Linking processes and pattern of land use change

Overmars, K.P.

2

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

Abstract

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

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

2.1 Introduction

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

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

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

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

'Socialising the pixel' can be described as moving from patterns to processes. Information within spatial imagery that is relevant for the social sciences is identified and used to inform concepts and theories (Lambin *et al*., 1999; Geoghegan *et al*., 1998). Some recent LUCC studies have presented preliminary results that link the pattern from geographical approaches to the human behaviour by incorporating landscape data in social data. A number of studies aim to link household level data directly to pixels in remote sensing images (e.g. Vance and Geoghegan, 2002; Walsh *et al*., 2003) to better understand the human-environment interaction. Mertens *et al*. (2000) aggregate household level data to the village level and combine the aggregated data at that level with spatial data. Walker *et al*. (2000) and Staal *et al*. (2002) base their analyses on household level data, but add spatial data to the household data using the geographical position of the households.

The other way around, 'pixelising the social' involves moving from processes to patterns. For example, socio-economic theory is tested in a spatially explicit way (e.g. Chomitz and Gray, 1996). Other approaches, like multi-agent modelling start with social and decision-making theories and move from there to construct spatial explicit models (Parker *et al*., 2002).

The approach applied in this study explores the results of statistical models based on socio-economic theories at the household level and uses the outcomes in the construction of geographical models in order to incorporate the theories about human decision-making in these spatially explicit models. This approach aims to link the widely used geographical approaches based on statistical models (Veldkamp and Fresco, 1997; Kok and Veldkamp, 2000; Serneels and Lambin, 2001; Nelson *et al*., 2001; Schneider and Pontius, 2001) and the socio-economic approaches using household level data (Walker *et al*., 2000; Staal *et al*., 2002; Vance and Geoghegan, 2002).

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

The socio-economic approach and the geographical approach often work at different scales and at different organisational levels. This alternative approach aims to provide tools and methods to facilitate the exchange of information between the two approaches.

2.2 Study area and data collection

2.2.1 Study area

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

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

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

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

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

2.2.2 Data collection

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

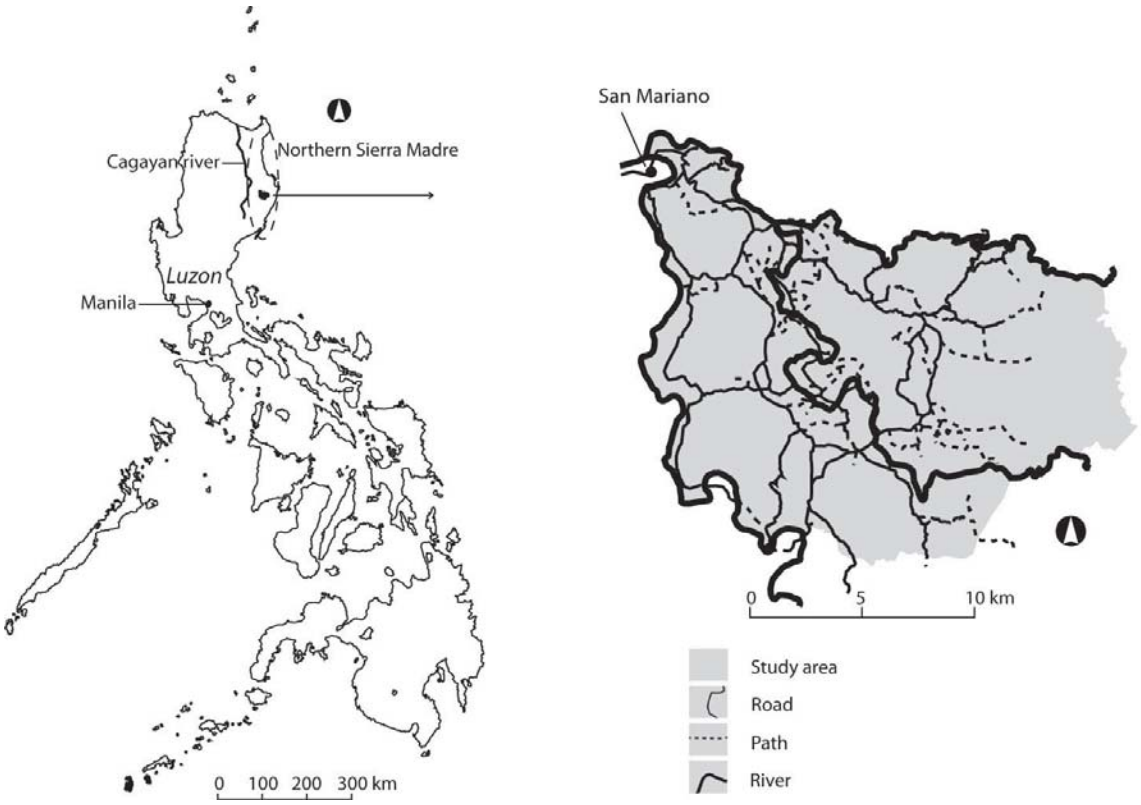

Figure 2.1: Location of the study area in the Philippines (left) and a close up of the study area (right)

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

Land use data for the two spatial approaches

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

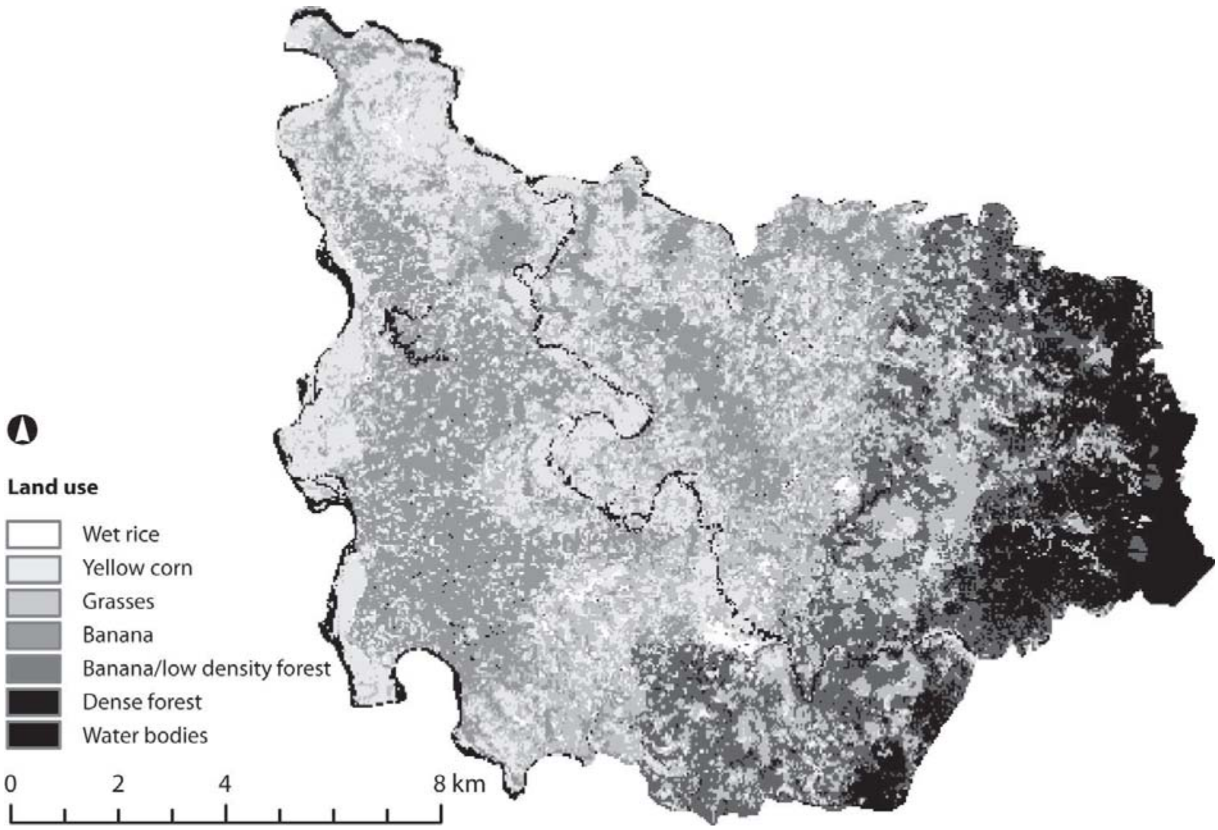

Figure 2.2: Land use map

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

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

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

Household level data

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

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

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

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

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

Table 2.1: (Continued)

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

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

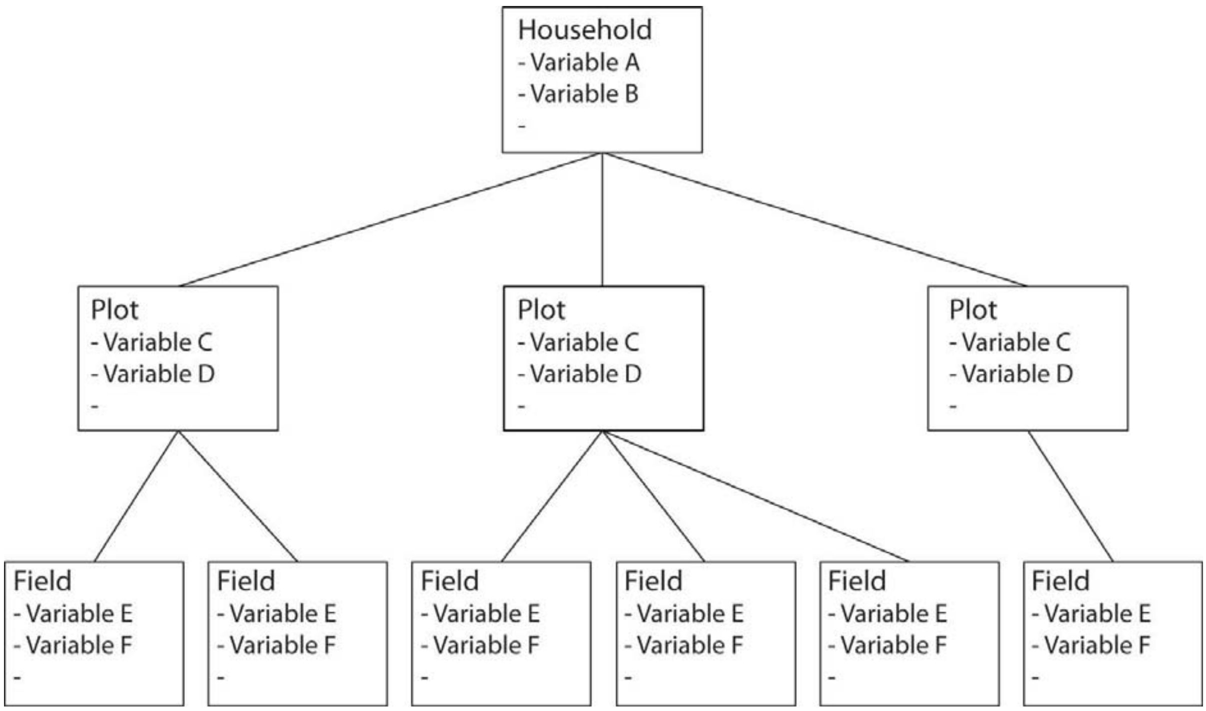

Figure 2.3: Hierarchical structure of the household level dataset presenting the relation between the levels Household, Plot and Field

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

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

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

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

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

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

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

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

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

Table 2.2: (Continued)

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

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

2.3 Methods

2.3.1 Analysis

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

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

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

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

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

Mainstream geographic approach

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

Household model

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

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

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

Enhanced spatial model

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

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

2.3.2 Logistic regression

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

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

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

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

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

2.4 Results

2.4.1 Models based on mainstream geographic approach

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

Table 2.3: Results of the mainstream geographic models

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

** 95% confidence interval*

2.4.2 Household models

The result of the household analysis is presented in Table 2.4. For all land use types a clear relation with slope can be observed, which is in line with the hypotheses. The variables *steep slope* and *moderate to steep slope* have a negative effect on the probability for a field to have corn. The model for wet rice shows a positive relation with *flat slope* and *flat to moderate slope*. On the contrary banana has a positive relation with steep and a negative relation with flat slopes. So, slope is a good determinant to make the distinction between corn and wet rice on the flatter fields and banana on the steeper fields. An explanation for this is that for corn and wet rice regular tillage is necessary and for wet rice irrigation has to be applied. Both activities can be best performed in the flatter areas. The remaining steep parts are cultivated with banana in which no regular tillage is applied.

The different preferences of ethnic groups turned out to be significant as well as the variable *place of birth*, which is partly related with the ethnic groups, because some groups have a distinct migration period. In general, the Ibanag have a tradition in growing corn and are less focussed on rice cultivation. The Tagalog speaking people, the Ifugaos and also the Ilocano people have a tradition in rice cultivation. In the corn model a positive relation was found with the Ibanag people and a negative relation with the Ifugao people. In the wet rice model a positive relation with the Ifugao and Ilocano people was found. But, besides ethnicity also the variable *place of birth* turned out to contribute significantly to the model. Being born in San Mariano has a positive effect on the probability of corn and a negative

Table 2.4: Results of the household models

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

* 95% confidence interval

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

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

Besides the strong relation with slope banana cultivation has a negative relation with the variable *'first year the farmer started on the plot'*, which means that recently acquired plots are often not cultivated with banana, but with other crops like corn and rice. Banana is negatively related with *plot size*. This is remarkable, because the hypothesis was that small field are cultivated with corn and rice and bananas are cultivated extensively on the larger plots. A possible explanation is that in the limited capacity for farmers to transport bananas from their fields leading to acreages planted with banana that do not exceed the transportation capacity.

The variables included in the models can be explained in how they affect land use decisions. On the other hand some variables that were hypothesised to be important to land use were not included in any of the models. Especially transportation costs and tenure were expected to be important. Transportation costs were hypothesised to influence the decisions for cash crops like yellow corn. Another strong relationship that was not included in these models is the positive relation between tenancy and corn.

The independent variables of the three models were investigated on any correlation to see if there is any collinearity within the models. This resulted in excluding *transportation cost* in the wet rice model, because this variable was highly correlated with *ethnicity Ifugao*. The other variables did not show high correlations among each other.

The ROC values that are used to give an indication for the goodness-of-fit are high for all three models, but especially for the land use types wet rice (0.92) and banana (0.92). The ROC for the corn model is 0.84. The high ROCs indicate that the variables used in the model can very well predict the occurrence of the land use types corn, wet rice and banana at the household level.

2.4.3 Enhanced spatial models

The enhanced spatial models are presented in Table 2.5 and the results are visualised in Figure 2.4. This figure shows the actual land use pattern (left) and the predicted probability (right). Especially for land use types corn and banana the predicted probabilities seem to fit the observed data well. For wet rice, the visual interpretation of the goodness-of-fit is less easy, because there are not many rice fields present in the area. The ROC values for the spatial models are 0.78, 0.76 and 0.74 for respectively the corn, wet rice and banana model.

The spatial models, incorporating the improved variables, have characteristics of both the household and mainstream geographic models (Table 2.5). The variable *improved distance to village* appears in the models for wet rice and banana. Wet rice is an intensive crop and is cultivated close to the villages and banana is situated further away. So, people grow the intensive crops close to their houses, or the other way around, people have settled where the conditions for growing their main crops is best. The variable *improved distance to market* wet is negatively related with corn. The (time) distance to market represents the costs to transport in- and outputs to the market and the negative relation indicates that the crop is grown at places with low transportation cost, generating the largest profits. The variable *distance to small river* turned out to be positively correlated with corn and negatively with banana. Only the small rivers are used for irrigation, since hardly any technology is used to use the water from the big rivers. Therefore, the variable *distance to small river* was created to approximate suitable locations for irrigation. *Slope* was selected in the corn and in the banana model and the relations are as hypothesised: negative for corn and wet rice

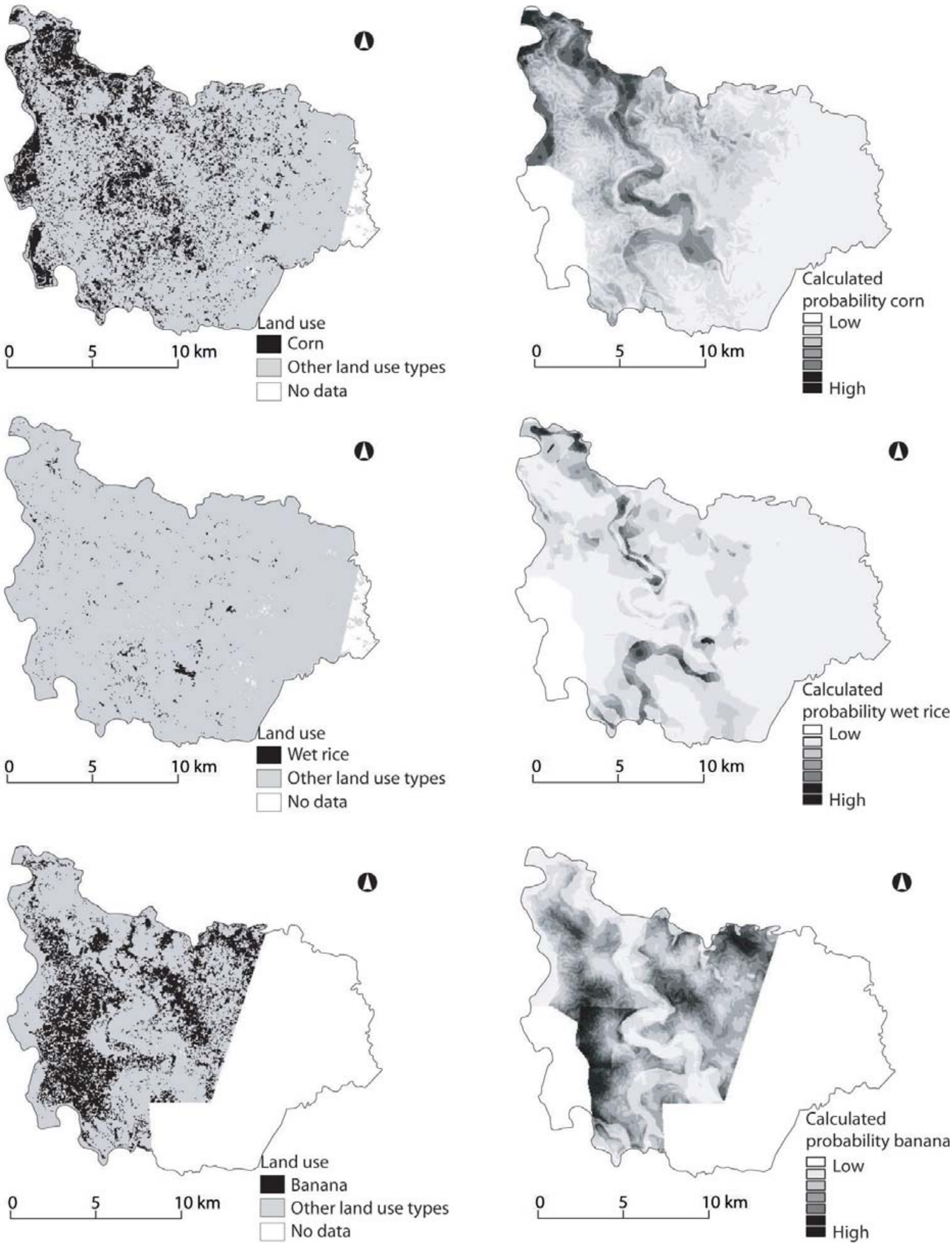

Figure 2.4: Actual land use (left) and predicted probabilities of the enhanced spatial models (right) for the three land use types corn (top), wet rice (centre) and banana (bottom)

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

Table 2.5: Results of the enhanced spatial models

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

** 95% confidence interval*

In general the active floodplain and terraces are flat and have better conditions for crop production. On these sites the intensive crops, yellow corn and wet rice, are cultivated. The less favourable areas are left for banana, which is cultivated in an extensive manner in this region. The variable indicating place of birth (*original municipality*), which was introduced based upon the experiences with the household level analysis, appeared to be significant in the model for wet rice. The relation is the same as in the household model. The variable *tax declaration* also turned out to be relevant in the wet rice model. The positive relation suggests means that wet rice is cultivated more on plots with a relatively strong tenural security. The positive relations with forest policies in the wet rice (*project ISF*) and banana model (*project IFMA*) are the contrary to what was hypothesised. It was expected that all agricultural land use types would occur less or equal at locations designated for forest

policies compared to other areas, because the forest policies are mainly focussed on reforestation and agroforestry. Though, the ISF program aims at developing only 20 percent of the area with forest trees and the remaining 80 percent with agroforestry and it aims at increasing upland production (Balagtas-Mangabat, 2002). So, the higher probability of wet rice in the ISF area could be well caused by the program. The relation between banana and IFMA could be caused by misclassification of the remote sensing images, since banana and secondary forest or reforestation are difficult to distinguish from each other.
The variables selected initially by the stepwise procedure were checked on collinearity. This resulted in excluding *improved distance to village* from the corn model (because of correlation with the variable *improved distance to market*) and *ethnicity Ibanag* from the banana model (correlated with *Geo. Rock*).
In contrast with the household models the spatial models do not incorporate any ethnicity variables, but do incorporate accessibility measures. These differences are discussed in the following section.

2.5 Discussion and conclusions

The mainstream geographic models explaining LUCC are based on readily available biophysical and geographical data, or on data that are easy to calculate from basic data, like distance measures. From the results of the mainstream geographic model in this study it could be concluded that the location of the land use types yellow corn, wet rice and banana are primarily determined by distance to village, road and market, and slope. These are location specific measures, which are basically physical characteristics. Though, the distance measures can be proxies for other process related variables like transportation cost and travel time to the field.
The household models show somewhat different results. From the household analysis it is clear that variables from all three levels present in the household survey (household, plot and field) play a significant role. Biophysical characteristics of the field, like slope and the presence of a creek, as well as social-economic characteristics of the household, like ethnicity and place of birth, are important. This advocates the incorporation of household characteristics in the spatial model, in addition to the geographical characteristics. The household model also shows the relevance of using the field as the unit of analysis in the household study. By taking the smallest unit as the level of analysis and linking the higher-level characteristics to this level the biophysical and socio-economic variables can be incorporated jointly in the statistical analysis. The result shows that both types of variables play a role in explaining the occurring land use.
The spatial model presented in this chapter combines the best of both. The dataset for this analysis was a combination of newly specified spatial variables that best represent variables that were significant in the household model and improved spatial variables complemented with spatial variables that are considered to have explanatory power at the watershed level. In these models a combination of accessibility measures, social and physical variables turned out to be the explanatory factors for land use. The main differences between the household models and the spatial models are that (1) the variables that are representing the same do not always have the same relative importance (indicated by the standardised b (b^*) in the models) and (2) that the goodness-of-fit (indicated by the ROC value) of the spatial models is substantially lower than the household models. The

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

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

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

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

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

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

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

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

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