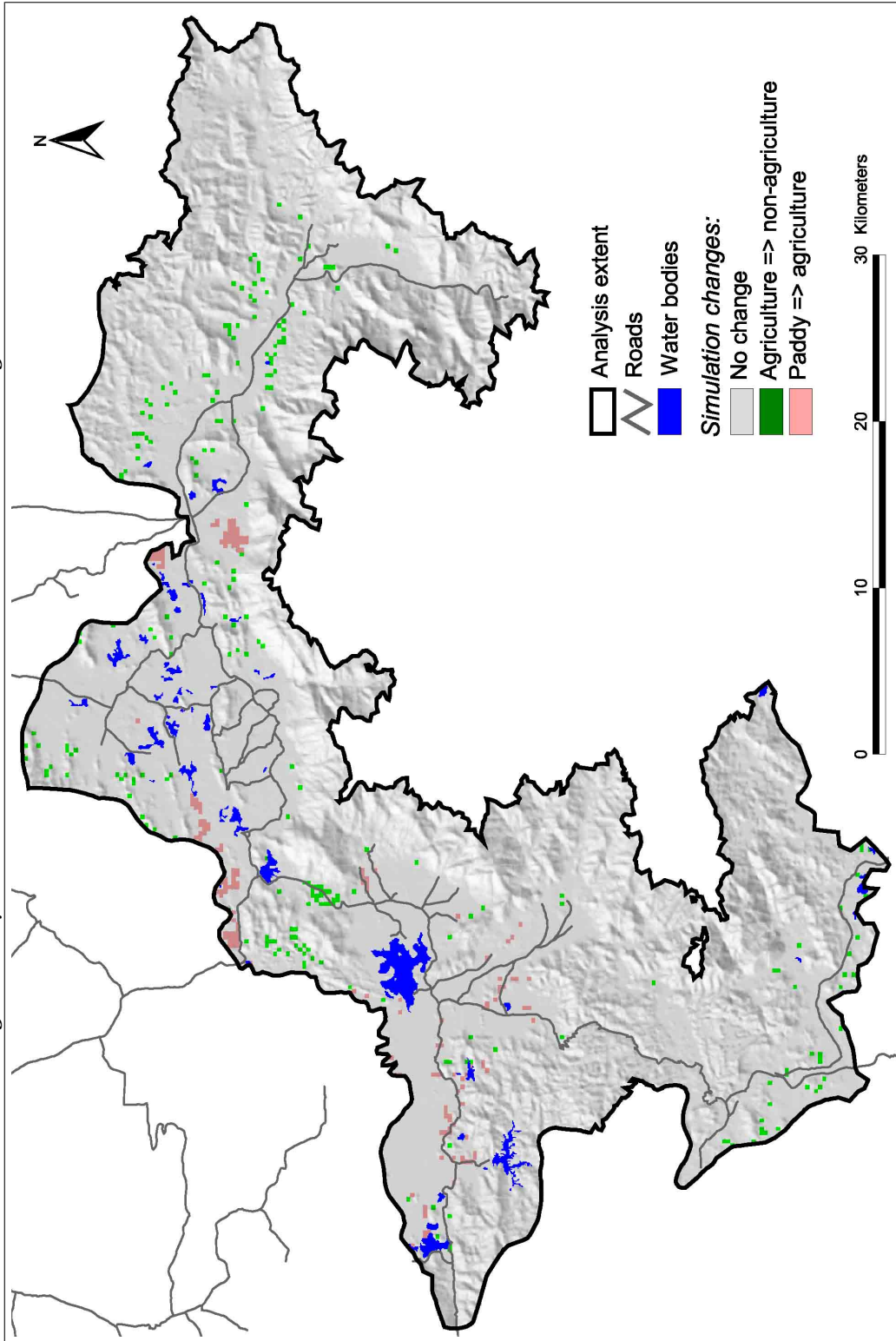
**Table A.9: Multinomial logit estimation for base prediction
(non-agricultural land as comparison group)**

Variable	Mixed agriculture	Paddy
Slope (degrees)	-0.088 (12.39)**	-0.207 (9.00)**
Spatial lag slope (degrees)	-0.046 (12.46)**	-0.072 (7.92)**
Elevation (100m)	-2.621 (24.30)**	-6.695 (19.06)**
Soil suitability for paddy (0/1)	1.306 (12.98)**	2.543 (20.38)**
Soil suitability for agriculture (0/1)	1.5 (9.59)**	1.047 (4.99)**
Rainfall amount (100mm), 1992-1999	0.504 (7.94)**	1.293 (7.93)**
Rainfall variance (100mm), 1992-1999	-0.116 (14.32)**	-0.265 (16.05)**
Population (households), 1990	0.007 (3.87)**	0.013 (5.15)**
Ethnic minority village (0/1)	-0.123 -1.78	-0.433 (3.95)**
Km to all-year road, 1998	-0.001 -0.1	-0.004 -0.17
Years since establishment of New Economic Zone	0.041 (7.30)**	-0.022 (2.79)**
Years since introduction of NPK, 2000	0.062 (6.78)**	0.128 (10.36)**
Irrigated area * age of reservoir, 2000	0 (2.42)*	0.001 (6.52)**
Protection dummy Nam Ca (0/1)	-2.708 (15.44)**	-2.74 (9.85)**
Landscape fragmentation, 1992	-0.017 (14.36)**	-0.02 (5.58)**
Constant	6.488 (5.36)**	11.168 (4.24)**
Observations	22,255	22,255

Note: Robust z statistics in parentheses; * significant at 5%; ** significant at 1%

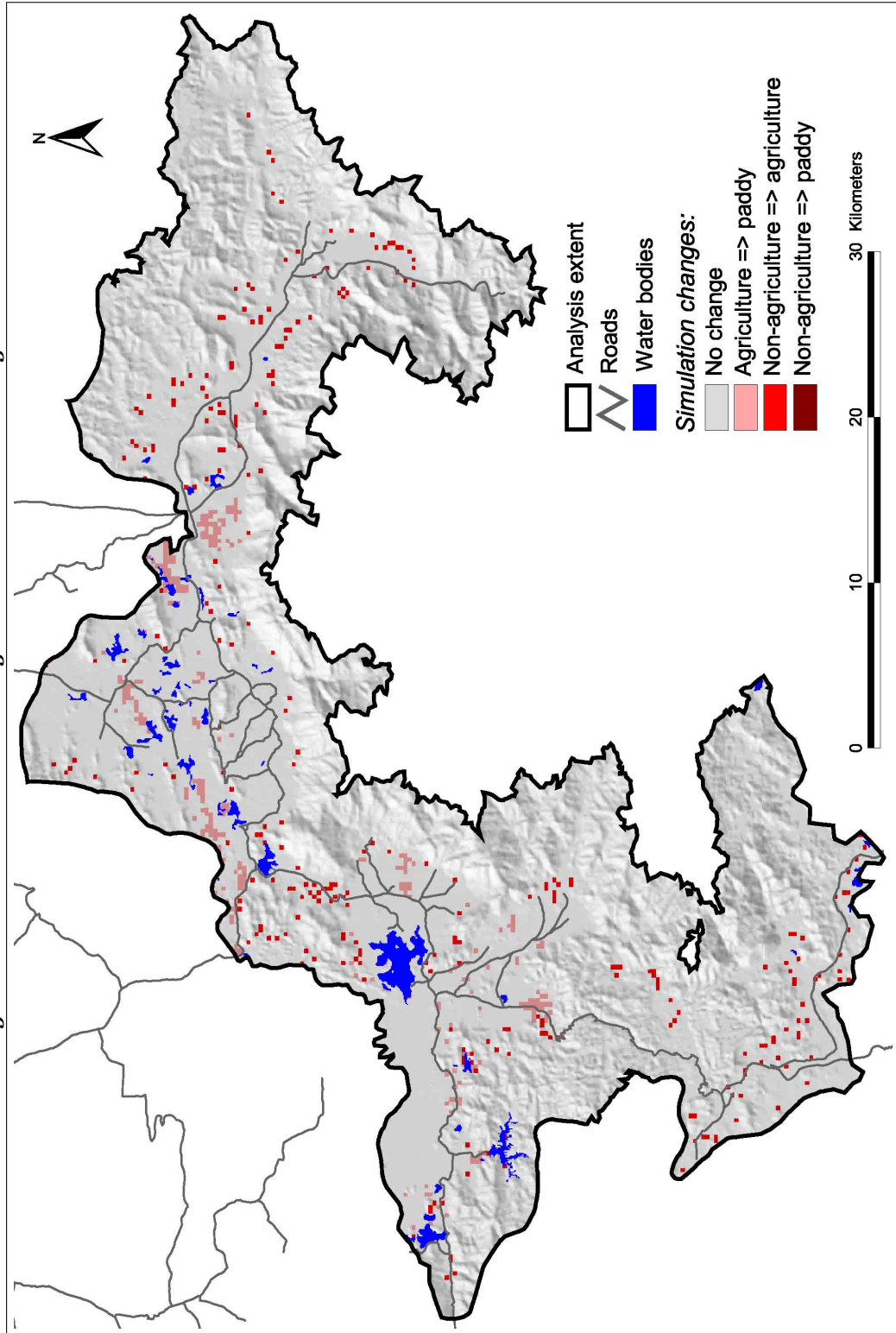
Source: Own calculations

Figure A.4: Scenario 1 – Low investments in irrigation



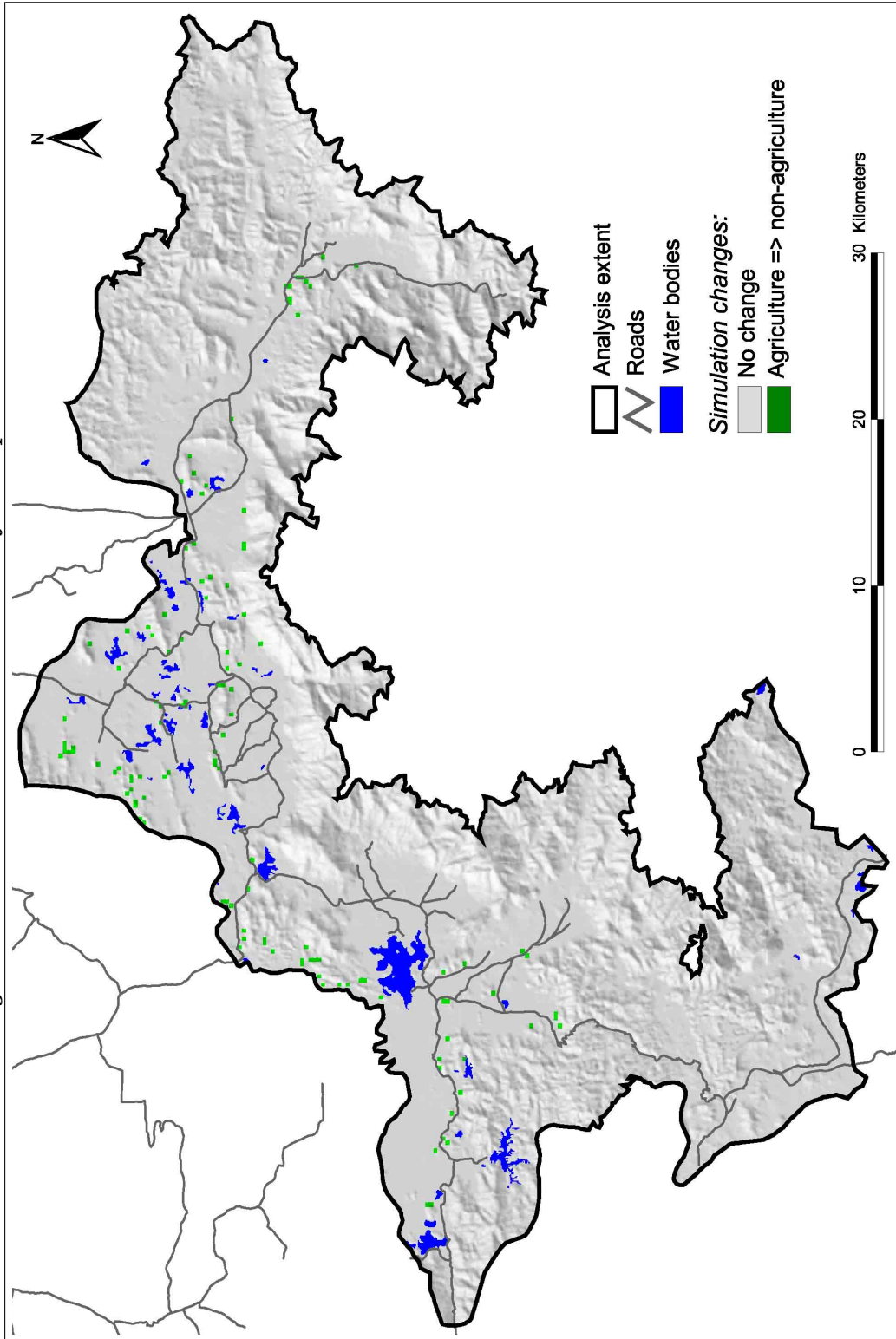
Source: Author

Figure A.5: Scenario 2 – High investments in irrigation

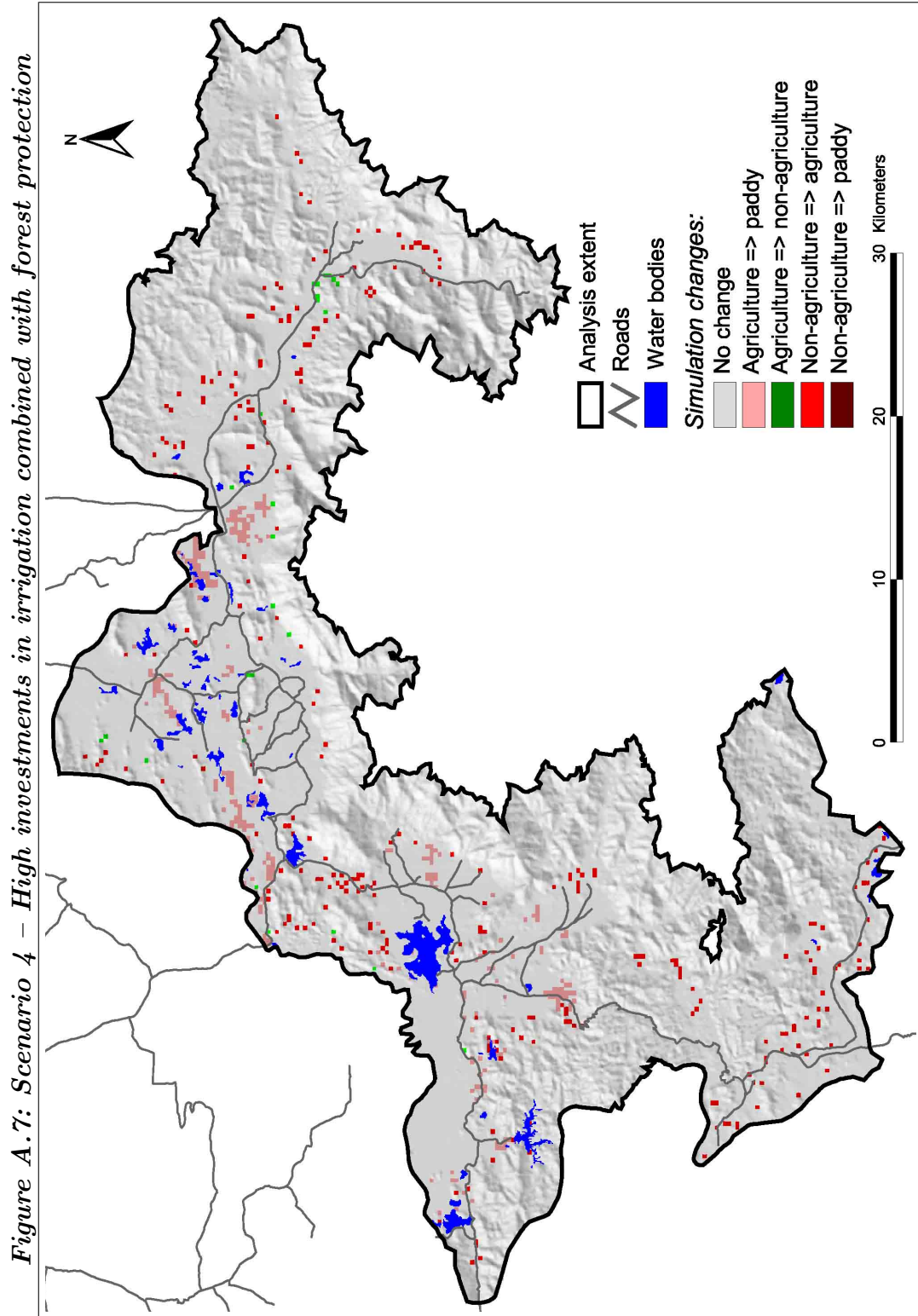


Source: Author

Figure A.6: Scenario 3 – Increased forest protection



Source: Author



Source: Author

Glossary

Absolute distance	The shortest path between two places measured on a standard unit of length (e.g. kilometers);
Data	Streams of raw facts representing events occurring in organizations or the physical environment before they have been organized and arranged into a form that people can understand and use (see also information);
Decision-support system (DSS)	Information system at the management level of an organization that combine data and analytical models to support semistructured and unstructured decision-making;
Digital elevation model (DEM)	A digital file in raster format where each georeferenced pixel has an elevation value;
Ecological fallacy	Estimation of parameters and other characteristics of a distribution at aggregate levels, but behavioral and socioeconomic relations are inferred for another, disaggregate level; leads to cross-level bias;
Electromagnetic energy (EM)	Energy with electric and magnetic components; measurement of reflected and emitted EM is the basis of remote sensing;

Endogenous	A variable in a mathematical relationship, whose value is calculated by using this relationship, i.e. it is the output of the model;
Exogenous	Independent variable, which is assumed to influence another (endogenous) variable;
Geographical information system (GIS)	A software package that accommodates the capture, analysis, manipulation, and presentation of georeferenced data; a generic tool applicable to many different types of use, termed GIS-applications;
Georeferencing	Process of referencing data to known coordinate system such as Universal Transverse Mercator (UTM);
Global positioning system (GPS)	System of 24 satellites that orbit the earth and are used to compute a location on earth; a GPS receiver tracks these satellites and calculates the position on the ground using geometry; used to collect ground control points (GCP);
Ground truthing	Field verification of image interpretation, usually by employing GPS;
Heteroskedasticity	Variance of the error term differs across observations;
Image classification	Process of assigning pixels to thematic classes; input is a image and output is a raster where each cell has a thematic code (e.g. a land-cover class); two techniques: supervised and unsupervised classification;
Information	Data that have been shaped into a form that is meaningful and useful to human beings (see also data);
Joint count statistic	Test for nominal data to detect both positive (similar values clusters together) and negative (dissimilar values appear in close association) spatial autocorrelation;

Land cover	(Bio-) physical cover of the land surface and immediate subsurface;
Land use	Describes human employment on land like paddy cultivation, hunting, or firewood collection;
Landsat	Satellite system initially operated by the US National Aeronautics and Space Administration and now by a private firm (Earth Observation Satellite Company); Landsat collects multispectral data;
Landscape	Interacting mosaic of patches relevant to the phenomenon under consideration (at any scale);
Latent variable	Unobserved variable that can be hypothesized along with a measurement model relating the latent variable to the observed, categorical outcome;
Moran's I	Specification test for spatial autocorrelation derived as a two-dimensional analog of a test for univariate time series correlation;
Multicollinearity	Relationship among independent variables, which may result in insignificant coefficients;
Multispectral scanner	Simultaneous collection of data from different wavelengths in the electromagnetic spectrum;
Patches	Basic elements or units that make up a landscape;
Pixel	Smallest unit that can be detected by a sensor; refers to the area on the ground represented by a digital number;
Polygon	A closed feature formed by a bounded series of straight lines represented by X and Y coordinates;
Raster GIS	Spatial data model with features represented by pixels; each pixel is assigned a value that corresponds to a feature;

Relative distance	Transformation of absolute distance into relative measures such as time or monetary costs; yields different measures than linear (absolute) distances; distances between objects are constant by absolute terms, but relative distances may vary with improvements of transportation;
Remote sensing	Use of electromagnetic radiation sensors to record images of the environment; aerial photographs, satellite images and GPS data are types of remotely sensed data;
Resampling	Process to create another raster with different orientation and/or cell size; different methods such as nearest neighbor selection or bilinear interpolation;
Shifting cultivation	Any spatially and temporally cyclical agricultural system that involves clearing of land followed by phases of cultivation and fallow periods;
Spatial autocorrelation	Coincidence of value similarity with locational similarity;
Spatial heterogeneity	Structural instability in the form of non-constant error variances (heteroskedasticity, e.g. groupwise heteroskedasticity) or model coefficients;
Spatial interpolation	Estimation of unknown values of a variable or parameter from known values;
Spatial lag operator	Weighted average of a random variables at neighboring locations;
Spatial resolution	Level of detail that can be detected with a sensor, often measured by the dimensions of a pixel;
Thiessen polygons	Define individual areas of influence around each of a set of points. TP are constructed in such a way that each location within a polygon is closer to its contained point than to any other point;

Universal Transverse Mercator (UTM)	Map projection of a curved reference surface onto a flat plane; UTM is a rectangular coordinate system and divides the world into 60 zones; with a separate projection for each UTM zone, a high degree of accuracy is achieved;
Vector	Spatial data with features represented as points, lines or polygons;

Source: Anselin (2000, 2002); Aronoff (1995); Bakker et al. (2000); ESRI (1999); ERDAS (1999); Fellmann et al. (2002); Haining (1990); Laudon & Laudon (1996); Long & Cheng (forthcoming); McGarigal & Marks (1995); Patterson (1998); Wise (2002)

