



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

Koen P. Overmars^{a,b,*}, Peter H. Verburg^b

^a *Leiden University, Institute of Environmental Sciences (CML), P.O. Box 9518, 2300RA Leiden, The Netherlands*

^b *Wageningen University, Environmental Sciences Department, P.O. Box 37, 6700AA Wageningen, The Netherlands*

Received 21 January 2005; received in revised form 25 August 2005; accepted 14 October 2005

Abstract

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

* Corresponding author. Tel.: +31 71 527 1495; fax: +31 71 527 7496.
E-mail address: overmars@cml.leidenuniv.nl (K.P. Overmars).

can address many important land use issues involving scales and levels. Therefore, it is important in land use change research to formulate hypotheses that explicitly take scale and levels into account and then collect the appropriate data to answer these questions with approaches such as multilevel analysis.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Land use; Multilevel analysis; Philippines; Households; Scale; Level; Regression

1. Introduction

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

Within the LUCC discipline as a whole and in reference to regression approaches in particular, LUCC scientists have argued the importance of scale and levels (e.g., McConnell and Moran, 2001; Veldkamp and Lambin, 2001; Walsh et al., 2001; Nelson, 2002; Rindfuss et al., 2004). Gibson et al. (2000) state that scale is the spatial, temporal, quantitative, or analytical dimension used by scientists to measure and study objects and processes and level refers to specific locations along a scale. For this paper the following definitions are used: Levels refer to organisational levels originating from social context, for example, household level, village level and municipality level and scale is used for artificial resolution and extent originating from a geographic representation of reality in maps. The following issues regarding scales and levels that are important in land use (change) analysis can be identified (Gibson et al., 2000; Verburg et al., 2004b). First, land use is the result of processes that act at different scales and levels, which ideally would be addressed simultaneously. The choices that are made in a study about the extent and the unit of analysis determine to a large extent what patterns will be observed and which correlation will be found. Often, these choices are different between disciplines (Verburg et al., 2003). Second, scale and levels are important in identifying relations, but the fact that a relation occurs at a certain scale or level does not explain the phenomenon. Therefore, causal statements between variables should be made explicit and tested.

Within these causal statements scale and level are important factors, because different relations occur at different scales and levels. Moreover, causal relations can occur between different scales and levels. For example, village level variables like population or leadership capacity of the village head can influence land use at field level. Third, aggregation of processes to a higher level does not straightforwardly lead to a proper representation of these higher level processes because relations identified at the micro-level (or fine resolution) does not automatically translate into the same relation at the macro-level (or coarse resolution) (Robinson, 1950; Jones and Duncan, 1995; Easterling, 1997). The other way around the same phenomenon occurs: Inferences made on higher levels can often not be directly translated to lower level processes. Finally, all analyses, and therefore the insights from these analyses, are bounded by resolution or level of analysis and extent, which are determined by data structure and choices made by the researcher.

Mostly, scale and level issues are identified by comparing analyses at different resolutions and levels. Geoghegan et al. (2001) and Overmars and Verburg (2005) compared an analysis of land use decisions based on a household dataset with an analysis using a spatial dataset. Walsh et al. (2001) and Veldkamp et al. (2001) analysed the relation between land use and its explanatory factors at different resolutions created by aggregating grid data. However, the incorporation of multiple scales and levels in the analysis and including interactions between levels is often lacking. So far, the statistical tools that explicitly deal with these issues are not often applied as noted by Pan and Bilsborrow (2005) and Polsky and Easterling (2001). Multilevel modelling, which is the approach used in this study, is one of the statistical tools that are potentially capable to integrate artificial scales and organisational levels and to include interactions between these scales and levels. Multilevel statistical modelling allows for the analysis of data with complex patterns of variability that originate from hierarchical structure (Snijders and Bosker, 1999).

Multilevel modelling has mainly been used in the social sciences, for example, in sociology, education, psychology, economics, criminology (Snijders and Bosker, 1999), and is becoming more popular in geographic applications (e.g., in studying transport and land values (Schwanen et al., 2004; Polsky and Easterling, 2001)). In most of these applications multilevel modelling is used to study the effects of social context on the individual behaviour and to study the confusion between aggregate and individual effects. Land use studies can potentially benefit much from multilevel analysis, because land use data often has a very clear hierarchical structure (e.g., administrative levels, agro-ecological divisions and subdivisions, societal levels, artificial scales). Therefore, it is remarkable that multilevel modelling is not (yet) widely applied in land use studies. Some land use studies do incorporate data from multiple levels, but only few actually use multilevel modelling (Hoshino, 2001; Pan and Bilsborrow, 2005).

This paper aims to use multilevel analysis as the methodology to construct a predictive statistical model for the occurrence of land use that is statistically sound and which integrates different scales and levels. On the basis of a case study from a municipality in the Philippines different multilevel models will be presented that explain the occurrence of two major crops on individual fields in the area. In the discussion we

explore and describe the (surplus) value of multilevel modelling for land use studies regarding the issues of scale and levels in LUCC research and describe the implications of multilevel modelling for data collection.

2. Multilevel analysis

In this section, a short introduction of multilevel models is given in respect to land use issues in a general manner regardless of the outcome variable. Specific differences exist between models with a continuous, binary or multinomial outcome variable regarding estimation, model formation and the interpretation of coefficients. For the case study the logistic approach was adopted and the model specification is given in Section 3.3.

Multilevel analysis (e.g., Goldstein, 1995; Snijders and Bosker, 1999) is a methodology designed for the statistical analysis of hierarchically structured data. Multilevel regression models explicitly take the variability at different levels into account. Therefore, it is potentially a valuable tool in dealing with scaling issues in land use analysis. Multilevel modelling can address the scales and levels that are important to the land use system simultaneously, it can test hypothesis between scales and the modeller is not forced to aggregate or disaggregate data to one unit of analysis. Multilevel modelling can deal with nested data, such as hierarchically structured administrative units (e.g., farms in municipalities), as well as handle cases with observations that are structured differently, like lower level observations that are member of several groups at the higher level (e.g., farmers that have several buyers for their products).

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

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

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

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

3. Materials and methods

3.1. Study area

The study area is situated in Cagayan Valley in the north-eastern part of the island Luzon, the Philippines (Fig. 1). The study area includes 13 villages (barangays) in the municipality of San Mariano, in the province of Isabela, and comprises approximately 230 km². It is situated between the town of San Mariano in the west and the forested mountains of the Sierra Madre in the east. The population is approximately 20,000 persons (about 4000 households) of various ethnic groups,

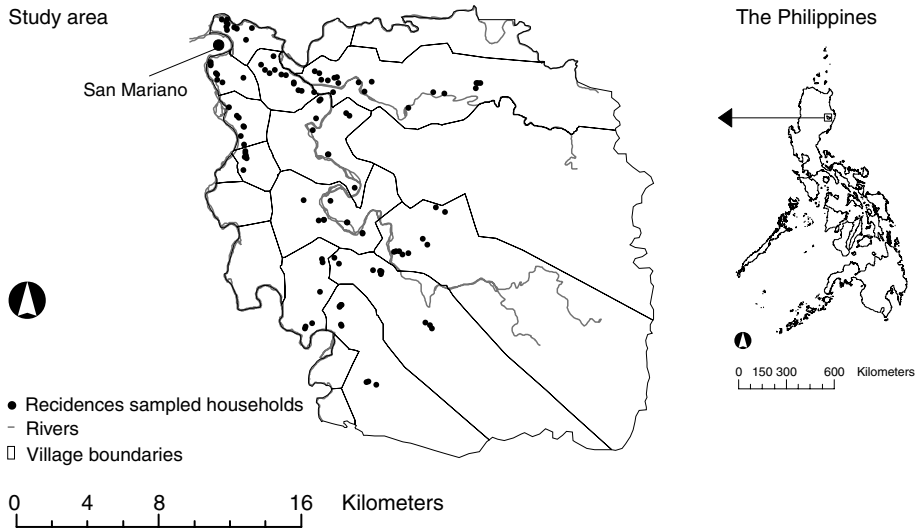


Fig. 1. Location of the study area in the Philippines and the location of the households' homes within the area.

among whom the Ilocano, Ibanag and Ifugao, who are all migrants or descendants of migrants that came to the area from the 1900s onwards, and the Kalinga and Agta, who are the indigenous inhabitants. Before immigration started, the area was completely forested with tropical lowland forest. At present, the study area shows a clear land use gradient ranging from intensive agriculture (mainly wet rice and yellow corn) near San Mariano via a scattered pattern of wet rice, yellow corn, banana, grasses, and (fruit) trees in the foothills to residual and primary forest in the eastern part. In the area a village unit actually consists of a group of settlements (sitios). The people live in these settlements, while their fields are often located in the surroundings of the settlement at an average distance of about 30 min walking.

3.2. Data

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

The selection of households to be interviewed was based on systematic random sampling using population data available at the POPMAT (POPulation Manipulation Action Team) member in the village. In all 13 villages every twentieth household

was selected (systematic random sampling with sampling interval 20) from the POP-MAT's list. From a total of approximately 3150 households in the 13 villages, 151 households were interviewed. The number of interviews per village ranges from 6 in the least populated village to 20 in the most populated. For the selected households the relevant characteristics were recorded for all fields (where a field is defined as a piece of land of a single owner used for one crop type). A household often owns or uses a number of fields at different locations and which are cultivated with different crops.

The most detailed (nested) hierarchy in the area, relevant to the land use system, could be constructed as follows (from the lowest level to the higher level): fields – plots (where a plot consists of a number of adjacent fields from the same owner) – households – sitios (the settlements) – villages – municipality. For the analysis only the field, household and village level were used (see Fig. 2). This is the most functional grouping, because the plots consist mostly of only one field and the dataset does not contain enough observations to discriminate between sitio and household level. Most sitios have only one or two households within the sample, which is insufficient for a proper multilevel analysis. Each of the variables was collected at its corresponding level, e.g., soil characteristics and slope at field level and household structure at the household level. Village level variables were derived from census data of 1997 (data about ethnicity and the percentage of the population that is born in the municipality of San Mariano).

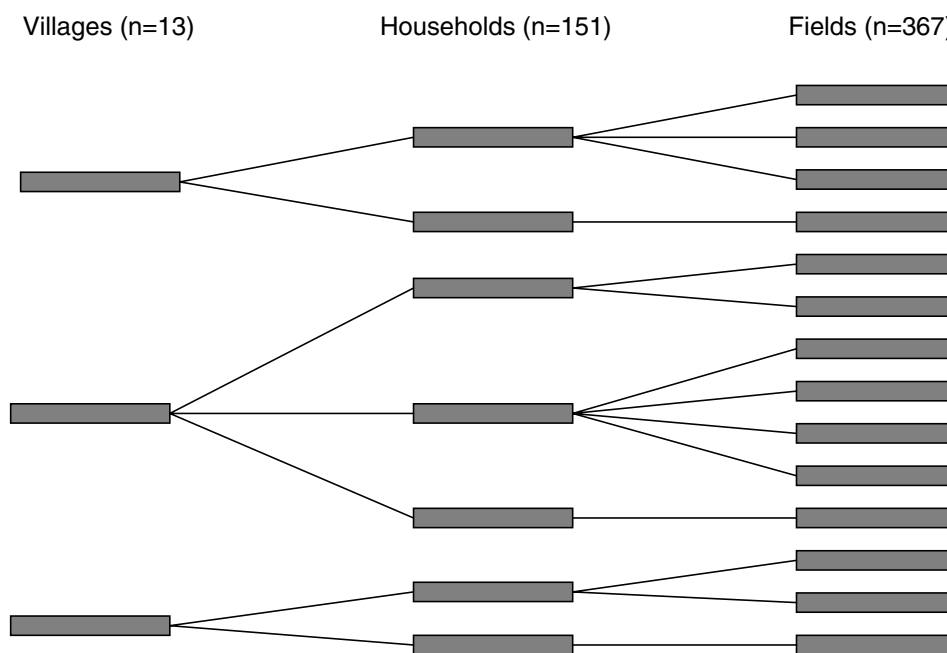


Fig. 2. Schematic representation of the hierarchical structure of the dataset.

Table 1
Description of the variables in the dataset used in this study

Variable name	Description	Min.	Max.	Mean	SD
<i>Dependent variables (field level: level 1, n = 297)</i>					
Yellow corn	1 if yellow corn, 0 otherwise	0	1	0.532	0.500
Banana	1 if banana, 0 otherwise	0	1	0.215	0.412
<i>Independent variables at field level (level 1, n = 297)</i>					
Slope 1	1 if slope category is flat, 0 otherwise	0	1	0.380	0.486
Slope 2	1 if slope category is flat to rolling/moderate, 0 otherwise	0	1	0.229	0.421
Slope 3	1 if slope category is rolling/moderate, 0 otherwise	0	1	0.283	0.451
Slope 4	1 if slope category is rolling/moderate to steep/hilly, 0 otherwise	0	1	0.081	0.273
Slope 5	1 if slope category is steep/hilly, 0 otherwise	0	1	0.027	0.162
Creek	1 if there is a creek or spring trough or bordering the plot, 0 otherwise	0	1	0.593	0.492
Plot distance	Hours walking from the residence of the household to the plot (h)	0	10	0.511	1.043
<i>Independent variables at household level (level 2, n = 115)</i>					
Ethnicity Ilocano	1 if male household head is Ilocano (or Tagalog speaking), 0 otherwise	0	1	0.539	0.501
Ethnicity Ifugao	1 if male household head is Ifugao, 0 otherwise	0	1	0.087	0.283
Ethnicity rest	0 if ethnicity is Ilocano or Ifugao, 1 otherwise	0	1	0.374	0.486
Transportation cost	Cost to transport a bag of corn from the residence to San Mariano (pesos)	7	45	22.652	12.214
Municipality of origin 0	1 if both male and female were not born in San Mariano, 0 otherwise	0	1	0.244	0.431
Municipality of origin 1	1 if male or female head is born in San Mariano, 0 otherwise	0	1	0.322	0.469
Municipality of origin 2	1 if both male and female were born in San Mariano, 1 otherwise	0	1	0.435	0.498
<i>Independent variables at village level (level 3, n = 12)</i>					
Ethnicity Ilocano (village)	Fraction of the population of the village that is Ilocano (or Tagalog speaking)	0.021	0.900	0.573	0.259
Ethnicity Ifugao (village)	Fraction of the population of the village that is Ifugao	0.000	0.404	0.076	0.147
Municipality of origin (village)	% of the population of the village born in San Mariano	64.899	99.007	84.479	9.748

Records with missing data were omitted from the dataset. Table 1 presents the dataset as it was used in the analysis, which is a subset of the original dataset and includes the most relevant variables based on preceding research and field experience (Overmars and Verburg, 2005).

3.3. Multilevel model specification

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

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

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

In this model p is the probability for the occurrence of the event, which in this study is the occurrence of a land use type on a field, β_0 is an intercept, β_n are regression coefficients to be estimated, and the x_n are exogenous explanatory variables.

The simplest imaginable way to incorporate levels would be to identify explanatory variables at the lower and the higher level:

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \alpha_1 z_1 + \dots + \alpha_m z_m. \quad (2)$$

Here, β_n and α_m are regression coefficients to be estimated, and the x_n are exogenous explanatory variables at the lower level (e.g., field) and z_m are explanatory variables at the higher level (e.g., a household).

Actually, many studies apply this approach by including variables from different levels in the regression model (e.g., Müller and Zeller, 2002; Overmars and Verburg, 2005) but do not report this explicitly. This model is typically called a fixed effect model since it lacks the random effects corresponding to higher levels in a multilevel model (Snijders and Bosker, 1999). The assumptions that belong to this model are that the residuals are mutually independent and have a zero mean. An additional assumption that is often made is that all groups have the same variances (homoskedasticity assumption). Implicitly the assumption is made that all group structure is represented by the explanatory variables. If this is not the case the residuals will be heteroskedastic. A second problem with this approach is that the higher level data is often disaggregated to the lowest level. As said before, this will lead to type I errors. The following models describe how the effects of the different levels can be incorporated into the regression model. With these models the assumptions stated above can be tested.

The model in Eq. (3) incorporates group effects but as yet without any explanatory variables. Besides the general intercept a random term U_{0j} is introduced, which is a group dependent intercept, in other words, an error term at the group level. With

this random term the variance that exists between groups is modelled explicitly. The effect of being a ‘member’ of a specific group is taken into account. Introducing this term will help to prevent the residuals from being heteroskedastic.

For reasons of clarity indices mark the different levels: i for level 1, j for level 2 (and k for level 3) and a zero indicates that a parameter is not variable at that level

$$\log \left(\frac{p_j}{1-p_j} \right) = \gamma_{00} + U_{0j}. \quad (3)$$

In Eq. (3) γ_{00} is the general intercept and U_{0j} is the group dependent deviation. The deviations U_{0j} are assumed to be independent and normally distributed with a zero mean and a variance of τ_0^2 (Snijders and Bosker, 1999).

This model is called the ‘pure random effects model’, ‘empty model’ or ‘unconditional model’. The empty model is a random intercept model without explanatory variables. With this model the variance of the dependent variable can be decomposed in a part caused by the individual level and a part caused by the group level (Snijders and Bosker, 1999; Polsky and Easterling, 2001). We will use this model as the base model to estimate if the group level variance in the dependent variable is significant. In the case study this is called model 1.

Including explanatory variables leads to the following model:

$$\log \left(\frac{p_{ij}}{1-p_{ij}} \right) = \gamma_{00} + \gamma_{10}x_{1ij} + \dots + \gamma_{q0}x_{qij} + \gamma_{01}z_{1j} + \dots + \gamma_{0r}z_{rj} + U_{0j}, \quad (4)$$

where x_{qij} are q explanatory level-1 variables and z_{rj} are r explanatory level-2 variables. Again, the deviations U_{0j} are assumed to have zero mean (given the values of the explanatory variables) and a variance of τ_0^2 (Snijders and Bosker, 1999). This model (Eq. (4)) is a random intercept model: a model where the intercept varies randomly between groups. The first part, $\gamma_{00} + \gamma_{10}x_{1ij} + \dots + \gamma_{q0}x_{qij} + \gamma_{01}z_{1j} + \dots + \gamma_{0r}z_{rj}$, is called the fixed part of the model and the second part, U_{0j} , is the random part of the model. In the case study analysis model 2, 3, 4 and 5 are based on this model (note that in the case study the model is extended to a model with 3 levels).

The interpretation of the regression coefficients is similar to ordinary logistic regression and is facilitated by the odds ratio ($\exp(\gamma)$). 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 $\exp(\gamma)$ for every unit increase of the variable corresponding to γ (Neter et al., 1996).

Starting from the empty model variables can be added at all levels. Variables at the individual level can explain part of the individual level variability as well as part of the group level variability, in the case when the values of the level one variable are consistently higher or lower than the general mean. For example, the slopes of the fields can be consistently higher in some of the villages and lower in others. Incorporating this field level variable can account for village level variability detected with the empty model.

Variables at the higher level(s) can be grouped in higher level variables that can only be observed at the higher level (e.g., the presence of a secondary school in a village) and aggregates of lower level variables (e.g., the average income of the inhabitants, which is an aggregate of observations at a lower level). Including these aggregates allows for the separation of the effect at the individual level and the effect at group level, which gives insight in the way a variable influences the outcome. In a model with only the level 1 data of that variable included the effect at both levels is forced to be equal (Snijders and Bosker, 1999). This difference is important while interpreting the regression coefficients. As described in Section 1 processes at the aggregate level can be substantially different from processes at the individual level. The village level variables in this study are of the aggregated type. Although they were calculated from census data, they have their equivalent at the household level.

The random intercept model (Eq. (4)) can be expanded by introducing random slopes. In a model with random slopes the regression coefficients (γ_{q0}) that act on the explanatory (level 1) variables are subdivided in a fixed and a random part. The addition of random slopes allows specific variables to differ by group. Even more complexity can be modelled by introducing level 2 variables in these slopes to explain (part of) the differences in slope. This is actually the same as a cross-product with an explanatory variable from level 1 and an explanatory variable from level 2. In multilevel modelling this cross-product is called cross-level interaction (Snijders and Bosker, 1999). In this study, random slopes and cross-level interactions were not included in the models. This will be explained in greater detail in the final discussion.

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

$$\log\left(\frac{\pi}{1-\pi}\right) = \gamma_{000} + \gamma_{100}x_{1ijk} + \cdots + \gamma_{q00}x_{qijk} + \gamma_{010}z_{1jk} + \cdots + \gamma_{0r0}z_{rjk} \\ + \gamma_{001}a_{1k} + \cdots + \gamma_{00s}a_{sk} + R_{0jk} + U_{00k}. \quad (5)$$

In Eq. (5) the a_{sk} are s explanatory level 3 variables, the R_{0jk} is the level 2 random part and U_{00k} is the level 3 random part. In this model fields are the unit of analysis at level 1, level 2 consists of households and level 3 are the villages. The dependent variable Y is land use. If $Y = 1$ the land use occurs, if $Y = 0$ the land use does not occur and p is the probability that the land use is found on that field.

Two analyses will be presented: one explaining the occurrence of yellow corn and one explaining the occurrence of banana. These are the most dominant crops in the study area (53% of the fields were cultivated with corn and 22% with banana). In the analysis, we present five different random intercept models per land use type. The first model is the empty model, which informs about the variability at the different levels. In the subsequent models variables will be added per level to see the influence of these groups of variables on the variance component of the higher levels.

The variables included were selected by studying prior analyses (Overmars and Verburg, 2005; Overmars et al., 2006) and field experience. For the corn model variables from the following list were added in different compositions: slope, creek and

plot distance at the field level; transportation cost, ethnicity and municipality of origin at the household level; and averages of municipality of origin and ethnicity at the municipal level. For the banana model the same variables were used except for presence of creeks, because this was considered to be of no influence to the occurrence of banana.

The analysis is performed with HLM software (Raudenbush et al., 2004). All models were estimated using the PQL (Penalized Quasi likelihood) routine. In HLM6 all 3-level hierarchical generalised linear models are estimated by full PQL by default (Snijders and Bosker, 1999; Raudenbush et al., 2004).

To indicate the proportion of variance that is accounted for by the group level the intraclass correlation coefficients (ρ_R and ρ_U , for the household and village level, respectively) are calculated. Eq. (6) shows the calculation of the intraclass correlation coefficient for the household level. (Snijders and Bosker, 1999; Browne et al., 2005).

$$\rho_R = \text{var}(R_{0jk}) / (\text{var}(R_{0jk}) + \text{var}(U_{00k}) + \pi^2/3), \quad (6)$$

where ρ_R is the intraclass coefficient for the household level, $\text{var}(R_{0jk})$ is the variance of the random intercept at household level and $\text{var}(U_{00k})$ is the variance of the random intercept at village level. A logistic distribution for the level one residual implies a variance of $\pi/3$, which appears as the level 1 variance in Eq. (6) (Snijders and Bosker, 1999). In an linear multilevel model this would be the level 1 variance σ^2 .

To assess the goodness-of-fit of the models the relative operating characteristic (ROC) (Swets, 1988) was used. This measure is capable to assess the quality of the predictor and can be compared between different models. The ROC summarises the performance of a logistic regression model over a range of cutoff values classifying the probabilities. The value of the ROC is defined as the area under the curve linking the relation between the proportion of true positives versus the proportion of false positives for an infinite number of cutoff values. The ROC statistic varies between 0.5 (completely random) and 1 (perfect discrimination).

4. Results

4.1. Corn models

This section presents various multilevel models, with different sets of explanatory variables, predicting the occurrence of yellow corn on a field. Model 1 is the empty model, which does not include any explanatory variables, but only includes random effects at the higher levels. Model 1 (Table 2) shows that the variance is significant ($p < 0.05$) at both level 2 and 3. The intraclass correlation coefficients (ρ_R and ρ_U , Table 2) indicate that 10% of the variance can be attributed to the household level and 4% to the village level. The remaining variance is in level 1, which is fixed in this modelling approach to $\pi/3$. Thus, both the households and the villages show significant clustering of the occurrence of corn. The variance detected in this model might be accounted for by explanatory variables. This is studied with the models 2, 3, 4 and 5.

Table 2
Multi-level models for yellow corn

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

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

Model 2 introduces a set of geographic and biophysical variables that are known explanatory variables for the occurrence of corn in the study area. These are slope, presence of a creek, hours walking from the residence of the household to the plot, and the cost to transport a bag of corn from the residence to San Mariano.

Table 2 (model 2) shows that almost all explanatory variables (the fixed effects) have significant coefficients. Corn is more likely to occur on field that are flatter, not close to a creek, close to the household's residence and close to the market town of San Mariano. The random part of level 3 turns out to be lower. So, these variables explain some of the variance at the village level detected in the empty model. This might be caused, for example, by the fact the transportation costs vary on average per village because the villages are situated at different distance from the market place. By introducing this variable (or perhaps one of the other variables) the

variability disappeared from the village level. Although the level 3 random part is not significant and, theoretically, level 3 could be excluded, the structure with three levels is maintained in order to study the level 3 behaviour in the following models. At the household level the variance component is still significant and similar to the variance component of the empty model. Thus, the variables included do not account for any of the household level variability.

Model 3 adds household variables to model 2. Additional to the relations in model 2 corn turned out to be negatively related with households where both the male and female are born outside the municipality and negatively with people of Ilocano origin. After including the household level variables the random part of level 2 (the household level) decreased substantially. Apparently, the variance at level 2 is captured by the included variables. The geographical/biophysical variables are still significant. The level 3 variance increased in comparison with model 2.

Model 4 adds the village level variables to the model 3 configuration. This model investigates if there is a fixed effect of the village level variables besides the variables included in model 3. For example, one can imagine that a village dominated by one ethnic group has an extra village level effect besides the effect of ethnicity at household level for the whole study area. The village level variables are aggregated values of variables at level 2 (ethnicity and municipality of origin). Instead of using the survey data to derive these level 3 variables census data of the complete population was used. Table 2 shows that there are no significant effects for the variables at village level. Including these variables results in a similar random part at the village level as model 3. Thus the level 3 variables did not explain any of the variance in level 3.

Model 5 was constructed to see if including the household variables at village level instead of at the household level would be a good alternative. This would be convenient because census data at village level is often more easily available than household level data. However, like in model 4, none of the village level variables are significant in model 5. Besides that, the variance component at household level is the same as in the models 1 and 2. This shows that the aggregated variables do not capture any of the variability at the household level. As theory suggests (Robinson, 1950; Jones and Duncan, 1995) the effect of the aggregate variable is quite different than that of its lower level equivalent.

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

4.2. *Banana models*

The analysis of the occurrence of banana shows a different result (Table 3). In the empty model (model 1) there are no signs of significant between-group variances. Model 2, which incorporates the geographical/biophysical variables, shows a significant relation between slope of a field and the choice to cultivate banana and a sig-

Table 3
Multi-level models for banana

Banana	Model 1	Model 2	Model 3	Model 4	Model 5
<i>Fixed effects</i>					
Level 1					
Intercept	-1.289***	-3.430***	-3.260**	-4.050	-4.120
Slope 3		2.389***	2.435***	2.432***	2.397***
Slope 4		5.022***	5.006***	4.975***	4.970***
Slope 5		5.634***	5.971***	5.765***	5.461***
Plot distance		0.0064	0.0076	0.0231	0.0207
Level 2					
Transportation cost		0.015	0.012	0.017	0.019
Ethnicity Ilocano			-0.297	-0.236	
Ethnicity Ifugao			-0.613	-0.500	
Municipality of origin 0			0.532	0.528	
Municipality of origin 1			-0.242	-0.247	
Level 3					
Municipality of origin village				1.253	1.198
Ethnicity Ilocano village				-0.380	-0.414
Ethnicity Ifugao village				-0.055	0.114
<i>Random effects</i>					
Level 2					
var (R_{0jk})	0.003	0.006	0.003	0.004	0.006
ρ_R	0.001	0.002	0.001	0.001	0.002
Level 3					
var (U_{00k})	0.107	0.546**	0.724**	0.637***	0.487**
ρ_U	0.031	0.142	0.180	0.162	0.129
ROC	0.694	0.906	0.909	0.908	0.903

* $p < 0.05$.

** $p < 0.01$.

*** $p < 0.001$.

nificant random effect at the village level. Thus, including variables results in a large and significant random part at the village level. This could not be explained completely. Part of the explanation is that in general changes in the fixed effects part can cause big changes in the random part while changes in the random part usually do not cause big changes in the fixed effect part. The mean structure the model can change the stationarity of the mean causing a shift in variance and making the random effects part significant.

Model 3 introduces household level variables ethnicity and municipality of origin, model 4 includes also village averages of these variables and model 5 includes only the village level and field level variables. All the coefficients of these variables do not differ significantly from zero, neither do they influence the levels 2 and 3 random parts significantly. Therefore, we conclude that these variables do not influence banana cultivation significantly and that this is predominantly determined by slope. To find out what process might cause the differences between villages the random

intercepts of the village level were examined. This did not show a clear pattern. Furthermore, models with additional explanatory variables and models with random slopes were tested, but this did not result in a satisfying explanation of the variability at the village level in model 2.

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

5. Discussion and conclusions

5.1. Multilevel statistics for land use studies

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

For corn cultivation the empty model indicated significant between-group variability at two higher levels (household and village). Explanatory variables at the household level turned out to account for that variability at that level (Table 2, model 3). Replacing some of the household level variables with their village level aggregates did show a significant variance component at the household level. From this it can be concluded that the household level variables cannot be substituted by village level aggregates in this case. The explanatory variables at the household level can explain a significant part of the occurrence of corn at field level. The exploratory procedure applied has revealed at which levels important variables explain land use decisions. Models 3 and 4 show that within these models a significant part of variability is left at the village level, which is left unexplained in these models.

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

The multilevel analysis of the land use system in the study area provided additional information to previous analyses based on conventional regression models (Overmars and Verburg, 2005). In the case of corn the analysis confirmed the hypothesis that the household level plays an important role. In this analysis the municipality of origin of the household (which is proxy for migration history) in combination with the ethnicity variables turned out to be a significant explanatory variables that account for the variability at the household level. For the case of banana the analysis confirmed the idea that household level characteristics did not play an important role in the decision to cultivate bananas. Bananas occur mostly on sites that are less productive or too steep for arable crops like corn, rice or vegetables. Significant village level variance indicates the importance of village level conditions in explaining the decision to cultivate banana.

Like in any other statistical analysis, drawing conclusions about the causality of the relations from the regression analysis should be done with care. For example, the positive relation between slope of a field and bananas results from the fact that the flatter areas are devoted to arable crops, not because bananas perform better on the steep slopes. Like any other regression analysis multilevel models can only reveal associations between variables and partition variance. Additional research is needed to study the causality of the relations. An example of such a method for the case study area is described in Overmars et al. (2006).

In this paper, random slopes were not incorporated in any of the models. Although there were no strong arguments suggesting that the coefficients for the explanatory variables were different, some experiments were carried out to study the behaviour of models that include random slopes. This resulted in either insignificant random slopes or models that did not converge. Most likely the data structure and the amount of observations made the estimation of the random slopes complicated. The number of observations (fields) per households is low and this complicates the determination of the random slopes.

The results indicate that the household level can be crucial in explaining land use at the field level. However, in many studies household level data is not available because in many regional studies the analysis is based on remote sensing, maps and census data (e.g., Nelson et al., 2001; Walsh et al., 2001; Müller and Zeller, 2002). As shows from this study simply substituting household level variables with their village level equivalents, which can be calculated from widely available census data, will most often not account for the household level variability because of errors due to aggregation. Disaggregating higher level variables to the level of analysis can lead to erroneous conclusions. In any case, disregarding the household level variables while explaining land use at field conclusion of Rindfuss et al. (2003) that the household level is the central level to be included in explanations of land use.

Data availability and data structure play an important role in land use studies. As illustrated in this paper, data availability determines at what level land use can be studied, and therefore at what level one can draw conclusions. If the hierarchical structure of the data is important to the land use system under study and the research questions that arise from this, this structure should be considered in the sample design to take full advantage of the multilevel modelling technique. Ideally, at every

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

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

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

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

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

Another important aspect to consider in land use studies is spatial dependency, which refers to the geographic law that nearby things are more related than distant things (Tobler, 1970). Spatial dependency in land use patterns can be caused by dependence of the land use pattern on an explanatory factor that is spatially struc-

tured (trend) or a spatial interaction process of the land use variable itself, like competition or imitation (Anselin, 1988; Irwin and Geoghegan, 2001; Overmars et al., 2003; Polsky, 2004). Both Polsky and Easterling (2001) and Pan and Bilbrough (2005) mention that multilevel modelling can partly reduce the effect of spatial autocorrelation when neighbouring observations are nested within one group. If the spatial dependency is only related to the nested hierarchy this might even correct for all spatial autocorrelation. However, often spatial dependency is structured differently than the nested hierarchy of the dataset. In this case the neighbourhood effects can be incorporated in the multilevel model as cross-random effects (where lower level observation can be member of different groups at the higher level). For example, each observation can be part of a group with all its neighbours. This approach would correct for spatial autocorrelation but is not yet studied in land use research. In this study this approach was not applied because the observed fields are relatively far apart due to the relatively small sample size and spatial autocorrelation is therefore assumed to be minimal.

5.2. Conclusions

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

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

use research toolbox that allows the exploration of a number of cross-scale propositions.

Acknowledgements

The authors acknowledge the Foundation for the Advancement of Tropical Research (WOTRO) of the Netherlands Organisation for Scientific research (NWO) for the funding of the research on which this paper is based. Special thanks go to Noel G. Perez who was the research assistant for this project during the fieldwork in the Philippines. Finally, we want to thank the CVPED (Cagayan Valley Programme on Environment and Development) staff in Cabagan, Isabela and the LGU (Local Government Unit) of San Mariano, Isabela for their support. We thank two anonymous reviewers for their constructive comments.

References

- Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht, 284p.
- Browne, W.J., Subramanian, S.V., Jones, K., Goldstein, H., 2005. Variance partitioning in multilevel logistic models that exhibit over-dispersion. *Journal of Royal Statistical Society A* 168, 599–613.
- Centre for multilevel modelling, 2005. <http://multilevel.ioe.ac.uk/softrev/index.html>.
- Chomitz, K.M., Thomas, T.S., 2003. Determinants of land use in Amazonia: A fine-scale spatial analysis. *American Journal of Agricultural Economics* 85 (4), 1016–1028.
- Easterling, W.E., 1997. Why regional studies are needed in the development of full-scale integrated assessment modelling of global change processes. *Global Environmental Change Part A* 7 (4), 337–356.
- Geist, H.J., Lambin, E.F., 2002. Proximate causes and underlying driving forces of tropical deforestation. *BioScience* 52 (2), 143–150.
- Geoghegan, J., Cortina Villar, S., Klepeis, P., Macario Mendoza, P., Ogneva-Himmelberger, Y., Chowdhury, R.R., Turner II, B.L., Vance, C., 2001. Modeling tropical deforestation in the southern Yucatán peninsular region: comparing survey and satellite data. *Agriculture, Ecosystems & Environment* 85 (1–3), 25–46.
- Gibson, C.C., Ostrom, E., Ahn, T.K., 2000. The concept of scale and the human dimensions of global change: a survey. *Ecological Economics* 32 (2), 217–239.
- Goldstein, H., 1995. *Multilevel statistical models* Kendall's Library of Statistics, vol. 3. Edward Arnold, London.
- Hoshino, S., 2001. Multilevel modeling on farmland distribution in Japan. *Land Use Policy* 18 (1), 75–90.
- Irwin, E.G., Geoghegan, J., 2001. Theory, data, methods: developing spatially explicit economic models of land use change. *Agriculture, Ecosystems & Environment* 85 (1–3), 7–24.
- Jones, K., Duncan, C., 1995. Individuals and their ecologies: analysing the geography of chronic illness within a multilevel modelling framework. *Health & Place* 1 (1), 27–40.
- Lambin, E.F., Bauleis, X., Bockstael, N., Fischer, G., Krug, T., Leemans, R., Moran, E.F., Rindfuss, R.R., Sato, Y., Skole II, D., Turner II, B.L., Vogel, C., 1999. Land-use and land-cover change (lucc): implementation strategy. Technical Report, IGBP Report No. 48 and IHDP Report No. 10. International Geosphere-Biosphere Programme (IGBP) and International Human Dimensions Programme on Global Environmental Change (IHDP), Stockholm.
- McConnell, W., Moran, E.F., 2001. Meeting in the Middle: The Challenge of Meso-Level Integration. An International Workshop on the Harmonization of Land Use and Land Cover Classification. LUCC Report Series No. 5. Indiana University, Bloomington, USA.

- Müller, D., Zeller, M., 2002. Land use dynamics in the central highlands of Vietnam: a spatial model combining village survey data with satellite imagery interpretation. *Agricultural Economics* 27 (3), 333–354.
- Nelson, G.C., 2002. Introduction to the special issue on spatial analysis for agricultural economists. *Agricultural Economics* 27 (3), 197–200.
- Nelson, G.C., Harris, V., Stone, S.W., 2001. Deforestation, land use and property rights: empirical evidence from Darien, Panama. *Land Economics* 77 (2), 187–205.
- Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W., 1996. *Applied Linear Statistical Model*, fourth ed. Irwin, Burr Ridge, IL, p. 1408.
- Overmars, K.P., Verburg, P.H., 2005. Analysis of land use drivers at the watershed and household level: Linking two paradigms at the Philippine forest fringe. *International Journal of Geographical Information Science* 19 (2), 125–152.
- Overmars, K.P., De Koning, G.H.J., Veldkamp, A., 2003. Spatial autocorrelation in multi-scale land use models. *Ecological Modelling* 164 (2–3), 257–270.
- Overmars, K.P., De Groot, W.T., Huigen, M.G.A., 2006. Comparing Inductive and Deductive Modeling of Land Use Decisions: Principles, a Model and an Illustration from the Philippines. *Human Ecology* (accepted).
- Pan, W., Bilborrow, R.E., 2005. The use of a multilevel statistical model to analyze factors influencing land use: A study of the Ecuadorian Amazon. *Global and Planetary Change* 47 (2–4), 232–252.
- Perz, S.G., Skole, D.L., 2003. Social determinants of secondary forests in the Brazilian Amazon. *Social Science Research* 32 (1), 25–60.
- Polsky, C., 2004. Putting space and time in Ricardian climate change impact studies: agriculture in the U.S. great plains, 1969–1992. *Annals of the Association of American Geographers* 94 (3), 549–564.
- Polsky, C., Easterling, W.E., 2001. Adaptation to climate variability and change in the US Great Plains: A multi-scale analysis of Ricardian climate sensitivities. *Agriculture, Ecosystems and Environment* 85 (1–3), 133–144.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., Lewis, T., 2000. *A User's Guide to MLwiN*, second ed. Institute of Education, London.
- Raudenbush, S., Bryk, A., Congdon, K., 2004. *HLM 6, Hierarchical Linear and Non-linear Modelling*. SSI, Scientific Software international.
- Rindfuss, R.R., Walsh, S., Mishra, V., Fox, J., Dolcemascolo, G.P., 2003. Linking household and remotely sensed data. In: Fox, J., Rindfuss, R.R., Walsh, S.J., Mishra, V. (Eds.), *People and the Environment: Approaches for Linking Household and Community Surveys to Remote Sensing and GIS*. Kluwer Academic Publishers, Boston, p. 344.
- Rindfuss, R.R., Walsh, S.J., Turner II, B.L., Fox, J., Mishra, V., 2004. Developing a science of land change: Challenges and methodological issues. *PNAS* 101 (39), 13976–13981.
- Robinson, W., 1950. Ecological correlations and the behaviour of individuals'. *American Sociological Review* (15), 351–357.
- Schwanen, T., Dieleman, F.M., Dijst, M., 2004. The impact of metropolitan structure on commute behavior in the Netherlands: A multilevel approach. *Growth and Change* 35 (3), 304–333.
- Snijders, T.A.B., Bosker, R.J., 1999. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel modelling*. Sage, New York, 266p.
- Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293.
- Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46, 234–240.
- Veldkamp, A., Lambin, E.F., 2001. Predicting land-use change. *Agriculture, Ecosystems & Environment* 85 (1–3), 1–6.
- Veldkamp, A., Verburg, P.H., 2004. Modelling land use change and environmental impact. *Journal of Environmental Management* 72 (1–2), 1–3.
- Veldkamp, A., Kok, K., De Koning, G.H.J., Schoorl, J.M., Sonneveld, M.P.W., Verburg, P.H., 2001. Multi-scale system approaches in agronomic research at the landscape level. *Soil and Tillage Research* 58 (3–4), 129–140.

- Verburg, P.H., Chen, Y.Q., 2000. Multiscale characterization of land-use patterns in China. *Ecosystems* 3 (4), 369–385.
- Verburg, P.H., De Groot, W.T., Veldkamp, A., 2003. Methodology for multi-scale land-use modelling: concepts and challenges. In: Dolman, A.J., Verhagen, A., Rovers, C.A. (Eds.), *Global Environmental Change and Land Use*. Kluwer Academic Publishers, Dordrecht/Boston/London.
- Verburg, P.H., Ritsema van Eck, J., de Nijs, T.C.M., Dijst, M.J., Schot, P., 2004a. Determinants of land use change patterns in the Netherlands. *Environment and Planning B* 31, 125–150.
- Verburg, P.H., Schot, P., Dijst, M., Veldkamp, A., 2004b. Land use change modelling: current practice and research priorities. *GeoJournal* 61 (4), 309–324.
- Walsh, S.J., Crawford, T.W., Welsh, W.F., Crews-Meyer, K.A., 2001. A multiscale analysis of LULC and NDVI variation in Nang Rong district, northeast Thailand. *Agriculture, Ecosystems & Environment* 85 (1–3), 47–64.