Statistical methods for analysing the spatial dimension of changes in land use and farming systems



Jan Peter Lesschen, Peter H. Verburg, Steven J. Staal

LUCC Report Series No. 7

Ecoregional Fund

WAGENINGEN UNIVERSITY
WAGENINGEN UR

ILRI
INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE

Statistical methods for analysing the spatial dimension of changes in land use and farming systems

Jan Peter Lesschen, Peter H. Verburg, Steven J. Staal

LUCC Report Series No. 7

Published by:

The International Livestock Research Institute, Nairobi, Kenya
&
LUCC Focus 3 Office, Wageningen University, the Netherlands

Statistical methods for analysing the spatial dimension of changes in land use and farming systems – LUCC Report Series 7

Bibliographical references
I. Lesschen, Jan Peter. II. Verburg, Peter H. III. Land-Use and Land-Cover Change (LUCC) Project. IV. International Human Dimensions Programme on Global Environmental Change (IHDP) V. International Geosphere-Biosphere Programme (IGBP)

Published by:

International Livestock Research Institute
P.O.Box 30709
Nairobi
Kenya

and

LUCC Focus 3 Office
Department of Environmental Sciences
Wageningen University
P.O. Box 37
6700 AA Wageningen
The Netherlands
http://www.lucc.nl

ISBN: 92-9146-178-4

ISSN: 1138-7424

Contents

Abbreviations and acronyms

AIC	Akaike information criterion
AUC	area under the curve
COV	coefficient of variation
GIS	geographic information systems
GPS	Global Positioning System
HYV	high-yielding variety
LUCC	land use and land cover change
NDVI	normalized difference vegetation index
OLS	ordinary least squares
PCA	principal component analysis
ROC	relative or receiver operating characteristic
SC	Schwartz criterion
SSE	error sum of squares
SSR	regression sum of squares
SST	total sum of squares
TLU	tropical livestock unit

Acknowledgements

This report is written as part of the project 'Transregional analysis of crop-livestock systems: understanding intensification and evolution across three continents' commissioned by the Ecoregional Fund and implemented by the International Livestock Research Institute (ILRI) in collaboration with the Department of Environmental Sciences of Wageningen University, and the Kenya Agricultural Research Institute. This report reviews a large amount of knowledge from the Land Use and Land Cover Change research community (LUCC, a joint IGBP/IHDP project) and is one of the activities of the LUCC Focus 3 Office hosted by the Department of Environmental Sciences, Wageningen University. Part of the writing was funded by the Foundation for the Advancement of Tropical Research (WOTRO) of the Netherlands Organization for Scientific Research (NWO) within the project 'Integrating macro-modelling and actor-oriented research in studying the dynamics of land use change in North-East Luzon, Philippines'.
The authors would like to thank all who contributed to this report, especially the contributions of Isabelle Baltenweck, Jeannette van de Steeg and Koen Overmars and the thoughtful reviews of two anonymous referees.

1 Introduction

Land use and land cover change (LUCC) has important impacts on the functioning of socio-economic and environmental systems with important tradeoffs for sustainability, food security, biodiversity and the vulnerability of people and ecosystems to global change impacts. Land cover change refers to the complete replacement of one cover type by another, e.g. deforestation. Land use change includes the modification of land cover types, e.g. intensification of agricultural management or other changes in the farming system. Land use and land cover changes are the result of the interplay between socio-economic, institutional and environmental factors. Key to understanding LUCC is to recognize the role of individual decision makers bringing about change, through their choices, on land resources and technologies. A unifying hypothesis that links the ecological and social realms, and an important reason for pursuing integrated modelling of LUCC, is that humans respond to cues both from the physical environment and from their sociocultural and economic contexts. Therefore, much LUCC research is devoted to the analysis of relations between land use and the socio-economic and biophysical variables that act as the 'driving forces' of land use change (Turner II et al. 1993; Turner II et al. 1995; Lambin et al. 2001). Driving forces are generally subdivided into two groups: proximate causes and underlying causes. Proximate causes are the activities and actions that directly affect land use, e.g. wood extraction or road building. Underlying causes are the 'fundamental forces' that underpin the proximate causes, including demographic, economic, technological, institutional and cultural factors (Geist and Lambin 2002). In most cases, a wide range of factors is used to represent the underlying causes; examples include soil suitability, population density, rainfall and accessibility. They can also be differentiated into 'driving' forces that are expected to change over time, such as population density and market conditions, and 'conditioning' factors that are relatively stable over time but may be spatially differentiated, such as agroclimate and cultural context. This allows differentiation into spatial and temporal expectations of change. At different scales of analysis different driving forces have a dominant influence on the land use system: at the local level this can be the local policy or the presence of small ecologically valuable areas; at the regional level distance to the market, port or airport might be the main determinant of land use change (Verburg et al. 2003).
Where driving and conditioning factors exhibit a high degree of spatial variation, such as in the cases of soil conditions and market access, this spatial variation gives rise to spatially distinct land use patterns related to the variations in social, economic and environmental context. Given the importance of spatial variation, LUCC research frequently uses techniques that analyse the relationship between land use and its supposed driving and conditioning factors based on spatially differentiated data. Empirical techniques are used to verify hypotheses of driving factors and quantify relations between driving factors, the decision maker and land use. The actual use of empirical techniques differs: often the prime interest of social scientists is explanation of observed land use changes, while ecologists focus on prediction. Whereas the use of spatial analysis for explanation enhances our understanding of the processes underlying LUCC (Nelson 2002), temporal prediction helps to explore the past and future importance of driving factors and model future land use dynamics (Briassoulis 2000). By predicting the spatial and temporal distribution of changes we can target areas for intervention, and develop appropriate, location-specific, intervention strategies. Therefore, spatially explicit analysis is increasingly used to predict landscape change and is often evaluated both in terms of conventional inference on variable coefficients and goodness of fit, and with respect to ability to predict actual landscape change (Pontius et al. 2004; Nelson and Geoghegan 2002).

It should be noted that the same spatial methods are generally equally relevant for farming system and technology choices that are not necessarily closely related to land use, such as changes in choices of livestock breed or crop variety. The spatial and temporal changes in driving and

conditioning factors that lead to LUCC also lead to changes in determinants of technology choice. For that reason, these analytical methods should be considered as important for a wide variety of analyses of farming systems change.

Scope of this report

A wide range of methods for spatial analysis exists and LUCC researchers often face similar problems (Rindfuss et al. 2004). Also, the use of statistical methods in the field of LUCC requires specific strategies, as well as a framework for understanding the role of the decision maker. In statistical textbooks the different techniques are often described in detail, without, however, a specific focus on land use change issues. This report intends to provide an overview of empirical methods that are frequently used for the analysis of spatial patterns of LUCC based on a survey of recent literature. Because these methods are relevant for wider analysis of system change beyond LUCC, some examples of analysis of livestock systems are included. The descriptions are not detailed, with the emphasis being instead on explaining the concepts in simple terms, with the aid of illustrations of these methods in land use research. The references refer to detailed descriptions, applications or textbooks. The methods discussed in this report aim at uses of different types of spatially differentiated data at different scales, including both household- and pixel-level analysis. Furthermore, a number of issues important to spatial analysis of land use and farming system change are discussed, including data representation, spatial autocorrelation and validation issues. This information should facilitate the application of these methods in LUCC and other studies and provide an overview of the possibilities and limitations of empirical methods to unravel the complexity of spatial variation in land use and farming system change.

History of spatial analysis of LUCC

Scholars have offered different explanations for variations in land use. In the early 19th century, von Thünen (1966) analysed the location of agricultural land as a function of distance to market centres and transport cost. According to him, agricultural intensity decreases with increasing distance from market centres. This explanation is a foundation upon which theories and explanations of land use are built, though it overlooks several other biophysical, socio-economic and institutional factors influencing land use (Rasul et al. 2004). The basic von Thünen model is illustrated in Figure 1.1. In a featureless plain, surrounding a central market, it is supposed that two crops are grown – wheat and vegetables. All locations have identical production characteristics, but transport costs to the central market, with exogenously determined prices, differ per crop. The price of vegetables in the central market (a) is higher than the price of wheat (b), but vegetables are more expensive to transport. Hence, the farm vegetable price falls more quickly than the farm wheat price as distance from the market increases. Beyond point (c), the farm price of wheat is higher than the price of vegetables. The result is a series of concentric rings of land use around the central market, indicated in the bottom part of the graph. In the shaded area, vegetables are grown; in the next ring, wheat. Beyond (d), neither crop is profitable and land is left in its natural state. Central to this model is the assumed rational economic behaviour of the farmer decision maker.

Even though the productive characteristics of the farms are assumed to be identical, the effect of transport cost on farmer incentives and choices causes a decline of land rents with distance from the central market. The basic model of von Thünen represents a featureless plain. Later the model has been adapted to take into account differences in land productivity, prices, transport costs and multiple markets, which makes the analysis more complex but the basic insights of the importance of location and transport cost in determining farmer incentives and so land use remain (Alonso 1964; Nelson 2002). The von Thünen model provides a simplified representation of the location decisions that have led to the diverse patterns of land use that cover the earth's

surface. The complexity of processes determining the current land use patterns and the locations of future changes in land use cannot be fully understood from a such a simple theoretical model, although it established the basic spatial economic framework for understanding land use and farming system change (Figure 1.1). Different studies have elaborated on the model of von Thünen to include a wider range of conditions (Alonso 1964; Walker 2004; Walker and Solecki 2004). Other disciplines have also contributed to a theoretical analysis of the processes leading to land use patterns, such as the 'new' economic theory of Krugman (Fujita et al. 1999; Krugman 1999). Although these theories do contribute to the explanation of LUCC, there is no single all-embracing theory to explain the variety in land use patterns. Therefore empirical methods are frequently used to explore land use change data to find evidence for the proximate causes of land use change and its location (Turner II et al. 1990).
Different empirical approaches have been suggested to explain the spatial patterns of LUCC, some strictly following the (economic) theoretical framework while others allow a broader exploration of the correlative structure between land use patterns and the spatial patterns of land use. A well-cited example of an empirical analysis based on the von Thünen model is the study of Chomitz and Gray (1996) for land use in Belize.

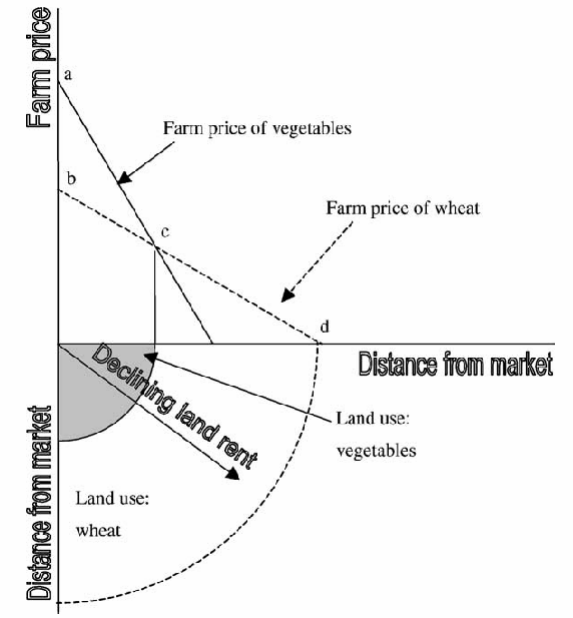

Figure 1.1. The von Thünen model of farm price, land use and land rent (Nelson 2002)

Contents of this report

This report provides an overview of the statistical and empirical techniques used in the spatial analysis of LUCC. The analysis of land cover patterns and changes is discussed, as well as changes in land use and farming systems. Several livestock system examples are included as representative of farming system change that may not be manifested in LUCC, but to which the same analytical methods can be applied. Chapter 2 provides an overview of data types that are used for the analysis of spatial patterns and problems related to the collection and processing of these data. Chapter 3 describes and discusses different methods that can be used for the analysis of spatial patterns. The main focus is on techniques for data reduction, structure detection and regression analysis, but Bayesian statistics, multilevel statistics and neural networks are also discussed. All methods are illustrated with examples from peer-reviewed articles of LUCC studies that adopted the techniques. Chapter 4 discusses several topics that are frequently encountered in spatial modelling studies, such as multicollinearity and spatial autocorrelation. The final chapter gives a short overview of the challenges of spatial analysis of land use patterns.

2 Data sets for spatial analysis of land use and cover change

2.1 Introduction

This chapter provides a short overview of the different types of data that are used in spatial analysis of LUCC. Essential to the discussion of spatial data is the definition of spatial scale. Scale is the spatial, temporal, quantitative or analytical dimension used to measure and study any phenomenon (Gibson et al. 2000). Definitions of other key terms related to scale are listed in Table 2.1. Issues related to scale are of fundamental importance for LUCC studies (Veldkamp et al. 2001; Verburg and Chen 2000; Walsh et al. 2001; McConnell and Moran 2001). Choices concerning scale, extent and resolution critically affect the type of patterns that will be observed, because patterns that appear at one level of resolution or extent may be lost at lower or higher levels. In addition to the explanations derived for phenomena at any one level, scale is central to attempts to generalize from one level or scale to another, i.e. upscaling or downscaling (Gibson et al. 2000).
As a result of the many interacting processes land use systems rarely or never produce a single scale that can be regarded as correct or optimal for measurement and prediction. Although for a specific data set an optimal scale of analysis might exist where predictability is highest, unfortunately this is not consistent through analysis. Therefore, it might be better not to use a priori scales of observation, but rather extract the observational scales from a careful analysis of the data.

Table 2.1. Definitions of key terms related to the concept of scales following Gibson et al. (2000)

Term	Definition
Scale	The spatial, temporal, quantitative or analytical dimensions used to measure and study any phenomenon.
Extent	The size of the spatial, temporal, quantitative or analytical dimensions of a scale.
Resolution (grain)	The precision used in measurement.
Hierarchy	A conceptually or causally linked system of grouping objects or processes along an analytical scale.
Levels	The units of analysis that are located at the same position on a scale. Many conceptual scales contain levels that are ordered hierarchically, but not all levels are linked to one another in a hierarchical system.
Absolute scale	The distance, time or quantity measured on an objectively calibrated measurement device.
Relative scale	A transformation of an absolute scale to one that describes the functional relationship of one object or process to another (e.g. the relative distance between two locations based on the time required by an organism to move between them).

Hierarchies
A fundamental difference exists between functional levels, e.g. household or plot, and spatial units, e.g. polygons or pixels (Table 2.2). Functional levels cannot always directly be linked to spatial units of analysis. Both have a different hierarchy: for example, a household has an influence at the plot, field and farm level and sometimes even at the watershed level. Scale is therefore a continuum, because it moves between discrete yet often unknown or not recognized levels of organization.

The most obvious conclusion from a quick scan through quantitative LUCC studies is that most studies opt for one level of analysis exclusively (Verburg, Schot et al. 2004). Often, this choice is based on arbitrary, subjective reasons or the disciplinary background of the researcher (Gibson et al. 2000; Watson 1978). Researchers in the social sciences have a long tradition of studying individual behaviour at the human-environment interface, making the individual the level of

analysis. Others even analyse the different interacting processes that lead to decision making by individuals, e.g. those involved with social psychology (Wester-Herber 2004).
Rooted in the natural sciences rather than the social, geographers and ecologists have focused on land cover and land use at the macro scale, spatially explicated through remote sensing and geographic information systems (GIS), and using properties of social organization and the environment at the meso and macro scales in order to identify factors connected to observed land use patterns. Doing this they focus on the system dynamics rather than on the behaviour of the individual components that make up the system. In the social sciences analysis at the meso and macro levels are sparser; examples are macroeconomic studies, and a number of macrosociological analyses.

Table 2.2. Hierarchies of observations

Functional/organizational level	Vector-based observations	Pixel-based observations
Individual	Plot	1 m
Household	Field	10 m
Population	Farm	20 m
Community	Watershed	100 m
Ecosystem	District	1 km
Landscape	Province	5 km
Region	Country	10 km

Spatial and temporal representation of LUCC data
The most familiar data source for LUCC research is maps, often derived from remote sensing information. However, other sources of data are also frequently used. While the concepts of extent and resolution directly apply to map data those are not terms most social scientists use frequently. But, all data implicitly have a resolution and an extent. The resolution is the smallest unit of analysis (individual, household, community) and the extent is the aerial dimension for which it is relevant (village, region, country). Sociologists often use data collected at the level of individual households, either aggregated to administrative regions as part of a census, or collected by specific questionnaires. Questionnaires are especially useful to obtain management-related data, e.g. crop rotations or years under fallow and household-specific conditions that might influence decision making. Questionnaires can also give insight into the driving factors of land use change. Issues related to questionnaires are gender, time coverage and sample size. Depending on the cultural circumstances the outcome of a questionnaire might be different when filled in by a woman instead of a man, who is normally the head of a household. One might therefore consider administering separately the household head and his (or her) spouse to acquire information (Pan et al. 2004).
Although it may be difficult to adequately represent land cover types on a map, farming systems that include livestock are even more difficult to map (Thornton et al. 2003). Livestock cannot directly be identified from remote sensing data, while at the same time livestock are able to move around in the landscape. At small scale some livestock systems such as zero grazing can be allocated relatively easily (point data), but other livestock systems, such as free-ranging cattle, are much more complex, since they cannot be assigned to a specific area. At a global scale Kruska et al. (2003) developed a livestock production system map according to the classification put forward by Seré and Steinfeld (1996). The method was based on agroclimatology (growing period length), land cover and human population density. The classification is based on two main classes: solely livestock systems and mixed livestock systems. Farming systems that are completely based on livestock are subdivided into grassland-based systems, landless monogastric systems (e.g. poultry enterprises in Asia) and landless ruminant systems (e.g. zero grazing in Kenya). The mixed farming systems are subdivided into rain-fed mixed farming systems and irrigated mixed farming systems. This classification was further elaborated by combining it with the agroclimatic breakdown. In mapping the classification use was made of data showing the spread of human population, land cover, length of growing period, distribution of irrigated areas, the Nighttime Lights of the World database and agricultural census data.

Verburg and Van Keulen (1999) used subnational data from agricultural surveys in combination with a simulation model to study the changes in the spread of livestock in China. At more detailed spatial scale Burnsilver et al. (2003) studied the actual movement of livestock throughout the season using Global Positioning System (GPS) tracking to obtain data for spatial analysis of the relation between environmental conditions and livestock. Another source of livestock and wildlife data is aerial surveys, as used in a study by De Leeuw et al. (2001).

From the perspective of survey research designs, many of the interesting research questions have to do with change over time, which in turn pushes us to have temporal depth in the variables of interest. If land cover data are coming from one of the frequently used sensors, such as Landsat TM, it is possible to achieve images from different years or time periods. On the social survey side, however, there are many sampling issues involved in obtaining temporal depth. Most household surveys are carried out only once; to detect changes in land use and driving factors, a survey with a longer timespan should be applied. One issue is whether to use a cross-sectional or longitudinal design. Cross-sectional designs are easier to administer because it is not necessary to follow over time sample households, which might move away from the study area or change in size, composition or character. Moreover, in a longitudinal design, repeated visits to the same household may affect the quality of household responses and, depending on the circumstances, the effect could be positive or negative. On the other hand, cross-sectional designs suffer from lack of comparability between sampled households at two time points, especially if the sample size is not large enough. Walsh et al. (2003) describe a good example of such a longitudinal survey. For a study area in north-east Ecuador a household survey was carried out in 1999, which covered precisely the same geographical sites as a prior 1990 study. According to a two-stage sampling design a sample of 878 households was interviewed.

2.2 Geographic representation of data

2.2.1 Point data

Point objects have a position in space but have no length, which makes them zero-dimensional objects that specify a geometric location. One coordinate pair of XY values specifies such a location. Points are used to locate geographical phenomena at that location on a map or to represent map features too small to be shown as lines or areas at the scale of the map. In LUCC-related studies most point data are based on household surveys. In many older studies no geographic coordinates of these households were noted, which makes it difficult to link these data with geographic features presented in remote sensing or maps. Nowadays, it is common practice for social scientists to georeference the locations of households with a GPS. Through a basic handheld receiver one can determine the X- and Y-coordinates with a precision of a few metres without the use of maps or any other equipment. The coordinates can be used to locate the household on a map or combine the data with other georeferenced information.

An example of a study using point data is provided by Staal, Baltenweck et al. (2002) who demonstrate that GIS-derived measures of spatially differentiated factors can be incorporated into a standard household adoption model, which is based on (georeferenced) household-level data, and can potentially differentiate the multiple impacts of location on choices of agricultural technology. The method is applied to smallholder dairy farming in Kenya. The approach integrates spatially referenced household data (point data) with information derived from digital surfaces and infrastructure maps (field data). The unit of observation is a household, rather than a spatial grid cell or administrative unit.

2.2.2 Polygon data

Within a vector-based geographic representation polygons are continuous two-dimensional objects, which may be homogeneous or divided internally into areas with different characteristics. Each polygon is encoded in the database as a sequence of locations that define the boundaries of each closed area in a specified coordinate system. The attributes of each polygon, such as land cover or soil type, are stored in the database as well. Besides normal polygon data such as soil and land use maps two other types of polygon data are commonly used in LUCC analysis: aggregated household data and census data.

Aggregated household data

This type of polygon data is created by the aggregation of household data to the village level. The approach will be illustrated with an example of Mertens et al. (2000) in the tropical forest zone of East Province, Cameroon. They performed statistical analyses that combine remotely sensed land cover change data and household information at the scale of villages. The household and remote sensing data were therefore aggregated to a common spatial representation at the village level. Village territories were represented by polygons. Two distances were considered in the definition of the boundaries that define the polygon representing the village territories: (i) the spatial extent of each village into the forest area, and (ii) the spatial extent of each village along the road network between each pair of villages. Since the villages are close to each other in the study area, the boundary between two adjacent villages along a road was set at a distance from each village such that the ratio of the distance from the village to that boundary and the population of that village was equal for the two villages. In other words, the spatial extent of the agricultural area of a village along a road was proportional to the population of the village. The boundary of the agricultural area of each village from the road into the forest area was defined at a 4-km distance from each village centroid. This distance is assumed to represent the maximum distance travelled from the village to the agricultural plots, based on a field survey. Different procedures were considered for the aggregation of the household data to village level, depending on the type of variable (i.e. continuous, categorical and binary) and the desired information (Table 2.3). The sum and the mean functions were applied for continuous variables. The mode, median and frequency of occurrence functions were applied for categorical variables.

Table 2.3. Example of methods for spatial aggregation of household survey variables to the village level (based on Mertens et al. 2000)

Types of variables	Values ranges	Method of aggregation
Continuous (type A) *(e.g. head of household age, fallow period)*	0 to n	Mean
Continuous (type B) *(e.g. number of created plots, area of coffee cultivated, production of cocoa, number of workers)*	0 to n	Sum (divided by the sampling coefficient)
Categorical (type A) *Categories (e.g. level of education, matrimonial status, main activity, origin of migrants, crop preferences)*	1 to n	Mode
Categorical (type B) *Binary variables (e.g. creation of plots, abandonment of plots, increased number of plots, native villager/migrant)*	Binary	Mode and frequency of occurrence
Categorical (type C) *State of a variable compared to the previous period (e.g. number of workers, area cultivated for plantain, production of food crops)*	Higher, equal or lower	Median and frequency of occurrence

Census data

Census data are mostly presented at the level of administrative units, which can be represented by a polygon. This can be the province, district, county, village or census block level, depending on the country, administrative organization and the type of data. Census blocks consist of polygons defined by relatively fixed features on the land (e.g. roads, rivers, railroad tracks, lake shores) and other features such as municipality boundaries, property lines and short, imaginary extensions of streets and roads. While having the disadvantage of being polygons of widely varying size and shape, they often represent the highest spatial resolution of any data published from a regular census (Radeloff et al. 2000). Population censuses attempt to enumerate all individuals within a census unit, and typically they are household based, which means that households are enumerated, and then information is collected on all individuals living in each household. There are three problems with censuses from the perspective of linking households to the land for which they are major decision makers. First, censuses are conducted infrequently, with once a decade being the most common periodicity. Second, in virtually all countries, household-level census data are considered confidential and not released to LUCC researchers. Rather, the data are aggregated to a high enough level (minimally census blocks) to protect confidentiality and then released, negating the possibility of examining individual households. Third, with the exception of some agricultural censuses, the typical census does not have links to the land the household owns or uses (Rindfuss, Walsh et al. 2003). Furthermore, the administrative units at which census data are presented often do not correspond to biogeophysical units (e.g. soil mapping units) or the units used to represent land use. Census data can be used for direct analysis as polygons, but might also be transformed to a common grid format for raster-based modelling.

A typical example of the use of census data is provided by Wood and Skole (1998), who describe a large-scale study on deforestation of the Brazilian Amazon region in which satellite-based estimates of deforestation were combined with census estimates of demographic and economic structure. A data set was constructed by merging, in a GIS, the satellite- and census-based variables for each of the municipios that comprise the Amazon region in Brazil. Indicators of demographic and economic structure were derived from two population and agricultural censuses. The satellite data were aggregated for each municipio polygon to correspond with the census data. Based on this data set linear regression models were constructed to determine the relation between deforestation and a set of demographic and economic indicators.

Another example is from Cardille and Foley (2003), who used a hybrid method combining remote sensing and census data. They used agricultural census data from Peru and Brazil to create municipio-level maps of three major categories of agricultural land use activity: cropland, natural pasture and planted pasture. For the cropland category basic census variables were available, such as the planted area of annual and perennial crops, temporarily fallow area associated with cropland and land harvested but not yet replanted. The notion of a systematic relationship between satellite-derived land cover categories and important ground-censused agricultural land use activities suggests a statistical blending that could distribute agricultural land use activity into those areas probably being used for cropland and pasture. A regression tree was generated that statistically linked the census data and land cover map. Unlike a simple renaming, e.g. the wooded grassland category to be planted pasture, the regression tree-based technique determined the statistical relationship of agricultural density and the fraction of each category within reporting polygon-based census units.

2.2.3 Raster data

Raster data are an abstraction of the real world whereby spatial data are represented by a matrix of cells or pixels, with spatial position implicit in the ordering of the pixels. With a raster data model, spatial data are not continuous but divided into discrete units. This makes raster data

particularly suitable for certain types of spatial analysis and modelling. Many models use raster data for the simulation of LUCC because grid cells are all equally sized, well-defined units, which makes modelling easy. Furthermore, most remote sensing-derived data, e.g. satellite images, are based on pixels, which provide the basis for raster data. Also other data types, e.g. digital elevation models, are nearly always in raster format. Vector data can be easily converted to raster data. Finally, for LUCC analysis there is no common spatial unit at which land use information and data on the socio-economic and biophysical conditions can be jointly presented, since all features have their own specific units of spatial representation (e.g. soil mapping units or administrative units) that do not overlap. The raster format is a common format by which all data can be represented.

Two different representations of land use and land cover by raster data can be distinguished (Verburg et al. 2002). In general, for study areas with a large extent the spatial resolution of analysis is coarse. This is a consequence of the impossibility of acquiring data for land use and all driving factors at finer spatial resolutions for large areas of land and the computational constraints of very large data sets. A coarse spatial resolution requires a different data representation than the common representation for data with a fine spatial resolution. In fine resolution, grid-based approaches, land use is defined by the most dominant land use type within the pixel. However, such a data representation would lead to large biases in the land use distribution at coarse scales as some class proportions will diminish and other will increase with scale, depending on the spatial and probability distributions of the cover types (Moody and Woodcock 1994). In applications at the national or continental level land use is represented by designating the relative cover of each land use type in each pixel, e.g. a pixel can contain 30% cultivated land, 40% grassland and 30% forest. This data representation is directly related to the information contained in census and other sources of aggregated data: for each administrative unit, census data denote the number of hectares devoted to different land use types. Furthermore, this data representation is not sensitive to aggregation errors. When studying areas with a relatively small spatial extent, land use data are often based on land use maps or remote sensing images that denote land use types respectively by homogeneous polygons or classified pixels. When converted to a raster format this results in only one dominant land use type, occupying one unit of analysis. The validity of this data representation depends on the patchiness of the landscape and the pixel size chosen. Most subnational land use studies use this representation of land use with pixel sizes varying between a few metres up to about 1x1 km (Verburg et al. 2002).

The two different data representations are shown in Figure 2.1. The CLUE approach (Verburg et al. 1999) is an example of a LUCC method that uses subpixel information on land use in its countrywide and continental applications. Other examples of LUCC studies that use mixed pixels are the land use scanner (Hilferink and Rietveld 1999) and the ATEAM land use model (Rounsevell et al. 2005). Most LUCC analysis is, however, based on the representation of land use by its dominant land use. Both data representations can theoretically also be applied on polygon data.

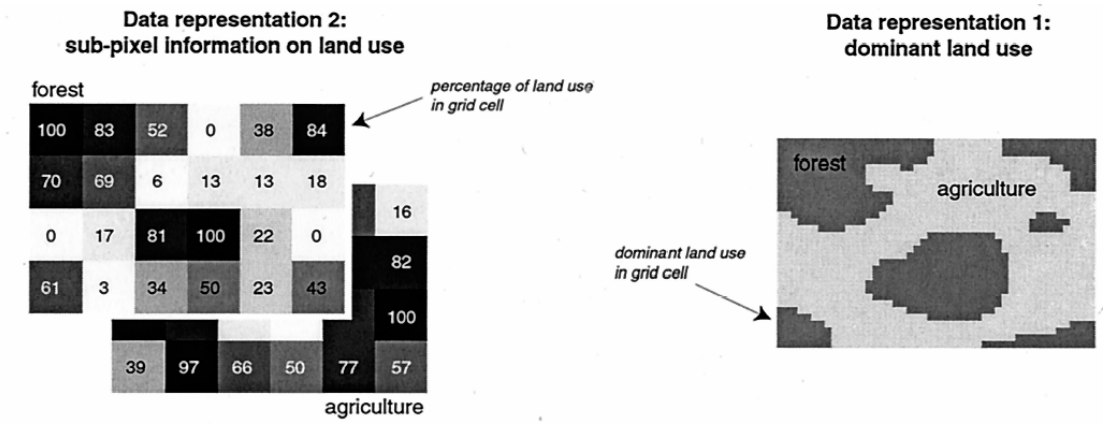

Figure 2.1. Example of different data representations of the same landscape (Verburg et al. 2002)

Remote sensing data

Among the most frequently used raster-based data in land cover analysis are remote sensing data. Therefore they deserve some extra attention. Remotely sensed data include information gathered digitally by aerial photography and satellites. Solar radiation is reflected from the surface of the earth, e.g. from soil, water, vegetation and buildings, to sensors that measure the intensity of different frequencies. Each type of surface reflects or absorbs different frequencies. Hence by a careful choice of sensor type it is possible to make inferences about what is on the surface of the earth. The most commonly used satellites and sensors, with their characteristics, are listed in Table 2.4.

Table 2.4. Selected satellites and their characteristics (Nelson and Geoghegan 2002)

Satellite/sensor	Repeat rate	Area of image	Pixel dimension	Frequencies
Landsat/MSS	16–18 days	150x150 km	80 m	Green, red, infrared
Landsat/TM	16–18 days	150x150 km	30/15 m	Blue, green, red, near infrared, mid infrared, thermal
AVHRR	0.5 day	800x800 km	1.1 km	Green, red infrared, lower frequencies
IKONOS	1–3 days	Variable	1–4 m	Green, red, infrared
Quickbird	1–3 days	16x16 km	0.6 m	Blue, green, red, near infrared
SPOT	3–6 days	60x60 km	10–20 m	Green, red, near infrared

All remote sensing techniques primarily deliver images of land cover and not of land use. Land use is characterized by the arrangements, activities and inputs people undertake in a certain land cover type to produce, change or maintain it, while land cover is the observed (bio)physical cover on the earth's surface (FAO 1997). Land use defined in this way establishes a direct link between land cover and the actions of people in their environment. Land use typically causes distinctive patterns of land cover. To some degree inferences about land use can be made from these patterns of land cover, but for a full land use classification ground information is essential.

Each land cover type has different spectral characteristics, absorbing some frequencies of light and reflecting others. With an understanding of the reflectance characteristics and some ground observations it is possible to use remotely sensed data to make inferences about the type of land cover (and with some additional uncertainty land use). There are two common ways in which this is done for agriculture and related natural resource questions: vegetative indices and land cover clustering and classification techniques (Nelson and Geoghegan 2002).

The normalized difference vegetation index (NDVI) uses multispectral scanner (or equivalent) bands 2 (0.58–0.68 μm) and 4 (0.725–1.1 μm) to measure the absorption and reflectance of solar radiation. In most cases, NDVI is correlated with photosynthesis. Because photosynthesis occurs in the green parts of plant material the NDVI is normally used to estimate green vegetation.

Clustering techniques operate by assuming that pixels with similar spectral characteristics have the same land cover. Two general approaches are used: unsupervised and supervised classification. With unsupervised classification only spectral information is used in the analysis (no field observations are used). One or more algorithms are used to find locations with similar spectral (and sometimes other) characteristics. A more widely used set of algorithms involves distance measures. The general approach is to start with an initial sample, choose clusters so within-cluster distance is minimized and across-cluster distance is maximized, then assume a normal distribution and use a maximum likelihood estimator to assign remaining pixels to clusters. Supervised classification involves the use of ground-control points, called ground-truth, where the true land cover is identified. These locations are then used to guide the classification process, say by identifying all locations whose combinations of characteristics are within a certain spectral distance from those of the ground-truth points (Nelson and Geoghegan 2002).

2.3 Linking different data representations

The previous sections described and discussed three types of data representation. The choice of which data representation to use is based on three key questions. What is the objective of the study, what is the level of explanation and what type of source data is available? Table 2.5 gives a summary of some possible answers to these questions for the three data representations. Besides choosing a certain data representation, one might also be interested in a combination of different data types, which leads to the question of how to link those data representations. This often coincides with the issue of how to link people (social data) and pixels (spatial data) (Rindfuss et al. 2004). In this section some issues involved with the linking of different data types will be discussed.

Table 2.5. Examples of objectives, levels of explanation and source data for the choice of data representation

	Point data	**Polygon data**	**Grid data**
Objective	Household influence on LUCC Adoption of technologies by households	Understanding causes of LUCC	Spatial modelling Analysis of LUCC patterns
Level of explanation	Household Plot	Village Regional Country	Regional Country
Source data	Household survey GPS coordinates	Village data Census data Aggregated household data	Remote sensing data Maps with different units

A key to linking household and GIS data is to correctly define the spatial observation unit with respect to decision making. In other words: do we link human activities to land or do we link land to people? (Geoghegan et al. 2001). Administrative units or grid cells are not individual agents, but aggregates of them. Inferences as to outcomes in such units require simplifying assumptions about homogeneity of the decision makers and the dynamics comprising that aggregate. Proper inference of micro-level spatial behaviour is therefore more appropriately based on survey samples of individual agents, under the general principle of matching the spatial scale of the decision process and the scale at which measurement is carried out (Anselin 2002). This consideration should be taken into account by linking spatial measures to the perceived real decision makers, thus matching the spatial and behavioural units.

The linkage of socio-economic household-level data and remote sensing data at the household level in general captures best the actual level of decision making, except in rare cases where policy

decisions at an administrative unit level are a more important determinant. Linked household-level data also allow a representation of the heterogeneity between households operating in a spatial unit. However, linking remote sensing and socio-economic data at the household level comes at a certain cost as, depending on the detail of analysis required, it generally requires georeferencing every plot of the interviewed households. This operation is labour intensive, because one should travel with the household to every plot to collect GPS coordinates or a detailed map should be available on which the household can identify its parcels (Lambin 2003). In this respect also the organization of the dwelling units is of importance. In general dwelling units within villages are organized in a cluster surrounded by agricultural land, where the typical household uses several parcels of land (Figure 2.2 B). But also other non-clustered patterns are found, as in the Amazon (Figure 2.2 A). Cadastral information could be useful but only reflects ownership and not the actual land use. Furthermore, cadastral information is not available in many regions. Often administrative boundaries of villages (arranged in nuclear patterns or otherwise) are lacking or do not effectively describe the 'functional' use of land at the household level and the geographic distribution of households across the landscape (Rindfuss, Prasartkul et al. 2003). Overmars and Verburg (2005) provide an alternative to the labour-intensive methods of actively linking household to parcels by georeferencing all individual plots. They did not link households to their plots physically, but obtained plot and field characteristics during a household survey as part of a hierarchically structured questionnaire in which household data as well as plot and field data are registered. The problem with this approach is that it is rather subjective and the field data are subject to the perception of the farmer.

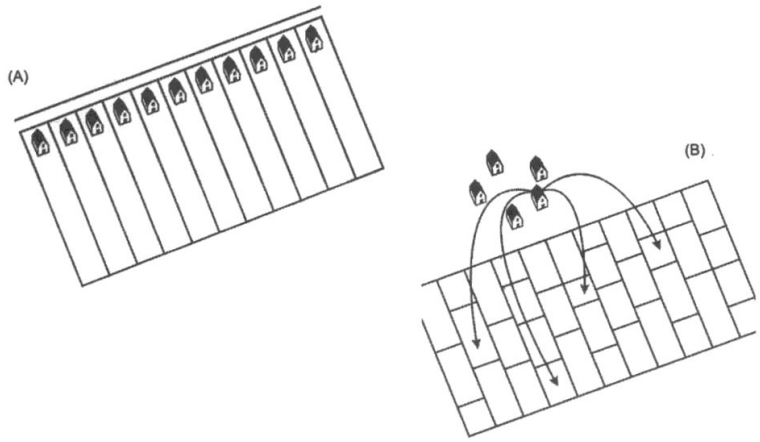

Figure 2.2. Illustration of households in non-clustered villages (A) and clustered villages (B) (Rindfuss, Prasartkul et al. 2003)

In a number of environments, the spatial representation of parcels may be hampered because the size of plots associated with a single household may be below the spatial resolution of remote sensing systems. In such situations, digital or analogue aircraft data might be considered, whereby the user can set the required minimum mapping unit for inter- and intra-plot mapping. Recent remote sensing systems having higher spatial resolutions are also online to render detailed (to approximately 1x1 metre spatial resolutions) land cover information. However, these high spatial resolution sensors have high data volumes and costs associated with them. These high data volumes might, in turn, create data management and budget issues for the researcher, as well as design considerations about how best to use the high spatial resolution data. One can use a continuous data set for broad area mapping or for only a subset or sampled region, using models to extend the effects to broader areas. Not only plots and pixels, but all digital spatial data involve mismatches with the reality, especially aggregated data. For example, the landscape unit river terrace, identified at a 100 metre resolution map, has to be aggregated to the more general term river valley when used at a 1 km resolution.

3 Empirical analysis techniques

This chapter discusses several multivariate techniques that are frequently used in LUCC and farming systems change research, including livestock systems and other examples of technology adoption. A number of methods discussed have not yet been frequently applied in LUCC research but are included because of the potential value for LUCC applications. We do not intend to cover all potentially interesting methods. A much larger selection of statistical techniques could be identified that have a potential value, but are outside the scope of this report. The first section describes four methods for exploratory spatial data analysis in terms of data reduction and structure detection, namely principal component analysis, factor analysis, canonical correlation analysis and cluster analysis. In section 3.2 different regression analysis methods are discussed: linear regression, (nested) logistic regression, multinomial regression, ordered logit, Tobit analysis and simultaneous regression. The last three sections respectively describe Bayesian statistics, for studies where prior knowledge of observations is available; multilevel statistics, for studies that involve different hierarchical levels; and artificial neural networks, useful for cases where a training procedure for pattern prediction or classification is applicable. Figure 3.1 shows an ordering for the different empirical analysis techniques for analysis of spatial patterns of LUCC. This diagram is not intended as a 'simple' decision tree for selecting the appropriate method for a specific case. The diversity in data structures, research questions and case study-specific conditions make a careful analysis of the requirements of the method necessary for each specific case study. Furthermore, different methods can be used at the same time or sequentially within one analysis to better explore the data structure.

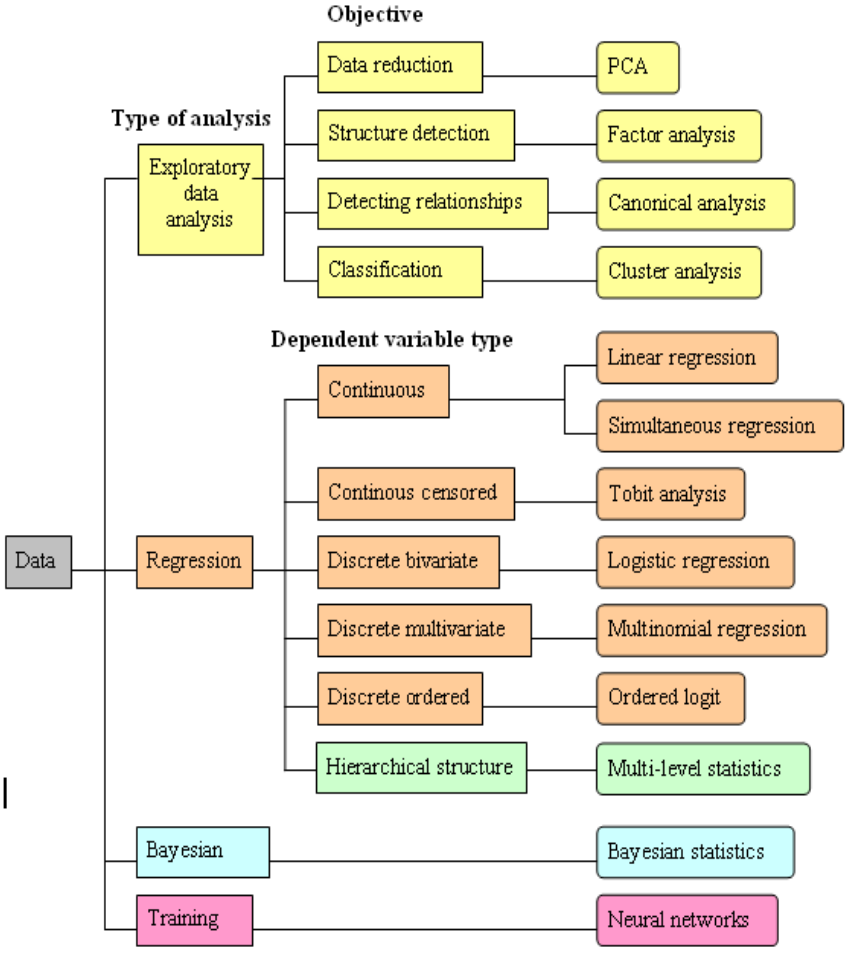

Figure 3.1. Classification of empirical analysis techniques for LUCC based on objective and data structure

3.1 Techniques for exploratory data analysis

The main uses of exploratory data analysis techniques are related to data reduction and structure detection. These methods aim (i) to reduce the number of variables; (ii) to describe the underlying structure between variables in the data; and (iii) to classify variables into groups. Factor analysis and principal component analysis (PCA) are applied as data reduction or structure detection methods and cluster analysis for classification. This is useful in LUCC analysis because land use change is often assumed to be influenced by a large set of driving and conditioning factors. PCA and factor analysis are suited to exploration of the structure of interrelationships between these different driving factors. Furthermore, the methods can also be used to characterize farming or land use systems based on a number of indicators.

3.1.1 Factor analysis

Factor analysis attempts to identify underlying variables, or factors, that explain the pattern of correlations within a set of observed variables. Factor analysis is often used in data reduction to identify a small number of factors that explain most of the variance observed in a much larger number of manifest variables. Factor analysis can also be used to generate hypotheses regarding mechanisms or to screen variables for subsequent analysis (e.g. to identify collinearity prior to performing a linear regression analysis).

Factor analysis is concerned with the internal relationships of a set of variables and is aimed at constructing a set of factors (hypothetical unobserved variables) from a set of observable variables. The factor analysis model specifies that variables are determined by common factors (the factors estimated by the model) and unique factors (which do not overlap between observed variables). The computed estimates are based on the assumption that all unique factors are uncorrelated with each other and with the common factors (SPSS 2000).
The factors are common when they contribute to the variance for at least two observed variables or unique when their contribution is only towards one variable. A correlation matrix for a set of observations (R-factor analysis) is prepared or, less frequently, for individuals for a set of variables (Q-factor analysis). Then the initial factors are extracted, which can be based on defined factors (principal component analysis) or on inferred factors (common factor analysis). As the exact configuration of the factor structure is not unique, one factor solution can be transformed into another one or rotated to a terminal solution. This can achieve simpler and more meaningful factor patterns, instead of the highly complex extracted factors that are related to many of the variables rather than to just a few (Comrey and Lee 1992).

Prior to factor analysis the collected information on the various variables can be processed. The variables in factor analysis should be quantitative at the interval or ratio level. Categorical data, e.g. ethnicity or soil type, are not suitable for factor analysis. The data should have a bivariate normal distribution for each pair of variables, and observations should be independent. Those variables that do not show variability can be discarded. First, any variable that makes little contribution, in terms of its variability, to the measure of distance being used to form clusters can be discarded. This is normally evaluated through the coefficient of variation. It is established on an a priori basis that the variables with a coefficient of variation of less than 50% are normally not considered (Köbrich et al. 2003). Second, some variables may not be relevant to the typification required for the purposes of a particular study and can therefore be discarded, even though the typology obtained initially is consistent with observations. Thus one has to assess if the information imparted by a variable is consistent with the research objectives. Third, highly correlated variables can be eliminated, as an uncritical use of such variables.

The usual interpretation of the factors is that they 'explain' the correlations that have been discovered among the original variables and that these factors are real factors in nature. Unfortunately, factor analysis encourages subjective overinterpretation of the data (James and McCulloch 1990).

Example
Veldkamp and Fresco (1997) used factor analysis in a LUCC study, in which Costa Rican land use and land cover were investigated at six different scales. Spatial distributions of potential biophysical and LUCC drivers were statistically related to the distribution of pastures, arable lands, permanent crops, and natural and secondary vegetation. The factor analysis demonstrated that factor contributions and compositions change with scale, confirming spatial scale dependence in the structure of the spatial data. The total variance in the data set could be described by four significant factors for all scales, describing between 68% and 81% of the total variance (Table 3.1).

Table 3.1. Example of a factor analysis (Veldkamp and Fresco 1997)

Spatial resolution (km)	**7.5x7.5**	**15x15**	**22.5x22.5**	**30x30**	**37.5x37.5**	**45x45**
Explained variance						
Factor 1:	28.2	27.9	30.5	31.6	37.3	36.1
Factor 2:	22.5	19.7	23.5	24.7	20.8	18.6
Factor 3:	11.6	10.7	12.0	11.6	13.7	14.7
Factor 4:	11.2	9.9	9.8	9.5	9.2	10.3
Total:	73.5	68.2	75.8	77.4	81.0	79.8
Factor composition *						
Factor 1:	PER	PER	PER	PER	PER	PER
	RUR	RUR	RUR	RUR	ARA	RUR
	URB	ALF	URB	ALF	RUR	ALF
	ALF		ALF		ALF	
Factor 2:	ARA	ARA	SOIL	REL	ARA	SOIL
	-NAT	SEC	PAS	SOIL	-NAT	ARA
	SEC		-NAT	-ALT	SEC	-NAT
					URB	SEC
Factor 3:	SOIL	REL	ARA	ARA	-REL	REL
	PAS	ALT	-NAT	SEC	ALT	-ALT
	-NAT		SEC			
Factor 4:	REL	PAS	REL	PAS	SOIL	PAS
	-ALT	NAT	-ALT	NAT	PAS	-NAT
				URB	-NAT	URB

* The factor analysis was made for the following data: altitude (ALT), relief (REL), soil drainage (SOIL), rural population (RUR), urban population (URB), agrarian labour force (ALF), permanent crops (PER), pasture (PAS), arable land (ARA), natural vegetation (NAT) and secondary vegetation/fallow (SEC).

3.1.2 Principal component analysis

Principal component analysis (PCA) reduces the dimensions of a single group of data by producing a smaller number of abstract variables. The combination of two correlated variables into one factor illustrates the basic idea of PCA. With multiple variables the computations become more involved, but the basic principle of expressing two or more variables by a single factor remains the same. Basically, the extraction of principal components amounts to a variance maximizing (varimax) rotation of the original variable space. For example, in a scatterplot we can think of the regression line as the original X axis, rotated so that it approximates the regression line. This type of rotation is called variance maximizing because the criterion for (goal of) the rotation is to maximize the variance (variability) of the 'new' variable (factor), while minimizing the variance around the new variable. When there are more than two variables, we can think of them as defining a 'space', just as two variables defined a plane. Thus, when we have three variables, we could plot a three-dimensional scatterplot, and again we could fit a plane through

the data. With more than three variables it becomes impossible to illustrate the points in a scatterplot; however, the logic of rotating the axes so as to maximize the variance of the new factor remains the same.

In PCA, after the first factor has been extracted, we continue and define another factor that maximizes the remaining variability, and so on. In this manner, consecutive factors are extracted. Because each consecutive factor is defined to maximize the variability that is not captured by the preceding factor, consecutive factors are independent of each other. Put another way, consecutive factors are uncorrelated or orthogonal to each other (StatSoft 2003).

PCA is very useful to detect the structure of data that describe the driving factors of LUCC. However, PCA has also been used in the processing of LUCC data, i.e. it is a frequently used technique in the classification of remote sensing images. Principal component analysis of a set of p images generally aims to summarize – and hopefully improve the interpretation of – the available information by a few new images that are orthogonal linear combinations of the original images. The first principal component 'explains' the largest part of the total variance included in all p images, the second component the second largest part, etc. The larger the correlation between the p images, the fewer components are required to explain a large part of the total variance of the original images. For this reason, PCA is a common method to improve the interpretation and classification of multispectral or multitemporal satellite images (Richards 1986).

Factor analysis versus principal component analysis

Basic computational similarities lead to confusion when distinguishing between PCA and factor analysis. The defining characteristic that distinguishes between the two factor analytic models is that in PCA it is assumed that all variability in a variable should be used in the analysis, while in factor analysis only the variability in a variable that it has in common with the other variables is used. Another difference between PCA and factor analysis is how the communalities are computed, that is the fraction of each variable's variance that is explained by the total of the extracted factors. Communality represents the extent of overlap between the extracted factors and the variable and it equals the sum of squares of the variable's loadings across factors (Comrey and Lee 1992). As PCA is based on statistical variance, the first chosen factor accounts for most of the variance in the data. The second is chosen in the same way but it has to be orthogonal to the first. The last factor explains all the residual variance (Kim 1970). Common factor analysis is a covariance or correlation-oriented method based on the assumption that each variable is influenced by a set of shared or common factors that determine the correlation between variables. The implied expectation is that the number of common factors will account for all the observed relations and that such factors will be less than the number of variables (Lawley and Maxwell 1971). In common factor analyses the correlation matrix is transformed before undertaking factor analysis (Kim 1970). In most cases, these two methods yield very similar results. However, PCA is often preferred as a method for data reduction, while factor analysis is often preferred when the goal of the analysis is to detect structure.

Determining how many factors should be retained is a problem, as with real data the actual number that merit retention is often considerably smaller than the number of variables. One test searches for a point where there is a break in the Eigenvalues, that is, the variation in the original group of variables, which is accounted for by a particular factor. As factors are extracted from large to small, their Eigenvalues are also decreasing. When they are plotted, a straight line can be drawn through the latter smaller values. The earlier, larger values will fall above the straight line. It is proposed that the number of factors to be retained is at the point where the last small factor is above the line, giving an indication of how many factors there are (Comrey and Lee 1992). Another test defines a threshold level for the residual correlation, beyond which it would be unnecessary to continue extracting, as any new factor would have very small loadings. A common rule is to extract all the factors with Eigenvalues of 1.0 or more (Kaiser's rule). Whatever rule is

used, it must be kept in mind that it is better to err on the side of extracting too many factors rather than too few, as the idea is to extract enough factors to be relatively certain that no more factors of any importance were discarded (Comrey and Lee 1992).

Examples
Köbrich et al. (2003) used PCA as a data reduction technique in a procedure for the typification of farming systems, to be used for the reconstruction of representative farm models. The farming system typification was applied for a region in the coastal mountains of Chile. A data set of 67 farmers with 25 different farming system variables (e.g. actual land use, income, available labour, livestock numbers) was collected. The data set was first analysed on missing data, variation, relevance and presence of correlation. The result of this analysis was that 14 variables were discarded and only 11 variables were kept for analysis. The high level of correlation between the variables means that a lot of information is redundant, which confirms the view that typification surveys should contain relatively few questions but many observations (Escobar and Berdegué 1990). The 11 variables were used in the principal component analysis, which resulted in seven factors that explained 85% of the total observed variation.
Another example is a study by Hary et al. (1996), who used PCA to identify a structure behind the occurrence of rangeland degradation in Kenya. Duchateau et al. (1997) used PCA analysis in an effective way to reduce the number of independent variables prior to regression analysis. In a study for Zimbabwe they tried to relate incidence of a livestock disease to a set of environmental variables. Many of the environmental variables are correlated, causing multicollinearity in the regression analysis. PCA analysis and varimax rotation of the principal components were first used to select a reduced number of variables to be taken into account in the regression analysis.

3.1.3 Canonical correlation analysis

Canonical correlation is a multivariate technique that has the same computational basis as factor analysis, but in its concept and objectives it is closely related to multiple regression. Multiple regression is concerned with the relationship between a single dependent variable Y and a set of predictor variables X_1, X_2, …, X_m. An extension of this concern is the relationship(s) between a set of Y variables and a second set of X variables measured on the same objects. These relationships may be investigated by finding linear combinations of the X and Y variables that give the highest correlation between the two sets. Such correlations are called canonical correlations and the linear combinations are called canonical variables. In effect, the set of X variables is converted into a single new variable and the set of Y variables into another single new variable. Then the correlation between these new variables is determined (Davis 1986).

The correlation coefficients can be interpreted as the square root of the Eigenvalues. Because the correlations pertain to the canonical variates, they are called canonical correlations. Like the Eigenvalues, the correlations between successively extracted canonical variates are smaller and smaller. Therefore, as an overall index of the canonical correlation between two sets of variables, it is customary to report the largest correlation, that is, the one for the first root. However, the other canonical variates can also be correlated in a meaningful and interpretable manner (StatSoft 2003).

This statistical method is particularly appropriate when the dependent variables themselves are correlated with each other. In such cases, canonical correlation analysis can uncover complex relationships that reflect the structure between predictor and dependent variables. In a study of Hoshino (1996), on land use in Japan, the predictor set was the natural and socio-economic conditions and the dependent set the percentages of the four major land use categories (Figure 3.2). Close relationships among different kinds of land use are normally expected, so in this case the application of canonical correlation analysis was very appropriate.

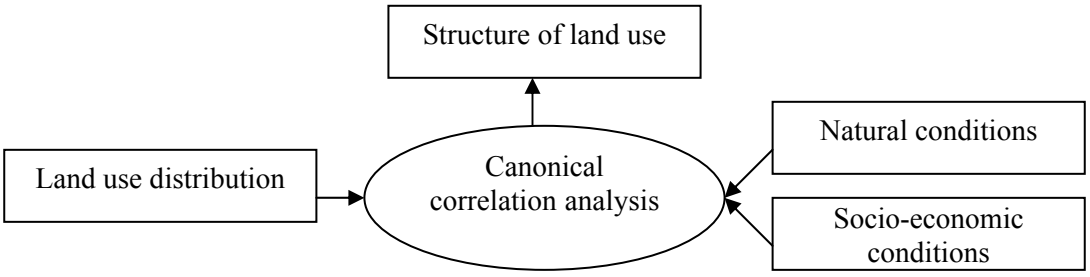

Figure 3.2. Framework of analysis for land use distribution (after Hoshino 1996)

Canonical correlation analysis versus factor analysis
Canonical correlation and factor analysis both create latent variables (variates) based on a linear combination of measured variables, but factor analysis is not usually focused on the correlation of these variates. In fact, normally the factors are uncorrelated in factor analysis. Factor analysis is a non-dependent procedure, whereas canonical correlation can be conceptualized in terms of an independent and a dependent set of variables. Variates are rarely rotated in canonical correlation, whereas rotation of factors is the norm in factor analysis.

Examples
Walsh et al. (2001) used canonical correlation analysis to examine the relationships between a group of population and environment variables as a consequence of variation in the scale of observation. The analysis was made for a case study in Thailand. To examine possible scale dependence, canonical coefficients were derived to relate biophysical variables to environmental axes and social variables to population axes. The results indicated that the relationship between the population and environmental variables, as a group, were scale dependent (Table 3.2). This meant that the numerical relationships as well as the geometric pattern of the relationships visualized through plots of the linkages between variables and defined axes of social and biophysical variables vary with scale.
Another example of a study that has used canonical correlation analysis is an analysis of the relationship between land cover change trajectories and environmental variables in Hesse, Germany (Hietel et al. 2004).

Table 3.2. Standardized canonical coefficients for environmental variables at two scales (Walsh et al. 2001)

Scale	Variable *	Axis 1	Axis 2	Axis 3
30 m	NDVI	0.220	*0.824*	0.132
	Elevation	*0.945*	-0.592	-0.059
	Slope	0.035	0.378	0.768
	Soil wetness	0.090	-0.373	*1.062*
1050 m	NDVI	*0.734*	-0.544	-0.384
	Elevation	0.092	*1.276*	-0.940
	Slope	0.484	-0.260	*1.544*
	Soil wetness	0.045	0.484	0.252

* The associated variables with the different environmental axes are indicated in italics.

3.1.4 Cluster analysis

Cluster analysis is used to classify observations by computing the similarity between any pair of observations though a distance coefficient (Sokal 1977). For LUCC research, cluster analysis is useful to group similar land use types or farming systems when a reduction of classes for further analysis is aimed at. Two main types of cluster analysis exist: K-means cluster analysis and hierarchical cluster analysis.

K-means cluster analysis

This procedure attempts to identify relatively homogeneous groups of cases based on selected characteristics, using an algorithm that can handle large numbers of cases. However, the algorithm requires specifying the number of clusters. Variables should be quantitative at the interval or ratio level. For binary variables the hierarchical cluster analysis procedure should be used. Distances are computed using simple Euclidean distance. Scaling of variables is an important consideration; if your variables are measured on different scales (for example, one variable is expressed in dollars and another is expressed in metres) the results may be misleading. In such cases, consider standardizing the variables before performing the K-means cluster analysis. The procedure assumes that the appropriate number of clusters has been selected and that all relevant variables are included. Cluster analysis produces clusters whether or not natural groupings exist.

Hierarchical cluster analysis

This procedure attempts to identify relatively homogeneous groups of cases (or variables) based on selected characteristics, using an algorithm that starts with each case (or variable) in a separate cluster and combines clusters until only one is left. It is possible to analyse raw variables or choose from a variety of standard transformations. The variables can be quantitative, binary, or count data. The analysis should include all relevant variables. Omission of influential variables can result in a misleading solution. Because hierarchical cluster analysis is an exploratory method, results should be treated as tentative until they are confirmed with an independent sample. Dendrograms can be used to assess the cohesiveness of the clusters formed and can provide information about the appropriate number of clusters to keep (Figure 3.3).

K-means versus hierarchical cluster analysis

The main advantage of the K-means cluster analysis procedure is that it is much faster than the hierarchical cluster analysis procedure. On the other hand, the hierarchical procedure allows much more flexibility in your cluster analysis: it is possible to use any of a number of distance or similarity measures, including options for binary and count data, and there is no need to specify the number of clusters a priori.

Examples

Köbrich et al. (2003) used hierarchical cluster analysis for the typification of farming systems, to be used for the reconstruction of representative farm models. The farming system typification was applied for a region in the coastal mountains of Chile. A data set of 67 farmers with 25 different farming system variables was collected and reduced with PCA. A hierarchical cluster analysis was applied on the seven retained factors. A dendrogram (Figure 3.3) was constructed to show the sequence by which the observations and clusters were merged. Line C was defined as the cutting line, which resulted in five clusters (farms 1–27, farms 28–50, farms 51–57, farms 58–61 and farms 62–63). Table 3.3 contains the typification of the resulting farming systems, which are the selected clusters of farms.

Table 3.3. Comparison of selected clusters of farms (Köbrich et al. 2003)

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Farmer on farm	One year	One year	One year	One year	Half a year
Additional labour	One woman	Marginal	Two men	Two women, two men	Marginal
Farm size	Small	Small	Medium	Large	Small
Herd	Small	Small	Small	Large	Large
Arable/available	57%	86%	67%	20%	-
Crop/arable	31%	17%	28%	30%	-
Sharecropping	Takes in	Takes in	Takes in	-	Gives out

The K-means clustering technique has been used in a study of land cover changes and wildlife decline in the Serengeti-Mara ecosystem in Kenya (Thompson et al. 2002). With K-means clustering households were clustered according to land use strategies. The following variables were used: (i) revenues from shares in tourist facilities and wildlife associations; (ii) acreage under maize cultivation; (iii) acreage under wheat cultivation; and (iv) wage labour. Then the determinants of the land use production choices were assessed, based on variables representing socio-economic status, landscape attributes and accessibility factors. The 278 households were clustered into four land use strategies (Table 3.4). Other examples of land use studies using cluster analysis are Hietel et al. (2004) and Rasul et al. (2004).

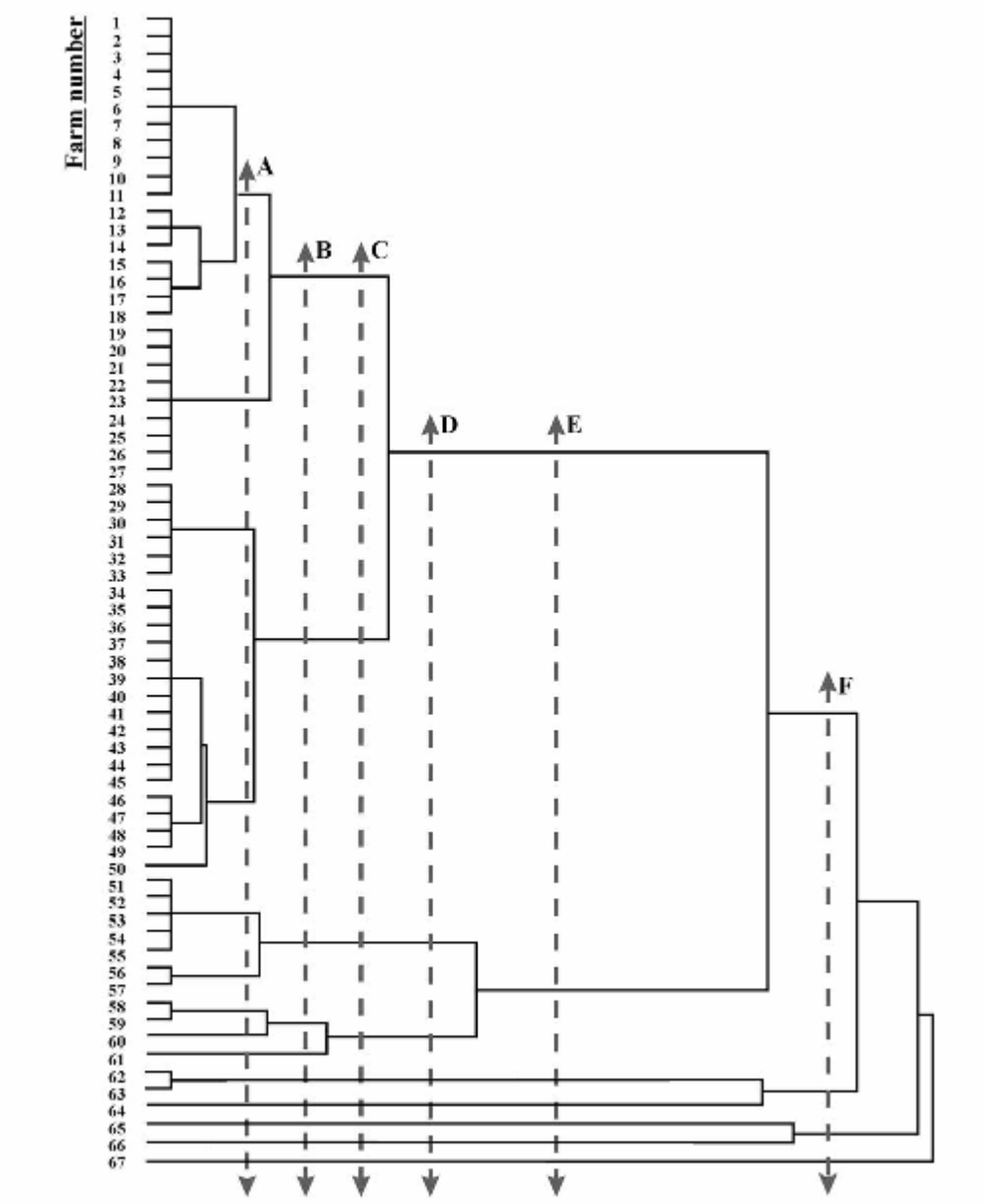

Figure 3.3. Dendrogram showing the full history of cluster construction and six possible cutting lines (Köbrich et al. 2003)

Table 3.4. Final clusters of land use strategies (Thompson et al. 2002)

Land use strategy	Maize	Wheat	Tourism	Wage
Livestock herding with some subsistence cultivation (N = 54)	Some	No	No	Some
Livestock herding with income from wildlife tourism (N = 136)	No	No	Yes	Some
Livestock herding with mechanized agriculture (N = 29)	Yes	Yes	No	Few
Livestock herding with income from tourism and subsistence cultivation (N = 59)	Yes	No	Yes	Some

3.2 Regression analysis

Regression analysis is used to investigate the association of a dependent variable with one or more independent variables. In linear regression a straight line is used to represent the association of the explanatory variables with the dependent variable. More complex methods of regression exist, intended for different types of dependent variables and data structures. Table 3.5 gives an overview of a selection of different regression methods based on the characteristics of the dependent variable. This section will discuss these methods in more detail as well as a number of methods to measure the goodness of fit for regression analysis.

Table 3.5. Summary of regression methods

Regression method	Dependent variable type or data structure
Linear regression	Continuous
Logistic regression	Discrete bivariate
Multinomial regression	Discrete multivariate
Ordered logit/probit	Discrete ordered
Tobit analysis	Censored continuous
Simultaneous regression	Interdependent/simultaneous relations
Multilevel models	Hierarchically organized data sets

3.2.1 Linear regression

Linear regression is a method that estimates the coefficients of a linear equation, involving one or more independent variables, that best predict the value of the dependent variable. Linear regression is a frequently used technique; however, in LUCC modelling, this regression is less popular because linear regression can only be applied for continuous dependent variables. Instead logistic or multinomial regression is used, because land use is normally expressed as a discrete variable. An exception is NDVI data, which range between -1 and 1 and belong therefore to continuous data. Linear regression can also be used to derive input data, e.g. trends of population growth out of census data, or for validation.

In linear regression analysis, it is possible to test whether two variables (or transformed variables to allow for non-linearity) are linearly related and to calculate the strength of the linear relationship if the relationship between the variables can be described by an equation of the form $Y = a + \beta X$. Y is the variable being predicted (the dependent, criterion, outcome or endogenous variable), X is a variable whose values are being used to predict Y (the independent, exogenous or predictor variable), and a and β are population parameters to be estimated. The parameter a, called the intercept, represents the value of Y when $X = 0$. The parameter β represents the change in Y associated with a one-unit increase in X or the slope of the line that provides the best linear estimate of Y from X. In multiple regression, there are several predictor variables. If k denotes the number of independent variables, the equation becomes $Y = a + \beta_1 X_1 + \beta_2 X_2 + \ldots +$

$\beta_k X_k$ and $\beta_1, \beta_2, ..., \beta_k$ are called partial slope coefficients, reflecting the fact that any one of the k predictor variables $X_1, X_2, \ldots, X_k$ provides only a partial explanation or prediction for the value of Y (Menard 2001).

Estimates of the intercepts a and the regression coefficients β are obtained mathematically using the method of ordinary least squares (OLS) estimation. For bivariate regression, the residuals can be visually or geometrically represented by the vertical distance between each point in a bivariate scatterplot and the regression line. For multiple regression, visual representation is much more difficult because it requires several dimensions.

Regression assumptions

To use the OLS method to estimate and make inferences about the coefficients in linear regression analysis, a number of assumptions must be satisfied (Menard 2001), including:

1. Measurement: all independent variables are interval, ratio or dichotomous, and the dependent variable is continuous, unbounded and measured on an interval or ratio scale. All variables are measured without error.
2. Specification: (i) all relevant predictors of the dependent variable are included in the analysis; (ii) no irrelevant predictors of the dependent variable are included in the analysis; and (iii) the form of the relationship (allowing for transformations of dependent or independent variables) is linear.
3. Expected value of error: the expected value of error is 0.
4. Homoscedasticity: the variance of the error term is the same or constant for all values of the independent variables.
5. Normality of errors: the errors are normally distributed for each set of values of the independent variables.
6. No autocorrelation: there is no correlation among the error terms produced by different values of the independent variables (see the section on spatial autocorrelation for more details).
7. No correlation between the error terms and the independent variables: the error terms are uncorrelated with the independent variables.
8. Absence of perfect multicollinearity: for multiple regression, none of the independent variables is a perfect linear combination of the other independent variables (see multicollinearity section).

Goodness of fit (R^2)

In linear regression analysis, evaluation of the overall model is based on two sums of squares. If we were concerned with minimizing the sum of the squared errors of prediction and if we knew only the values of the dependent variable (but not the cases to which those values belonged), we could minimize the sum of the squared errors of prediction by using the mean of Y as the predicted value of Y for all cases. The sum of squared errors based on this prediction would be $\sum(Y_j - \overline{Y})^2$, the total sum of squares (SST). If the independent variables are useful in predicting Y, then $\hat{Y}_j$, the value of Y predicted by the regression equation (the conditional mean of Y), will be a better predictor than $\overline{Y}$ of the values of Y, and the sum of squared errors $\sum(Y_j - \hat{Y})^2$ will be smaller than the sum of squared errors $\sum(Y_j - \overline{Y})^2$. $\sum(Y_j - \hat{Y})^2$ is called the error sum of squares (SSE) and is the quantity OLS selects parameters ($\beta_1, \beta_2, ..., \beta_k$) to minimize. A third sum of squares, the regression sum of squares (SSR), is simply the difference between SST and SSE: SSR = SST – SSE (Menard 2001).

The coefficient of determination (R^2), or 'explained variance', is an indicator of substantive significance; that is, whether the relationship is strong enough for us to be concerned about it. It

measures the proportion (or, multiplied by 100, the percentage) by which use of the regression equation reduces the error of prediction relative to predicting the mean, $\overline{Y}$. R^2 ranges from 0 (the independent variables are no help at all) to 1 (the independent variables allow us to predict the individual values Y_j perfectly). R^2 is calculated as:

$$R^2 = SSR / SST = (SST - SSE) / SST = 1 - (SSE / SST) \quad (3.2.1)$$

Examples
Weiss et al. (2001) used linear regression in a study with the objective of assessing the condition of rangelands in Saudi Arabia and evaluating the effects of grazing. The coefficient of variation (COV) of the monthly normalized difference vegetation index (NDVI) was used as a measure of vegetative biomass change. A higher NDVI COV for a given pixel represented a greater change in vegetation biomass for that area. The trend in COV values was assessed with linear regression over a 12-year period. The COV regression line for each pixel reflects the overall long-term trend in the data. A *t*-test of the value of the slope was performed to test whether the data used to compute the regression line were statistically significant at a certain confidence level.
Other examples of linear regression are Chen (2002), who used linear regression to test the correlations between census dwelling data and residential densities; López et al. (2001), who used linear regression between urban growth and population growth for the prediction of urban expansion in Morelia, Mexico; and de Wolff et al. (2000), who tested the role of accessibility in milk price formation in Kenya as a determinant of livestock adoption in farming systems.
Linear regression for the analysis of multiple land use types is only used when the land use data are represented as continuous values instead of dichotomous. Such a representation is used in the case of a coarse spatial resolution at which the data land use situation cannot adequately be presented by dichotomous data; see e.g. Verburg and Chen (2000) and Wood and Skole (1998).

3.2.2 Logistic regression

Logistic regression is useful for situations where the dependent variable has a binary output, e.g. the presence or absence of a characteristic or outcome. The method is useful to predict the probability that a case will be classified into one as opposed to the other of the two categories of the dependent variable. Several transformations are made to adequately deal with the binary structure of the dependent variable. The odds that $Y = 1$, written odds(Y=1), is the ratio of the probability that $Y = 1$ to the probability that $Y \neq 1$. The odds that $Y = 1$ is equal to P(Y=1) / [1 – P(Y=1)]. Unlike P(Y=1), the odds has no fixed maximum value, but like the probability, it has a minimum value of 0 (Menard 2001).

One further transformation of the odds produces a variable that varies, in principle, from negative infinity to positive infinity. The natural logarithm of the odds, ln{P(Y=1) / [1 – P(Y=1)]}, is called the logit of Y. The logit of Y, written logit(Y), becomes negative and increasingly large in absolute value as the odds decrease from 1 towards 0, and becomes increasingly large in the positive direction as the odds increase from 1 to infinity. If the natural logarithm of the odds that $Y = 1$ is used as the dependent variable, there is no longer the problem that the estimated probability may exceed the maximum or minimum possible values for the probability. The equation for the relationship between the dependent variable and the independent variables becomes:

$$\text{logit}(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k \quad (3.2.2)$$

The logit(Y) can be converted back to the odds by exponentiation. Then the odds can be converted back to the probability that (Y=1). This produces the equation:

$$P(Y=1) = \frac{e^{(\alpha+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_k X_k)}}{1+e^{(\alpha+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_k X_k)}} \quad (3.2.3)$$

It is important to understand that the probability, the odds, and the logit are three different ways to express exactly the same thing. Of the three measures, the probability is probably the most easily understood. Mathematically, however, the logit form of the probability best helps us to analyse dichotomous dependent variables (Menard 2001).

Categorical variables
Categorical independent variables, e.g. soil type or ethnicity, should be dummy or indicator coded. A dummy means that a column (map) indicating presence (1) or absence (0) of that variable has to be created for each value of the independent variable. In the case of a soil map describing five different soil types a map for each soil type should be made, indicating presence or absence of that independent variable. Most statistical programs have options to specify details of how the logistic regression procedure will handle categorical variables. A frequently used option in LUCC modelling is the 'simple' option, which means that each category of the predictor variable (except the reference category) is compared to the reference category. This is a good method to include the categorical variables such as soil suitability, which is often divided into classes such as very good, good, moderate, bad and unsuitable.

Standardized beta
Independent variables are often measured in different units or on different scales. When we want to compare the strength of the relationship between the dependent variable and different independent variables in linear regression we use standardized regression coefficients. For the same reasons, we may want to consider using standardized coefficients in logistic regression analysis. The use of standardized coefficients is especially appropriate for theory testing and when the focus is on comparing the effects of different variables for the same sample.

A standardized coefficient is a coefficient that has been calculated for variables measured in standard deviation units. A standardized coefficient indicates how many standard deviations of change in a dependent variable are associated with a 1 standard deviation increase in the independent variable. In logistic regression analysis, the calculation of standardized coefficients is more complicated than in linear regression because it is not the value of *Y*, but the probability that *Y* is 1 that is predicted by the logistic regression equation. The actual dependent variable in logistic regression in not *Y*, but logit(*Y*), whose observed values of logit(0) = -∞ and logit(+∞) = +∞ do not permit the calculation of means or standard deviations. Although we cannot calculate the standard deviation directly for the observed values of logit(*Y*), we can calculate the standard deviation directly for the observed values of logit(*Y*) and the explained variance, R^2. Dividing both the numerator and the divisor by N (N – 1 for a sample), we get R^2 = SSR/SST = (SSR/N)/(SST/N) = $s^2_{\bar{Y}} / s^2_Y$. The variance of logit(*Y*) can be calculated based on the standard deviation of the predicted values of logit(*Y*) and the explained variance. The standardized logistic regression coefficients can be estimated, because the standard deviation is the square root of the variance (Menard 2001).

$$b^*_{YX} = (b_{YX})(s_X) / \sqrt{s_{\log it(\bar{Y})} / R^2} = (b_{YX})(s_X)(R) / s_{\log it(\bar{Y})} \quad (3.2.4)$$

where b^*_{YX} is the standardized logistic regression coefficient, b_{YX} is the unstandardized logistic regression coefficient, s_X is the standard deviation of the independent variable *X*, $s^2_{\log it(\bar{Y})}$ is the

variance of logit($\widehat{Y}$), $s_{\log it(\bar{Y})}$ is the standard deviation of logit($\widehat{Y}$), and R^2 is the coefficient of determination.

To calculate standardized logistic regression coefficients with existing statistical software (e.g. SPSS), a guideline is provided by Menard (2001).

The interpretation of the standardized logistic regression coefficient is straightforward and closely parallels the interpretation of standardized coefficient in linear regression: a 1 standard deviation increase in *X* produces a *b** standard deviation change in logit(*Y*) (Menard 2001).

Goodness of fit (Classification tables/pseudo R^2/ROC)
In linear regression analysis, we need to know (i) whether knowing the values of all of the independent variables put together allows us to predict the dependent variable any better than if we had no information on any of the independent variables and, if so, (ii) how well the independent variables as a group explain the dependent variable. For logistic regression, we also may be interested in the frequency of correct as opposed to incorrect predictions of the exact value of the dependent variable, in addition to how well the model minimizes errors of prediction. In linear regression, when the dependent variable is assumed to be measured on an interval or ratio scale, it would be neither alarming nor unusual to find that none of the predicted values of the dependent variable exactly matched the observed value of the dependent variable. In logistic regression, with a finite number (usually only two) of possible values of the dependent variable, we may sometimes be more concerned with whether the predictions are correct or incorrect than with how close the predicted values (the predicted conditional means, which are equal to the predicted conditional probabilities) are to the observed (0 or 1) values of the dependent variable (Menard 2001). Therefore, different methods and measures to evaluate the performance of logistic regression models are used and discussed below (Boyce et al. 2002).

Classification table
Most statistical programs give a classification table as one of the outputs of a logistic regression. This classification table gives a comparison of observed and predicted values of cases. Table 3.6 is an example of the classification table of a logistic model for deforestation, in which + indicates remaining forest and – indicates deforestation. In this case the model predicted quite well the forested areas (90.7%), but the deforested areas were not predicted very well (54.3%). For the classification table normally a cut-off value of 0.5 is used, which means that probabilities < 0.5 are classified as 0 (–), and probabilities > 0.5 are classified as 1 (+). Since prevalence has a decisive influence on the maximum probability achieved by a model, the cut-off value might need to be adapted based on the prevalence of the dependent variable. Peppler-Lisbach (2003) provides an example of correcting the cut-off value of the predicted probabilities based on the prevalence of the land use types in the study area.

Table 3.6. Example of a classification table

	+	–	Percentage correct
+	13600	1400	90.7
–	3700	4400	54.3
		Overall:	77.9

Pseudo R^2
In non-linear regression models an R^2-type summary measure that expresses the degree to which the model and data agree cannot be determined directly. The rationale of R^2-type measures is to express the degree of variation in the data that is explained or unexplained by a particular model. That has led to the pseudo-R^2 measure suggested for non-linear models.

$$\text{pseudo } R^2 = 1 - \frac{SSR}{SST_m} \qquad (3.2.5)$$

Where $SST_m = \sum_{i=1}^{n}(y_i - \bar{y}_i)^2$ is the total sum of squares corrected for the mean and $SSR = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is the residual (error) sum of squares. Even in the absence of an intercept in the model, SST_m is the correct denominator since it is the sample mean $\bar{y}$ that would be used to predict y if the response were unrelated to the covariates in the model. The ratio SSR/SST_m can be interpreted as the proportion of variation unexplained by the model (Schabenberger and Pierce 2002).

ROC

The ROC (relative operating characteristic or receiver operating characteristic) measure is useful to evaluate the performance of models. It is normally used as a measure for the goodness of fit of a logistic regression model similar to the R^2 statistic in OLS regression, although it can be applied to any model that predicts a homogeneous category in each grid cell. The ROC works only for two land use and land cover types; with more land use and cover types an ROC can be determined for each type. This can be accomplished by reclassifying the maps into the category of interest versus all others, thus each category can have its own ROC. In an ROC curve the rate of true positives (sensitivity) is plotted on the vertical axis versus the rate of false positives (specificity) on the horizontal axis for each situation. An ROC curve summarizes the performance of a two-class classifier across the range of possible thresholds. Thus, this measure is 'threshold independent' and therefore not sensitive to the prevalence of the dependent variable. The ROC value is the area under the curve (AUC) that connects the plotted points. With Equation 3.2.6 the area under the curve is computed, where x_i is the rate of false positives for scenario i, y_i the rate of true positives for scenario i, and n the number of points (Pontius and Schneider 2001; Manel et al. 2001).

$$\text{AUC} = \sum_{i=1}^{n}[x_{i+1} - x_i][y_i + y_{i+1} - y_i / 2] \qquad (3.2.6)$$

An ideal model outcome hugs the left side and top side of the graph, and the area under the curve is 1. A random classifier should achieve approximately 0.5. Figure 3.4 shows an example of an ROC curve, in this case forest is the state variable and the predicted probability for forest of a logistic regression is the test variable.

The ROC curve is recommended for comparing model outcomes, as it does not merely summarize performance at a single arbitrarily selected decision threshold, but across all possible decision thresholds. The ROC curve can also be used to select an optimum decision threshold (StatSoft 2003). No general rules are available for judgement of the ROC values. However, any AUC above 0.50 is statistically better than random (Pontius and Schneider 2001), while a value higher than 0.7 is normally considered acceptable for LUCC modelling; Hosmer and Lemeshow (2000) consider AUC beyond 0.8 as excellent and > 0.9 as outstanding.

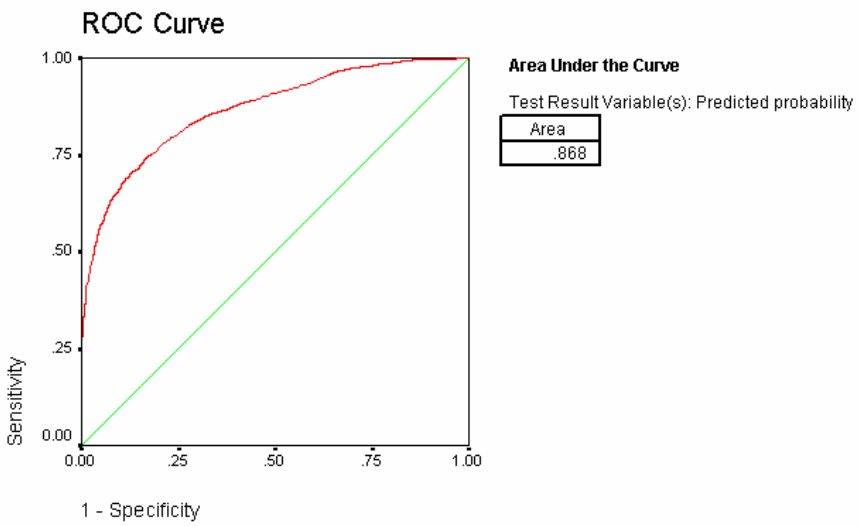

Figure 3.4. Example of an ROC curve of a logistic regression

Other measures for the goodness of fit of the spatial model are based on the likelihood function. These include the value of the maximized log likelihood, the Akaike information criterion (AIC) and the Schwartz criterion (SC). The model with the highest log likelihood, or with the lowest AIC or SC, has the best goodness of fit (Anselin 1992). An example of the comparison of land use change models based on AIC is presented by Aspinall (2004). By rescaling the AIC scores for a series of models against the model with the minimum AIC score the models can be ranked. Rescaled differences between models are transformed and converted to Akaike weights. The Akaike weight for a given model can be interpreted as the probability that the model is the best model, given the data and the set of models (Aspinall 2004).

Examples

Logistic regression is a frequently used methodology in LUCC research. Serneels and Lambin (2001) used logistic regression to identify how much understanding of the driving forces of land use changes can be gained through a spatial statistical analysis for the Mara ecosystem in Kenya. All explanatory variables suggested by the conceptual model for the study area were introduced in the statistical mode and, based on the full model information, they analysed which variables contribute significantly to the explanation of land use changes. Schneider and Pontius (2001) used logistic regression for modelling deforestation in the Ipswich watershed of Massachusetts. Geoghegan et al. (2001) used logistic regression to model tropical deforestation and land use intensification in the southern Yucatán peninsular region, in combination with household survey data on agricultural practices. Staal, Baltenweck et al. (2002) used a logit model to explain the adoption of technology in crop-livestock farming systems in Kenya.

Instead of applying logistic regression directly on the whole data set one can also apply a nested strategy, especially for land use types of which one is distinctly different from the others, e.g. urban areas versus maize, rice grassland and fallow, which can all be incorporated in agricultural land. Gobin et al. (2002) used this nested strategy for logistic modelling to derive agricultural land use determinants for a case study area in south-east Nigeria. First, a binary logistic model was used to predict local agricultural land use under private ownership on the basis of landform and spatial accessibility variables. Afterwards, an ordinal logistic model simulated probabilities of the four different levels of communal agricultural land use (Figure 3.5). This approach was chosen because the land use system for privately managed agricultural land is distinctly different from the communal land use system.

Another example of the use of logistic regression is provided in Box 1.

Box 1. Logistic regression to analyse patterns of land use change

Verburg, Ritsema van Eck, de Nijs, Dijst and Schot (2004) used logistic regression to analyse the factors determining land use patterns in the Netherlands. The method was based on an extensive database, including land use, biophysical, socio-economic, neighbourhood and policy characteristics. All data were aggregated to 500×500 metre grids covering the Netherlands. Historic and recent land use changes were studied. The long-term effects of land use changes were studied by analysing current land use patterns. Many factors that are commonly used to explain land use change patterns are endogenous at a long timescale, e.g. measures indicating current accessibility. Therefore the assumption was made that long-term land use change was mainly determined by biophysical factors. A binomial logit model was compiled for each land use type:

$$\text{Logit } P = \alpha + \beta_1 X_{\text{soil}} + \beta_2 X_{\text{altitude}} + \beta_3 X_{\text{dist-hist-town}}$$

Table 3.7 gives the exp(β) values (odds ratio) for the logit models describing the land use pattern for the main land use types in 1989. Values lower than 1 mean that the probability will decrease upon an increase in the value of the independent variable, while values higher than 1 indicate an increase in probability. A very clear association exists between the pH and the location of forest, which is mainly found on poor sandy soils. The last row gives the ROC values, which indicate the goodness of fit. Model fit for forest is good, while the independent variables for residential and industrial areas only explain a small fraction of the spatial variability. The logit models indicated which factors were important determinants of land use patterns in the Netherlands.

Table 3.7. Logistic regression estimates (exp(β) values) for land use patterns

	Forest	**Arable land**	**Grassland**	**Residential area**	**Industrial area**
Constant	0.14	0.10	0.62	0.70	0.06
Altitude	1.03	*	0.98	*	1.01
Organic matter topsoil	0.99	1.02	0.996	*	*
Organic matter subsoil	1.04	0.98	*	*	*
Calcium content topsoil	1.10	1.29	0.72	*	*
Loam content subsoil	0.93	1.02	1.01	1.01	*
Clay content topsoil	*	0.97	1.01	0.95	*
pH topsoil	0.64	*	1.17	*	*
pH subsoil	2.04	0.9	0.86	0.77	*
Distance to historic town		1.00004		0.9999	0.99998
Distance to open water				*	0.99
ROC	**0.82**	**0.73**	**0.73**	**0.67**	**0.65**

* Not significant, all other exp(β) values are significant at 0.05 level

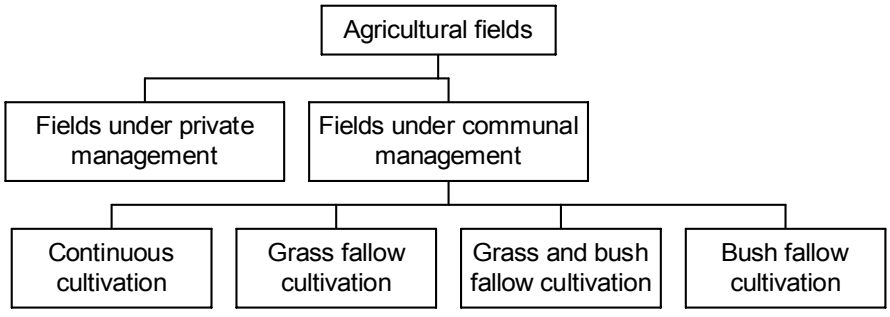

Figure 3.5. Partially nested strategy to predict the probabilities of local agricultural land use (Gobin et al. 2002)

3.2.3 Multinomial regression

Multinomial logit models are used for the case of a dependent variable with more than two categories (Jobson 1992). This type of regression is similar to logistic regression, but it is more

general because the dependent variable is not restricted to two categories. Each category is compared to a reference category, e.g. all types of forest conversion are compared to the stable forest category. The dependent variable should be categorical. Independent variables can be factors or covariates. In general, factors should be categorical variables and covariates should be continuous variables. It is assumed that the odds ratio of any two categories is independent of all other response categories.

Multinomial logit models estimate the direction and intensity of the explanatory variables on the categorical dependent variable by predicting a probability outcome associated with each category of the dependent variable. The probability that $Y = h$ can be stated as:

$$P(Y = h) = \frac{e^{\beta' x_{lh}}}{\sum_{m=1}^{M} e^{\beta' x_{lm}}} \qquad (3.2.7)$$

m denotes the land cover classes used for analysis, β a vector of estimation parameters and x_l are the exogenous variables for all Y and at all locations l. This equation holds, if the error terms are independently and identically distributed as log Weibull (McFadden 1973). Normalizing on all probabilities yields a log-odds ratio (Greene 2000):

$$\ln\left[\frac{p_{lh}}{p_{lm}}\right] = x'_l\,(\beta_h - \beta_m) \qquad (3.2.8)$$

The dependent variable is expressed as the log of the odds of one alternative relative to a base alternative. If model assumptions hold, the maximum likelihood estimators are asymptotically normally distributed, with a mean of zero and a variance of one for large samples. Significance of estimators is tested with z-statistics.

An alternative is the nested logit model, which assumes that the choices can be grouped into categories. Within a category, the independence of irrelevant alternatives condition holds, whereas across categories, it need not. The categories can be grouped into different choice sets, producing a tree structure. The conditional probability ($P_{ml|J}$) of a land use m in choice set J is:

$$P_{ml|J} = \frac{e^{\beta_m' x_l|J}}{\sum_{f=1}^{H_J} e^{\beta_f' x_l|J}} \qquad (3.2.9)$$

where x_l is a vector of location-specific attributes that are common to all land use choices. β_f is a vector of estimated parameters that are specific to a given land use choice. H_J is the set of land uses in choice set J (Nelson et al. 2004). This type of regression is useful for land use decisions that show some hierarchy or grouping.

Examples

Müller and Zeller (2002) used a multinomial logit model to estimate the influence of hypothesized determinants on land use and the probabilities a certain pixel has for one of five land classes. This analysis was conducted for the periods 1975–1992 or 1992–2000 for two districts of Dak Lak Province in the Central Highlands of Vietnam. Mertens et al. (2002) used multinomial regression to improve the understanding of deforestation processes by crossing spatial analyses and livestock economics studies and to characterize the role and impact of various natural and anthropogenic factors in the location and development of the main types of farmers, and their policy implications. Multinomial logit models were run in order to characterize the role and importance of the independent variables in explaining the forest conversion for specific types of activities or processes of colonization. Also Nelson et al. (2004) used multinomial logit regression for the econometrical estimation of a spatially explicit economic model of a proposed road improvement activity in Panama's Darién Province and the simulation of location-specific effects on land use.

The results of a study by Speybroeck et al. (2004) indicate some of the shortcomings of multinomial regression techniques in a study of farming system choice in the Republic of the Congo. Fitting a full model containing all possible interactions becomes an impossible task with 20 explanatory variables. Therefore, they propose to use a classification tree method to generate information on the structure of the data set before the multinomial regression. The advantages of this approach for the analysis of livestock and agricultural production systems are discussed.
Box 2 provides an example of using multinomial regression in the analysis of deforestation

Box 2. Multinomial regression to model deforestation

Chomitz and Gray (1996) developed a spatially explicit land use model to explore the tradeoff between rural roads and deforestation. The model was applied for southern Belize, an area that experienced rapid expansion of agriculture. The model was estimated using data on a sample of land points, with information on slope, distance to road and soil quality. Three land use types were distinguished: natural vegetation, subsistence agriculture and commercial agriculture. Therefore a multinomial logit model was used, which can handle multiple dependent variables. The multinomial logit model estimates the coefficients, provided that the coefficients of one land use type, in this case natural vegetation, are normalized to zero. A bootstrapping procedure was used to estimate the standard error of the coefficients taking spatial autocorrelation into account.

Table 3.8 gives the results of the multinomial logit estimations. Natural vegetation is used as the comparison class. Both forms of agriculture become less attractive as distance to market increases, though commercial agriculture is much more sensitive to distance. The soil characteristics strongly affect the probability of agricultural use, e.g. nitrogen is relatively more important for subsistence agriculture, due to lack of mineral fertilizers. The flood hazard dummy, which is an indicator of riverside location, is strongly related with the probability of subsistence agriculture. The results suggested that road construction in areas with agriculturally poor soils and low population densities were lose-lose situations, causing deforestation and providing low economic returns.

Table 3.8. Descriptive statistics and multinomial logit estimates of land use

Variable	Descriptive statistics		Commercial agriculture	Subsistence agriculture
	Mean	St. dev.		
Distance to market (km)	3.19	4.72	-2.25	-0.60
Nitrogen (%)	0.136	0.050	5.16	16.9
Slope (degrees)	16.3	15.7	-0.017	0.035
Available phosphorus (ppm)	5.26	7.14	0.043	0.042
pH	5.32	0.96	1.50	2.32
Wetness	1.71	2.19	-0.46	1.05
Flood hazard (dummy)	0.40	0.49	0.090	0.92
Rainfall (m)	2.30	0.62	-0.32	0.27
Forest reservation (dummy)	0.46	0.50	-3.21	-1.90
National land (dummy)	0.20	0.40	-0.72	0.72
Constant			-5.05	-14.0

3.2.4 Ordered logit

Ordered logit is not yet, as far as is known to us, used in LUCC-related studies. However, potentially it can be an interesting method for LUCC studies when the dependent variable is discrete, with a certain ordering among the categories. For example, ordered logit could be used to analyse the degree of crop cultivation intensity (low, medium, high) when it is not possible to capture this by a continuous variable or proxy. Especially when data are based on stakeholder perceptions, this may be a suitable approach. The ordered logit model is explained below for a

study of Sonneveld (2003). He evaluated and formalized the use of expert judgements to conduct a nationwide water erosion hazard assessment in Ethiopia. The expert opinions were reproduced with an ordered logit model that predicted the presence or absence of erosion.

The study applied an ordered logit model that uses a continuous but unobserved variable y (for example, soil loss in tonnes per ha per year) in a regression with a set of independent variables x (site characteristics and land use). The range of this y is subdivided into adjacent intervals representing classes (e.g. 1 = no erosion; 2 = moderate; 3 = severe; etc.) that represent an observed discrete variable z. In the logit model, additive error terms are used, so that the underlying process is given by:

$$y_i = \beta' x_i + \varepsilon_i \tag{3.2.10}$$

where β is the vector of parameters to be estimated; ε_i is the disturbance, assumed to be independent across observations; y_i can take any value and the subscript i refers to the observation number. The relation between z_i, given in ordered classes (1, 2, ..., n), and y_i is that adjacent intervals of y_i correspond with qualitative information z_i. This relation is given by:

$$\begin{aligned} z_i &= 1 \quad \text{if } y_i < \mu_1 \\ z_i &= 2 \quad \text{if } \mu_1 \le y_i < \mu_2 \\ &\dots \\ z_i &= n \quad \text{if } \mu_{n-1} \le y_i \end{aligned} \tag{3.2.11}$$

whereby the ordering requires that thresholds $(\mu_1,..,\mu_{n-1})$ satisfy $\mu_1 < \mu_2 < .. < \mu_{n-1}$. The maximum likelihood method is used to estimate parameters β and thresholds $(\mu_1, ..., \mu_{n-1})$, thereby maximizing the probability of correct classifications.

The probability (P) that $z_i = 1$ is calculated by:

$$P(z_i = 1) = P(y_i < \mu_1) = P(\varepsilon_1 < \mu_1 - \beta' x_i) = F(\mu_1 - \beta' x_i), \tag{3.2.12}$$

the probability that $z_i = 2$ by:

$$\begin{aligned} P(z_i = 2) &= P(\mu_1 \le y_i < \mu_2) = P(\mu_1 \le \beta' x_i + \varepsilon_i < \mu_2) \\ &= P(\varepsilon_i < \mu_2 - \beta' x_i) - P(\varepsilon_i < \mu_1 - \beta' x_i) \\ &= F(\mu_2 - \beta' x_i) - F(\mu_1 - \beta' x_i) \end{aligned} \tag{3.2.13}$$

and the probability that $z_i = n$ by:

$$P(z_i = n) = P(y_i \ge \mu_{n-1}) = P(\varepsilon_i \ge \mu_{n-1} - \beta' x_i) = F(\beta' x_i - \mu_{n-1}) \tag{3.2.14}$$

To meet the requirements of a probability model (monotonic-increasing cumulative distribution and results lie between 0 and 1), the disturbances ε_i are assumed to possess a logistic distribution, leading to a cumulative logistic transformation function

$$\Lambda = \frac{1}{1 + e^{-(\bullet)}} \tag{3.2.15}$$

which maps the admissible area of y, i.e. $(-\infty, \infty)$, to $[0,1]$, with a first derivative that is always positive. Thus, the likelihood function for the ordered logit model that consists of Equation 3.2.13 and 3.2.14 for n=N is given by:

$$L(\beta, \mu_1, \mu_2) = \prod_{y_i=1} \Lambda(\mu_1 - \beta' x_i) \times \prod_{y_i=2} (\Lambda(\mu_2 - \beta' x_i) - \Lambda(\mu_1 - \beta' x_i)) \times \dots \times \prod_{y_i=N} \Lambda(\beta' x_i - \mu_{N-1}) \tag{3.2.16}$$

The function L is minimized with respect to the parameters β and $\mu_1, \mu_2, \dots, \mu_n$ (Sonneveld 2002). Another example of an ordered logit is given in Box 3.

Box 3. Ordered logit to identify determinants of poverty

The method of ordered logit will be illustrated with a study on poverty analysis, since no LUCC-related studies have used ordered logit yet. Geda et al. (2001) used an ordered logit model to examine possible determinants of poverty status in Kenya. An ordered logit model was appropriate for the estimation of relevant probabilities because the categories have a natural order. The ordering of the population subsamples was based on total and food poverty lines as cut-off points in a cumulative distribution of expenditure. Three categories of poverty were distinguished: non-poor, poor and extremely poor.

The data were based on a welfare-monitoring survey for the whole country covering almost 10,000 households. Explanatory variables that were used comprised property-related data (land and livestock holding), household characteristics (employment, age, gender, educational level) and others (time to obtain water and energy and location). An income-based and a consumption-based model were used for the probability prediction. These models were fitted for the whole national sample, but also for two subsamples to differentiate between the rural and urban areas.

Table 3.9. Ordered logit estimates for the consumption-based model and the national sample

Variable*	β coefficient	Z value
Employment in agriculture	0.315	3.33
Primary education	-0.430	-5.54
Secondary education	-1.149	-11.22
University	-2.642	-4.81
Household size	0.199	14.82
Total land holding	-0.011	-2.55
Age	0.041	3.25
Cut-off point 1 ($\mu 1$)	2.379	0.425
Cut-off point 2 ($\mu 2$)	3.140	0.422

* Variables significant at 1% level

Table 3.9 presents the estimated coefficients for the consumption-based model. The coefficients show that the level of education and size of the household are the most important factors associated with poverty. The cutting lines determine the probabilities for each poverty class. This leads to the following probabilities for the ordered logit model of Table 3.9:

Non-poor	$P(\beta' x_i+\varepsilon_i < \mu 1)$	= 0.52
Poor	$P(\mu 1 < \beta' x_i+\varepsilon_i < \mu 2)$	= 0.15
Extremely poor	$P(\mu 2 < \beta' x_i+\varepsilon_i)$	= 0.33

The predicted probabilities for both models and the different samples are presented in Table 3.10. This shows that poverty is mainly concentrated in the rural areas of Kenya. This example can also be applied in a LUCC context; instead of the three poverty levels, one can use land use intensity or forest degradation classes.

Table 3.10. Predicted probabilities of being non-poor, poor or extremely poor

Sample	Income-based model			Consumption-based model		
	Non-poor	Poor	Extremely poor	Non-poor	Poor	Extremely poor
National	0.42	0.13	0.45	0.52	0.15	0.33
Rural	0.39	0.11	0.50	0.49	0.15	0.33
Urban	0.58	0.19	0.23	0.72	0.17	0.13

Box 4. Tobit analysis for farming system classification

Baltenweck et al. (2004) used Tobit analysis for the classification of farming systems in the Kenyan Highlands. The analysis was based on data from a household survey of 2810 agricultural households. Five dependent variables were hypothesized to represent the type of farming systems: percentage of land under coffee, under tea, under fodder, percentage of land fertilized and number of dairy cattle in tropical livestock units (TLU). Tobit analysis was appropriate because the variables were censored with an upper limit of 100 per cent and a lower limit of 0 per cent.

The following steps were taken for the Tobit analysis:

1. Each variable was first regressed on the same set of explanatory biophysical variables. From this first step, predicted levels for each of the dependent variables were computed.
2. The five dependent variables were regressed on the predicted values of the four other dependent variables plus the set of independent variables, which were household characteristics (sex, years of farming experience and education of the household head, dependency ratio), farm characteristics (land size, number of adults, ratio female over total adults), external factors (population density, altitude, annual precipitation) and market access (distance to Nairobi by three road types). Bootstrapping was used to control for a possible heteroscedasticity of the errors terms.
3. Linear predictions were made for the spatial predictions, based on only the significant GIS-derived variables at 5%. The other variables that could not be mapped were included in the constant (using mean values).
4. The results for each variable were mapped and afterwards combined to obtain a classification of the farming systems, e.g. dairy-based farming system (high percentage of land under fodder and high dairy TLU).

In Table 3.11 some results of the analysis are presented. Since the coefficients are not standardized they cannot be compared directly; however, the table gives an indication of which variables are important. For example and as expected, the results show that the decision on growing coffee is negatively correlated with the percentage of land under tea (farmers choose between growing either tea or coffee). On the other hand, dairy and coffee are usually complementary activities as reflected in the positive sign of the dairy TLU variable.
Using this approach, it was possible to capture not only the binary decision (e.g. having coffee or not) but also the extent (e.g. land allocated to coffee). It was also possible to take into account the fact that the decisions are taken simultaneously.

Table 3.11. Coefficients* of the Tobit analysis for variables representing farming system types

Independent variables	Percentage coffee	Percentage tea	Percentage fodder	Percentage fertilized	Dairy TLU
Percentage coffee				7.52	-7.96
Percentage tea	-1.40		-0.330	-2.36	4.24
Percentage fodder		13.4			
Percentage fertilized	-5.03	-3.90	-0.895		
Dairy TLU	0.962	-2.21	0.308	-1.83	
Population density	-0.000114	-0.000097	0.000077		
Altitude	-0.00176	0.00688	-0.000539	0.00945	
Annual precipitation	0.00296		0.000845	0.00305	
Distance Nairobi type 1	-0.00193		-0.000526		-0.00988
Distance Nairobi type 2		0.00365	-0.00220		-0.0231
Distance Nairobi type 3		0.0159		-0.00486	
Constant	1.40	-11.3	0.28	-20.1	2.37

* Significant at 5%

3.2.5 Tobit analysis

Although logit models (binary variable as dependent variable) are widely used, Tobit models should be preferred when the dependent variable is censored so as to avoid loss of information (Lynne et al. 1988; Holloway et al. 2004). A censored variable is a variable for which values in a certain range are all transformed to (or reported as) a single value (see Box 4 for an example).
The regression model that is based on a censored distribution is referred to as the Tobit model (Tobin 1958). The general formulation is usually given in terms of an index function, also called the latent variable (Greene 2000):

$$\begin{aligned} y_i^* &= \beta' x_i + \varepsilon_i \\ y_i &= 0 \text{ if } y_i^* \leq 0, \\ y_i &= y_i^* \text{ if } y_i^* > 0. \end{aligned} \qquad (3.2.17)$$

The dependent variable y_i equals 0 if the latent variable y_i* is below a certain threshold, usually 0. If the values of the latent variable are positive, the dependent variable is equal to the latent variable.

Examples
Baidu-Forson (1999) used Tobit analysis to identify factors that motivate both the level and intensity of adoption of specific soil and water conservation technologies for a case study in Niger. Non-adoption occurs, even in areas of diffusion of improved technologies. Therefore, there are some households with zero adoption of the improved technology at the limit. The application of Tobit analysis is preferred in such cases because it uses both data at the limit (non-adopters – zero values) as well as those above the limit (adopters – positive values) in the estimations (McDonald and Moffit 1980). A direct application of Tobit estimation sufficiently provides the needed information on adoption probability and intensity of use of technologies. It can be shown that the total change in elasticity of adoption can be disaggregated into (i) a change in probability of the expected level of use of the technology for farmers who already are adopters; and (ii) change in the elasticity of the probability of being an adopter.

Chomitz and Thomas (2003) used a Tobit model to explain spatial variation in land use for the Brazilian Amazon. Census tract-level data were used to relate forest conversion and pasture productivity to precipitation, soil quality, infrastructure and market access, proximity to past conversion and protection status. The use of the Tobit model was appropriate because censoring at zero captures the intuition that there will be no conversion in unprofitable areas and the reality that many census tracts lack any agricultural land and censoring at 1 was necessary because clearance of forest cannot exceed 100 per cent.

3.2.6 Simultaneous regression

Simultaneous regression is used to model (interdependent) processes, which occur simultaneously in the real world, e.g. the simultaneous determination of demand and supply or the simultaneous determination of land use and transportation characteristics in a metropolitan region. Simultaneous regression is normally only used for linear regression, although in theory it can be applied also to other regression types. An example is given below, in which deforestation is determined by the population (Y_p), the economic potential (Y_e) and another driving factor, e.g. altitude (Y_3). However, population and economic potential are also partly determined by the extent of deforestation. In this case simultaneous regression should be used.

Deforestation: $Y_d = \beta_0 + \beta_1 Y_p + \beta_2 Y_e + \beta_3 X_3$
Population: $Y_p = a_0 + a_1 Y_d + a_2 X_2$

Economic potential: $Y_e = \gamma_0 + \gamma_1 Y_d + \gamma_2 X_2$

The structural form of a linear simultaneous regression is:

$$\begin{aligned}
&\gamma_{11} y_{t1} + \gamma_{21} y_{t2} + \ldots + \gamma_{M1} y_{tM} + \beta_{11} x_{t1} + \ldots + \beta_{K1} x_{tK} = \varepsilon_{t1} \\
&\gamma_{12} y_{t1} + \gamma_{22} y_{t2} + \ldots + \gamma_{M2} y_{tM} + \beta_{12} x_{t1} + \ldots + \beta_{K2} x_{tK} = \varepsilon_{t12} \\
&\vdots \\
&\gamma_{1M} y_{t1} + \gamma_{2M} y_{t2} + \ldots + \gamma_{MM} y_{tM} + \beta_{1M} x_{t1} + \ldots + \beta_{KM} x_{tK} = \varepsilon_{tM}
\end{aligned} \tag{3.2.18}$$

There are M equations and M endogenous variables, denoted $y_1, \ldots, y_M$. There are K exogenous variables, $x_1, \ldots, x_K$, that may include predetermined values of $y_1, \ldots, y_M$ as well. The first element of x_t will usually be the constant, 1. Finally, $\varepsilon_{t1}, \ldots, \varepsilon_{tM}$ are the structural disturbances. The subscript t will be used to index observations, $t = 1, \ldots, T$. In matrix terms the system may be written as:

$$\begin{aligned}
&y_t' = -x_t' B\Gamma^{-1} + \varepsilon_t' \Gamma^{-1} \\
&y_t' \Gamma + x_t' B = \varepsilon_t'
\end{aligned} \tag{3.2.19}$$

Each column of the parameter matrices is the vector of coefficients in a particular equation, whereas each row applies to a specific variable (Greene 2000).

Potentially, simultaneous regression is a very useful technique to model land use change more realistically. However, as far as is known to us, no LUCC-related studies have yet used this statistical technique.

3.2.7 Multilevel statistics

Many kinds of data, including observational data collected in the human and biological sciences, have a hierarchical or clustered structure. LUCC-related data, for example soil and population data, also normally have this structure. Hierarchy is referred to as consisting of units grouped at different levels. These levels are either spatial, e.g. village territories and districts, or thematic, e.g. the field, household and village levels. The existence of such data hierarchies is neither accidental nor ignorable (Goldstein 1995). To analyse these hierarchical data, multilevel statistics are preferred to the more conventional OLS approach, because with nested data regression coefficients may exhibit dependency, which means that the data may provide less information than if they were distributed at random, as is assumed in OLS. Misleading regression results are therefore likely because systematic associations reduce the effective sample size and lead to understated standard error estimates (Snijders and Bosker 1999).

The outline of a multilevel model is shown in the following ordinary simple regression (Goldstein 1995):

$$y_i = b_0 + b_1 x_i + e_i \tag{3.2.20}$$

Suppose the data have a two-level hierarchical structure. The above equation is expressed with Equation 3.2.21 in the style of multilevel modelling:

$$y_{ij} = b_0 + b_1 x_{1ij} + e_{0ij} = \{b_0 + e_{0ij}\} x_{0ij} + b_1 x_{1ij} \tag{3.2.21}$$

Here, i and j mean subscripts of different levels. x_{0ij} is a dummy variable whose value is all 1. Mean value of the random term e_{0ij} is 0. If either intercept (b_0) or slope (b_1) has a level-2 random term, it becomes a multilevel model:

$$y_{ij} = \{b_0 + u_{0j}\} + \{b_1 + u_{1j}\}x_{1ij} + e_{0ij} = \{b_0 + u_{0j} + e_{0ij}\}x_{0ij} + \{b_1 + u_{1j}\}x_{1ij} \qquad (3.2.22)$$

Mean values of the level-2 random terms (u_{0j} and u_{1j} in this case) are 0. The level-2 random term for intercept (u_{0j}) means how far the intercept for level-2 unit j is apart from the entire intercept b_0. Large variance of u_{0j} means that the intercept differs widely by level-2 units. The right-hand side of Equation 3.2.22 can be divided into fixed part and random part:

$$y_{ij} = \{b_0 x_{0ij} + b_1 x_{1ij}\} + \{e_{0ij} x_{0ij} + u_{0j} x_{0ij} + u_{1j} x_{1ij}\} \qquad (3.2.23)$$

Multilevel models are estimated using an iterative algorithm based on maximum likelihood and generalized least squares. Starting values for regression coefficients are calculated based on preliminary information about the variance-covariance matrix. The matrix is then re-estimated using the starting values for the coefficients. The coefficient estimates are then improved using the new variance-covariance matrix, etc. This process continues until a suitable convergence is attained (Hox 1995). In this way results for both the fixed and random parts of the model are estimated more efficiently compared to OLS (Osgood and Smith 1995). There are two principal estimation algorithms, full maximum likelihood and restricted maximum likelihood (Snijders and Bosker 1999). In general restricted maximum likelihood provides more realistic estimates because it corrects for degrees of freedom lost in estimating error variance. More detailed information about multilevel statistics can be found in the textbook by Goldstein (1995).

Examples:
Multilevel statistics are most frequently applied to the analysis of educational issues: student-level data are nested within class-level data within school-level data. There are, however, some sparse examples of applications in LUCC. Polsky and Easterling (2001) analysed climate sensitivities at multiple spatial scales, at county and district levels. A multilevel model was used with agricultural land value per acre as the dependent variable. The model was a random coefficient framework modified to account for the issues of scale. In this case counties were nested within districts. This means that estimates for the average agricultural land value for a county of average climate, and/or the various climate sensitivities (the slopes), depend in part on characteristics of the district in which the county is located. The multilevel model was used to specify which level-1 units (counties) are nested within which level-2 units (districts). The common approach to multilevel modelling was followed by fitting a series of models, testing at each step the tenability of the hypothesized scale-based variation. The model estimated a non-linear, hill-shaped relationship between July maximum temperatures and agricultural land values, with initial increases beneficial in all counties but more beneficial in districts of high interannual temperature variability. Farmers in districts with high variability have adapted to be more resilient to variability than farmers in areas of comparatively stable climate.
Other examples of LUCC-related applications of multilevel statistics are reported by Hoshino (2001), Pan and Bilsborrow (2005) and Overmars and Verburg (in press).

3.3 Bayesian statistics

Bayesian statistics are not widely used in LUCC modelling research; however, we will briefly discuss the principles of Bayesian statistics and illustrate their use with an example of de Almeida et al. (2003), who applied it in modelling urban change with a cellular automata model.

Bayesian analysis is an approach to statistical analysis that is based on Bayes Law, which states that the posterior probability of a parameter p is proportional to the prior probability of parameter p multiplied by the likelihood of p derived from the data collected. This methodology represents an alternative to the traditional (or frequentist probability) approach: whereas the latter attempts to establish confidence intervals around parameters, and/or falsify a priori null hypotheses, the Bayesian approach attempts to keep track of how a priori expectations about some phenomenon of interest can be refined, and how observed data can be integrated with such a priori beliefs, to arrive at updated posterior expectations about the phenomenon.

A good metaphor (and actual application) for the Bayesian approach is that of a physician who applies consecutive examinations to a patient so as to refine the certainty of a particular diagnosis: The results of each individual examination or test should be combined with the a priori knowledge about the patient, and expectation that the respective diagnosis is correct. The goal is to arrive at a final diagnosis, which the physician believes to be correct with a known degree of certainty.

De Almeida et al. (2003) used Bayesian statistics to develop a structure for simulating urban change based on estimating land use transitions with a cellular automata model. The analysis is introduced simply, referring to a generic probability of land use change ΔN which is influenced

Box 5. Multilevel statistics to model farmland distribution

Hoshino (2001) used a multilevel model to analyse the factors that influence land use distribution, in particular farmland, in Japan. First, a single-level model (ordinary regression) was applied; however, the results were not satisfactory when the model was applied to a wider region. The reasons were the problem of scale effect and the variability of the indicators per region; as a consequence the estimated effects were too small or too large. Therefore a second analysis with a multilevel model was applied with the same set of indicators. A hierarchical structure of municipalities (level-1 units) nested within prefectures (level-2 units) was assumed.

The dependent variable was the percentage of farmland area in each municipality, as derived from a land use survey of 1989. The indicators (independent variables) were divided in six groups: topographic conditions, size of farm, land use intensity, income dependency on farming, off-farm labour and indicators of urbanization (infiltration of non-farm households into rural areas). Principal component analysis was used for some of these indicators to avoid multicollinearity.

The two-level model was run based on the most effective indicator of each group from the single-level analysis. A level-2 random term (u_j) was set when either the variance of the random term at level 2 or the ratio of the variance to the standard error was small enough. This resulted in four indicators (flat lowland topography, farm size, paddy field percentage and the constant), as shown in Table 3.12. All variables were standardized beforehand, which made it possible to compare the coefficients. The flat lowland factor is mostly determining the percentage of farmland. The variance of the level-1 random coefficient (e_{ij}) was significant, and the goodness of fit was much better than the single-level models.

Table 3.12. Results of two-level farmland model (standard error is indicated between brackets)

Indicators	β-coefficient	Variance of u_j	Variance of e_{ij}
Flat lowland topography	0.825 (0.032)	0.040 (0.009)	
Plateau area	0.192 (0.010)	-	
Logarithm of average farm size	0.231 (0.052)	0.130 (0.025)	
Paddy field / total farmland	0.173 (0.064)	0.309 (0.076)	
% part-time farm household	-0.232 (0.016)	-	
Number of family members	0.194 (0.014)	-	
Non-farm household ratio	-0.098 (0.013)	-	
Constant	0*	0.250 (0.061)	0.154 (0.004)

* The parameter for constant was set at zero because all variables were already standardized.

by a factor X, without taking into account specific land use, location and temporal notation. A prior probability of land use change from k to l for any cell i is assumed, which is called $P(\Delta N)$, but what needs to be estimated is the posterior probability of such change, which is influenced by the factor X in question (posterior $P(\Delta N|X)$). The standard form for updating a prior probability to a posterior, based on Bayes Rule (Whittle 1970), is:

$$P(\Delta N|X) = P(\Delta N)\frac{P(X|\Delta N)}{P(X)} \tag{3.3.1}$$

Equation 3.3.1 gives the probability when a change in land use takes place, that is when change is present or $\Delta N_i^{kl} = 1$. But in the case where there is no land use change, when change is absent or $\Delta N_i^{kl} = 0$, the probability must be written as:

$$P(\Delta \overline{N}|X) = P(\Delta \overline{N})\frac{P(X|\Delta \overline{N})}{P(X)} = 1 - P(\Delta N|X) \tag{3.3.2}$$

This formulation enables the use of factors which are in binary form (presence or absence). This is particularly suited to physical infrastructure with socio-economic implications, such as the presence or absence of transport routes, utilities, social housing and so on in different cells. The probability of something happening divided by the probability of it not happening (the odds O) can be obtained by dividing Equation 3.3.2 into Equation 3.3.1:

$$O(\Delta N|X) = O(\Delta N)\frac{P(X|\Delta N)}{P(X|\Delta \overline{N})} \tag{3.3.3}$$

The ratio $P(X|\Delta N)/P(X|\Delta \overline{N})$ is a likelihood that updates the odds of event ΔN taking place in the presence of the factor X, with the ratio being related to support for the event taking place. This equation is best represented in logit form as the 'positive weight of evidence' by taking the logarithms of Equation 3.3.3 to give:

$$\log it(\Delta N|X) = \log it(\Delta N) + \log\frac{P(X|\Delta N)}{P(X|\Delta \overline{N})} = \log it(\Delta N) + W^+ \tag{3.3.4}$$

where W^+ is the positive weight of evidence associated with X. An exactly symmetric analysis can be derived for the log odds associated with the absence of a factor $\overline{X}$. We are now in a position to generalize this to many different factors X^e. The probability equations that we use will depend strongly on the extent to which the multiple factors $\{X^e, e = 1, 2, \ldots, E\}$ are independent of one another. This must be tested prior to using these equations, and if there is strong spatial dependence or association between the factors then more complicated forms must be used, with this kind of analysis being less suitable. In fact, independence from irrelevant alternatives is necessary in logit analysis for if the factors are associated with one another, then the probability estimates are biased. Assuming independence, we can write the conditional or posterior probability as $P(\Delta N|X^1, X^2, \ldots, X^E)$. The generalized forms for positive and negative weights of evidence respectively can now be stated as:

$$\text{logit}(\Delta N|X^1, X^2, \ldots, X^E) = \text{logit}(\Delta N) + \sum_e W_e^+ \tag{3.3.5}$$

$$\text{logit}(\Delta N \left| \overline{X}^1, \overline{X}^2, ..., \overline{X}^E \right.) = \text{logit}(\Delta N) + \sum_e W_e^-$$

Another example of the use of Bayesian statistics is a deforestation study for Madagascar by Agarwal et al. (2005).

Box 6. Bayesian statistics to estimate 'neighbourhood effect' in technology adoption

Holloway et al. (2002) explained and applied Bayesian statistics to estimate the 'neighbourhood effect' in high-yielding variety (HYV) adoption of rice among Bangladeshi rice producers. The use of Bayesian statistics was appropriate because there was a priori evidence that village-level synergy existed in technology adoption in Bangladesh. An external reviewing team found that 'copy farmers' (secondary adopters) abounded in the areas where the project was based. This lead to the observation that spatial reach via secondary adoption had a radius of 2 to 3 km. So the attitude towards adoption of HYV by farmers depends not only on their own internal characteristics, but also on the influence of other farmers in the village. The effect of farmers in surrounding villages was assumed to be negligible.

Adoption of new HYV depends upon a set of price variables, a set of fixed factors (e.g. farm assets, land holding), a set of socio-economic characteristics (e.g. education, income) and neighbourhood influences. The first three sets of characteristics are standard in adoption models and were collected during an intensive farm survey among 407 households. The neighbourhood influences are modelled through the combination of a spatial weight matrix and the spatial correlation parameter ρ. The underlying model is a spatially autoregressive probit model, used as the framework of choice for modelling new technology adoption.

Education, farm size and rented land were the only significant variables in the analysis. Years of education had a negative effect on the adoption of HYV, which seems counterintuitive; however, other studies showed that education provides more off-farm opportunities, which compete with cultivation activities. The posterior means estimate of the neighbourhood correlation coefficient ρ was 0.54, which indicates a strong positive neighbourhood effect for technology adoption in Bangladesh.

3.4 Artificial neural networks

Artificial neural networks are powerful tools that use a machine learning approach to quantify and model complex behaviour and patterns. They are used for pattern recognition, prediction and classification in a variety of disciplines, such as economics, medicine, landscape classification and remote sensing. It is not a statistical technique, but in its functioning it is related to regression models. The use of neural networks has increased substantially over the last several years because of the advances in computing performance. A number of applications in land use-related research have been published recently (Pijanowski et al. 2002; Li and Yeh 2002; Pijanowski et al. 2005).

Artificial neural networks were developed to model the brain's interconnected system of neurons so that computers could be made to imitate the brain's ability to sort patterns and learn from trial and error, thus observing relationships in data. The basics of artificial neural networks are based on Rosenblatt (1958), who created the 'perceptron'. It consists of a single node (Figure 3.6 A), which receives weighted inputs and thresholds the results according to a defined rule. This type of simple neural machine is capable of classifying linearly separable data and performing linear functions. The multilayer perceptron neural net consists of three layers: input (*I*), hidden (*H*) and output (*O*) (Figure 3.6 B), and thus can identify relationships that are non-linear in nature.

Artificial neural network algorithms calculate weights for input values, input layer nodes, hidden layer nodes and output layer nodes by introducing the input in a feed-forward manner, which

propagates through the hidden layer and the output layer. The signals propagate from node to node and are modified by weights associated with each connection. The receiving node sums the weighted inputs from all of the nodes connected to it from the previous layer. The output of this node is then computed as the function of its input, called the 'activation function'. The data move forward from node to node with multiple weighted summations occurring before reaching the output layer.

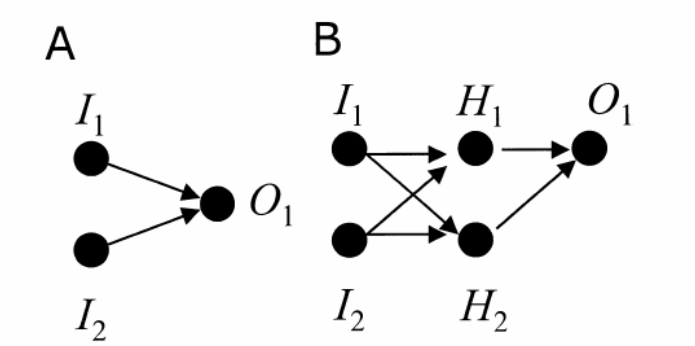

Figure 3.6. A simple perceptron (A) and a multilayer perceptron (B) illustrating input layers (*I*), hidden nodes (*H*) and output layers (*O*)

The determination of weights is critical to successful applications of neural networks. Weights are determined by using a training algorithm, the most popular of which is the back propagation algorithm. This algorithm randomly selects the initial weights, and then compares the calculated output for a given observation with the expected output for that observation. The difference between the expected and calculated output values across all observations is summarized using the mean squared error. After all observations are presented to the network, the weights are modified according to a generalized delta rule (Rumelhart et al. 1986), so that total error is distributed among the various nodes in the network. This process feeding forward signals and back-propagating the errors is repeated iteratively (in some cases, many thousands of times) until the error stabilizes at a low level.

In order to develop a network with adequate predictive capacity, it is necessary to train and test the neural network with different input data (Skapura 1996). Training involves presenting input values and adjusting the weights applied at each node according to the learning algorithm (e.g. back-propagation). Testing presents a separate data set to the trained network independently to calculate the error rate. The benefit of neural networks is their flexibility and non-linearity. However, a drawback is that neural networks function as a black box, which means that neural networks do not provide insight into the relations between variables. In the case of land use the relations between land use and its supposed drivers remains invisible.

Besides prediction applications, artificial neural networks are also widely used for classification of remote sensing images (Figure 3.7). It is generally agreed that artificial neural networks produce classifications with higher accuracies from fewer training samples. Kavzoglu and Mather (2003) constructed a number of guidelines for the effective design and use of artificial neural networks in the classification of remotely sensed image data.

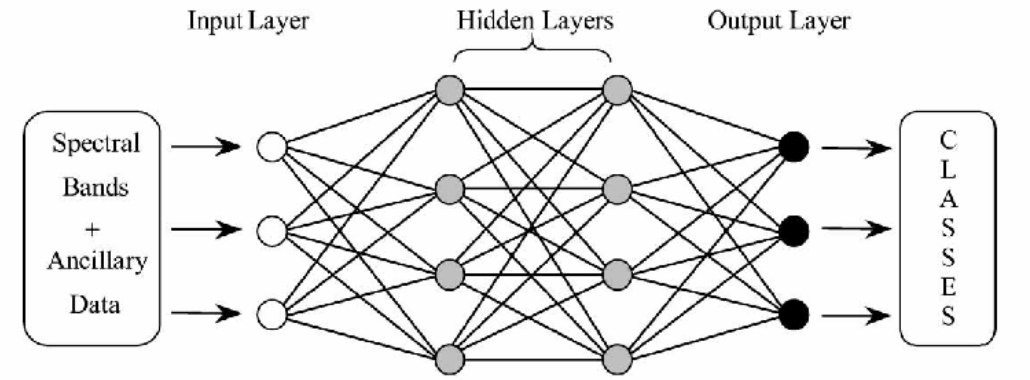

Figure 3.7. A simple four-layer fully connected free-forward neural network as used for image classification (Kavzoglu and Mather 2003)

Examples

Pijanowski et al. (2002) used artificial neural networks to determine the location of land use change using landscape-scale variables, given a certain amount of change determined by regional- and global-scale variables, for Michigan's Grand Traverse Bay watershed. Artificial neural networks were applied to the prediction of land use change in four phases: (i) design of the network and of inputs from historical data; (ii) network training using a subset of inputs; (iii) testing of the neural network using the full data set of the inputs; and (iv) using the information from the neural network to forecast changes. The artificial neural network was a feed-forward network with one input layer, one hidden layer and one output layer. The simple back-propagation algorithm was used as the learning process. The neural network was designed to have a flexible number of inputs depending on the number of predictor variables presented to it, an equal number of hidden units as input units and a single output. The neural network was incorporated in the Land Transformation Model, which explored how geographical factors (e.g. roads, highways, rivers, coastline, recreational facilities and agricultural density) can influence urbanization patterns.

Li and Yeh (2002) simulated the evolution of multiple land uses based on the integration of neural networks and cellular automata. Conventional cellular automata models have problems in defining simulation parameter values, transition rules and model structures. The integration of neural networks and cellular automata should be much better for the simulation of complex land use systems because neural networks are very good at coping with wrong and poor data and capturing non-linear complex features in modelling processes. A three-layer neural network with multiple output neurons was designed to calculate conversion probabilities for competing multiple land uses. The model was applied for a fast-growing urban area in southern China.

4 Special issues relevant to the spatial analysis of land use and farming systems

4.1 Multicollinearity

Dependencies between the explanatory variables are an important issue to account for in all multivariate methods. Data should be checked on multicollinearity before any regression analysis. Collinearity arises when independent variables are correlated with one another. Perfect collinearity means that an independent variable is a perfect linear combination of the other independent variables. If each independent variable in turn is treated as the dependent variable in a model with all of the other independent variables as predictors, perfect collinearity would result in $R^2 = 1$ for each of the independent variables. When perfect collinearity exists, it is impossible to obtain a unique estimate of the regression coefficients; any of an infinite number of possible combinations of linear or logistic regression coefficients will work equally well. Perfect collinearity is rare, except as an oversight: the inclusion of three variables, one of which is the sum of the other two, would be one example (Menard 2001).
Less than perfect collinearity is fairly common. Many variables that are frequently used in land use analysis, such as distance to roads and markets, tend to be highly correlated, making it difficult to distinguish their separate effects. Any correlation among the independent variables is indicative of collinearity. As collinearity increases among the independent variables, linear and logistic regression coefficients will be unbiased and as efficient as they can be (given the relationships among the independent variables), but the standard errors for linear or logistic regression coefficients will tend to be large. More efficient unbiased estimates may not be possible, but the level of efficiency of the estimates may be poor. Low levels of collinearity are not generally problematic, but high levels of collinearity may pose problems, and very high levels of collinearity almost certainly result in coefficients that are not statistically significant, even though they may be quite large. Collinearity also tends to produce linear and logistic regression coefficients that appear to be unreasonably high: as a rough guideline, standardized logistic or linear regression coefficients greater than 1 or unstandardized logistic regression coefficients greater than 2 should be examined to determine whether collinearity is present (Menard 2001).

Collinearity is easy to detect, but there are only few acceptable remedies for it (Dohoo et al. 1997). Deleting a variable involved in collinearity runs the risk of omitted variable bias. Methods to prevent multicollinearity include factor analysis, a priori correlation analysis and stepwise regression. Applying factor analysis before running a regression will combine collinear variables, if present, into a single factor. Instead of factor analysis one can also apply an even simpler form of correlation test by looking at the coefficient of determination (R^2) between all pairs of variables (Mertens et al. 2002). A weak association between explanatory variables implies a relatively low degree of collinearity, whereas one or more variables should be excluded when a strong association is found. Menard (2001) suggests that correlations > 0.8 between independent variables should be regarded as a high level of collinearity.
Stepwise regression is another method to reduce collinearity between the independent variables in a regression. The use of stepwise regression is, however, restricted to exploratory analysis, when we are more concerned with theory development than theory testing. Such research may occur in the early stages of the study of a phenomenon, when neither theory nor knowledge about correlates of the phenomenon is well developed. Stepwise regression refers to the use of decisions made by computer algorithms, rather that choices made directly by the research, to select a set of predictors for inclusion in or removal from a linear or logistic regression model. Stepwise procedures may be useful for purely predictive research and exploratory research. In purely predictive research, there is no concern with causality, only with identifying a model, including a set of predictors, that provides accurate predictions of some phenomenon. In

exploratory research, there may be a concern with theory construction and development to predict and explain a phenomenon, when the phenomenon is new or so little studied that existing 'theory' amounts to little more than empirically unsupported hunches about explanations for the phenomenon (Menard 2001).
Two methods exist for stepwise regression: backward elimination and forward inclusion. Usually, the backward elimination and forward inclusion methods will produce the same results, but when the results differ, backward elimination may uncover relationships missed by forward inclusion, due to the suppressor effect (Agresti and Finlay 1997). This means that in some cases a variable may appear to have a statistically significant effect only when another variable is controlled or held constant. With backward elimination, because both variables will already be in the model, there is less risk of failing to find a relationship when one exists.

Many authors have provided critical comments on the stepwise procedure. The procedure of screening variables by stepwise procedures may improve prediction, but it may also eliminate variables that are in fact important; stepwise procedures are not intended to rank variables by their importance (James and McCulloch 1990). Others have argued that stepwise regressions cannot necessarily find the best fitting model either. The best that can be hoped for, when a stepwise multiple regression is used, is the selection of a subset of the variables that does an adequate job of prediction. Users of stepwise regression should take the limitations of the method into account and refrain from conclusions based on the variable selection by these methods.

4.2 Spatial autocorrelation

The problem of using conventional statistical methods, like linear and logistic regression, in spatial land use analysis is that these methods assume the observations to be statistically independent and identically distributed (Cliff and Ord 1981). However, spatial land use data have the tendency to be dependent, a phenomenon known as spatial autocorrelation. Spatial autocorrelation may be defined as the property of random variables to take values over distance that are more similar or less similar than expected for randomly associated pairs of observations, due to geographic proximity (Legendre and Legendre 1998).

Spatial dependency could be seen as a methodological disadvantage because conventional statistics may lead to the wrong conclusions. On the other hand, the spatial relations actually provide information on spatial pattern, structure and processes. So, spatial dependency contains useful information but to deal with it statistically the appropriate methods have to be used. The effects of spatial dependence on conventional statistical methods are various, for example biased estimation of error variance, *t*-test significance levels and overestimation of R^2 (Anselin and Griffith 1988). All the usual statistical tests have the same behaviour: in the presence of positive autocorrelation, computed test statistics are too often declared significant under the null hypothesis. Negative autocorrelation may produce the opposite effect (Legendre and Legendre 1998). This is caused by the fact that an observation carries less information than an independent observation, since it is partly predictable from its neighbours and a new sample point does not bring with it one full degree of freedom (Cliff and Ord 1981; Legendre and Legendre 1998). Figure 4.1 illustrates the different types of spatial autocorrelation.

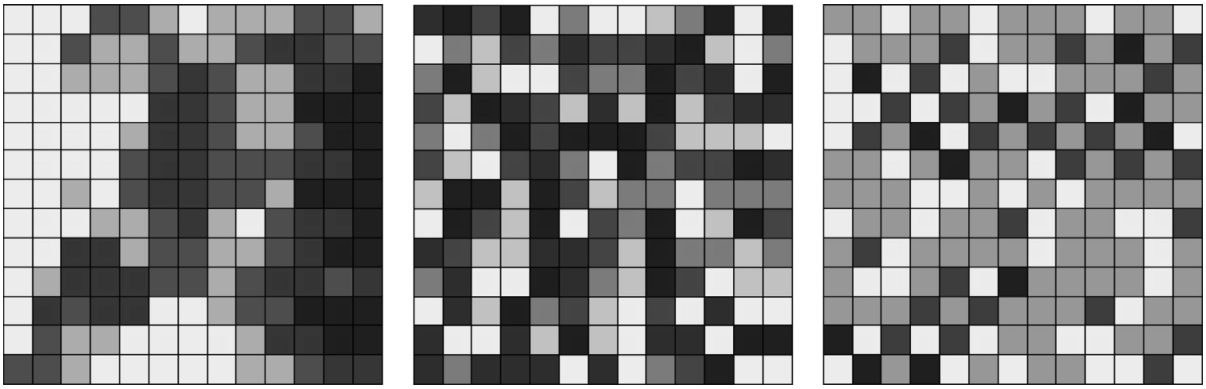

Figure 4.1. Types of spatial autocorrelation. Visualization of positive spatial autocorrelation (left), no spatial autocorrelation (middle), and negative spatial autocorrelation (right) in an imaginary 13x13 grid. The different tones of grey indicate different values of a variable (Overmars et al. 2003).

Detection of spatial autocorrelation

Spatial structures, like spatial dependency, can be described through structure functions. The most commonly used structure functions are correlograms, variograms and periodograms. These graphs show the spatial dependency per distance class (spatial lag). In correlograms autocorrelation values are plotted against distance classes. This can be computed for both univariate (Moran's *I* or Geary's *c*) and multivariate data (Mantel correlogram). Correlograms are one of the most frequently used structure functions in spatial autocorrelation. Correlograms are preferable over, for example, semi-variograms for two reasons. First, the significance of the correlation coefficient can be tested; and second, correlograms are standardized, so different cases can be compared (Meisel and Turner 1998).

The value of Moran's *I* generally varies between 1 and –1, although values lower than –1 or higher than +1 may occasionally be obtained. Positive autocorrelation in the data translates into positive values of *I*; negative autocorrelation produces negative values. No autocorrelation results in a value close to zero (Legendre and Legendre 1998). The following formula is used to calculate Moran's *I*:

$$\text{for } h \neq i \qquad I(d) = \frac{\frac{1}{W}\sum_{h=1}^{n}\sum_{i=1}^{n} w_{hi}(y_h - \overline{y})(y_i - \overline{y})}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2} \qquad (4.2.1)$$

in which y_h and y_i are the values of the observed variable at sites *h* and *i*. The values of w_{hi} are the weights. The weights w_{ij} are written in a (*n*x*n*) weight matrix *W*. *W* is the sum of the weights w_{hi} for a given distance class. The weight matrix depicts the relation between an element and its surrounding elements. Weight can be based, for example, on contiguity relations or on distance. In a weight matrix based on contiguity, a 1 in the matrix represents pairs of elements with a certain contiguity relation and a 0 represents pairs without contiguity relation. Two examples of weight matrices for a regular grid are rook contiguity and queen contiguity. The first takes only full neighbours into account and the latter all eight surrounding cells. The complete matrices contain the contiguity relations of all pairs of points. Besides this contiguity principle it is also imaginable to make weight matrices based on geographic distances between the centroids of the elements. The relations between the cells are then calculated comparing the distances between the centroids.

Accounting for spatial autocorrelation in regression models

Spatial autocorrelation can be analysed on unmodified data or on the residuals of a regression analysis. If autocorrelation is detected on the regression residuals, this can imply that the

regression model should have an autoregressive structure, or that non-linear relationships between the dependent and the independent variables (trend) are present, or that one or more important regressor variables are missing (Long 1998). The most general formulation of a spatial autoregressive model is Equation 4.2.2 (Anselin 1988; LeSage 1999).

$$\begin{aligned} y &= \rho W_1 y + X\beta + u \\ u &= \lambda W_2 u + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 I_n) \end{aligned} \qquad (4.2.2)$$

In this equation y contains an nx1 vector of cross-sectional dependent variables, X represents and nxk matrix of explanatory variables, and W_1 and W_2 are known nxn spatial weight matrices. The parameter ρ is a coefficient on the spatially lagged dependent variable and λ is a coefficient on the spatially correlated errors (LeSage 1999). β is a kx1 vector with linear regression coefficients, as in a standard linear regression model. The error term (ε) is an nx1 vector of independent identical normally distributed variables with zero mean and variance σ^2.

Specific models can be derived from the general model by imposing restrictions. Setting $X = 0$ and $W_2 = 0$ produces a first-order spatial autoregressive model, explaining variation in y as a linear combination of contiguous or neighbouring units with no other explanatory variables. Setting $W_2 = 0$ produces a mixed regressive-spatial autoregressive model. This model has additional explanatory variables in the matrix X to explain variation in y over the spatial sample of observations. This model is also called the simultaneous model (Anselin 1988) or simultaneous spatial autoregression (Kaluzny et al. 1997). $W_1 = 0$ results in a regression model with spatial autocorrelation in the disturbances. Anselin (2002) describes this model as a standard regression model with spatially filtered variables. A model known as the spatial Durban model contains a spatial lag in both the dependent variable (matrix W) and the independent variables (matrix X).

In case of a row-standardized W_1, the spatial part of the mixed regressive-autoregressive model functions as an extra variable equal to the (weighted) mean of observations from contiguous cells. If spatial dependence between the observations in the data set y is assumed, some part of the total variation in y across the spatial sample is explained by each observation's dependence on its neighbours. The parameter ρ would reflect that in the typical sense of regression (LeSage 1999).
For further reading about different ways to incorporate spatial effects in regression models see Anselin (2002).

Besides the above-discussed autoregressive model, two other methods exist to account for spatial autocorrelation. One is the inclusion of spatially lagged variables and the other is structured sampling. The spatial lag method explicitly includes the spatial context in the model. Each lag variable is the average of the values of the original variable in the eight cells surrounding the location. Nelson and Geoghegan (2002) used spatial lag variables, which included the latitude and longitude values, and average vegetative and soil quality indices in the surrounding locations.

Instead of explicitly correcting for spatial autocorrelation some authors use structured sampling methods to reduce the influence of spatial autocorrelation on the estimated model. Structured sampling can be based on three different techniques: the semi-variogram, the Besag approach and the jack-knife or bootstrap method. Structured sampling on the basis of a semi-variogram means that the range of a semi-variogram determines the sampling distance. The range is an indication for the area where variables are still spatially dependent. A sampling distance larger than the range supposes the variables to be spatially independent. The second method is a regular sampling procedure suggested by Besag (1974). The Besag approach includes only observations separated by sufficient distance in space that the autoregressive effect is absent. With a coding scheme a

sample from the full data set is selected, so no two sites in the sample are neighbours, e.g. neighbouring cells in the sample are 5 pixels apart in the full data set.
The last structured sampling technique is the jack-knife or bootstrap method. This is a technique to reduce bias and to obtain standard errors of the estimated autocorrelation coefficients. The jack-knife method is performed by fitting the regression model with *n* observations but one, leaving out successively one observation at a time. This procedure leads to the calculation of the empirical influence values *e* for each observation. These values can be plotted as a function of the observation number to detect possible outlying observations. Similar to this technique is the bootstrap, proposed by Efron (1982). It is a very general computing-intensive method to produce an approximation to the unknown distribution of a statistic. The bootstrap is a technique for estimating the standard error of a statistic using repeated samples from the original data set. This is done by sampling (with replacement) to get many samples of the same size as the original data set. The non-linear equation is estimated for each of these samples. The standard error of each parameter estimate is then calculated as the standard deviation of the bootstrapped estimates. Parameter values from the original data are used as starting values for each bootstrap sample.

Examples
Overmars et al. (2003) used correlograms of Moran's *I* to describe spatial autocorrelation for a data set of Ecuador. Positive spatial autocorrelation was detected in both dependent and independent variables (Figure 4.2). Also, the residuals of the original regression model showed positive autocorrelation. To overcome the positive autocorrelation a mixed regressive-spatial autoregressive model was used, which incorporated both regression and spatial autocorrelation. The models yielded residuals without spatial autocorrelation and had a better goodness of fit. Most autoregressive models are based on linear regression. No examples of autoregressive models are known in LUCC science. An example of an autologistic model in ecology is presented by Augustin et al. (1996).

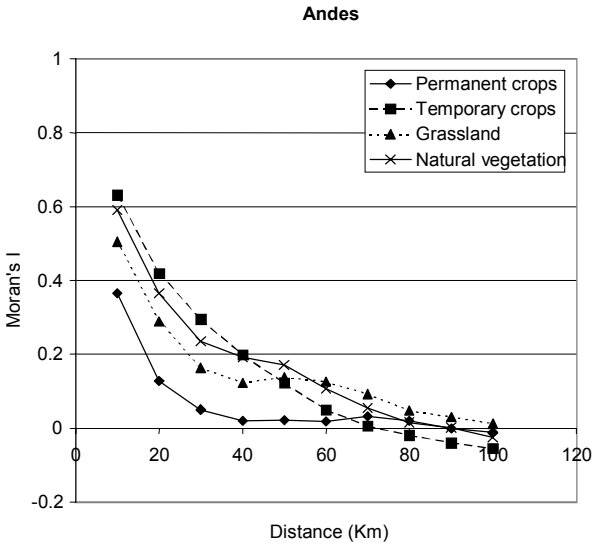

Figure 4.2. Correlogram with Moran's *I* comparing different land use types (Overmars et al. 2003)

In an example of addressing spatial autocorrelation in survey-based data, de Wolff et al. (2000), in a study on milk price formation in Kenya using both household survey and GIS-derived data, controlled for spatial autocorrelation by introducing the interactions between farmers as suggested by Anselin et al. (1993). In this case, the distances between each farmer and a common point (urban centres) are introduced in the regression analyses as an indirect measure of potential spatial relationships between farmers. Existence of autocorrelation is tested both with and without the distance variables. The authors show that residuals do exhibit spatial autocorrelation

when only household survey variables are introduced in the analysis but do not when the distance measures are included in the analysis.

Other examples of studies in which spatial dependence in land use patterns is explicitly addressed during the statistical analysis are the studies by Brown et al. (2002), who used geostatistical tools to quantify spatial dependence in land use patterns; Verburg, Ritsema van Eck, de Nijs, Visser and de Jong (2004), who developed a measure to explore spatial autocorrelation within a single land use category and between land use categories; and Polsky (2004), who used autoregressive models to model land use for the US Great Plains.

4.3 Validation techniques

Validation is the evaluation of the predicted modelling results. Validation is not the same as measuring the goodness of fit. Goodness of fit, as described for the different regression analysis methods, is based on the same data used for fitting the model, while validation is based on independent data. When a regression model is used to predict values of the dependent variable for other data, e.g. another subset or a different time-step or region, the performance of the model for extrapolation can be validated. The paragraphs below describe several validation techniques that are used in LUCC modelling.

4.3.1 Multiple resolution validation

No generally agreed-upon method for the quantitative evaluation of the goodness of fit for spatial models has evolved yet. The often-used visual comparison of modelled land use patterns with actual land use patterns is a rather subjective manner of validation. One objective manner is a multiple resolution procedure for the model goodness of fit developed by Costanza (1989). It is based on the measuring of similarity of the patterns, and the idea that measurement at one resolution is not sufficient to describe complex patterns. The method yields indices that summarize the way the fit changes as the resolution of measurement changes. An expanding 'window' is used to gradually degrade the resolution of the comparison.

With this method the near misses, besides the direct hits, receive some weight and tell whether the pattern matches. The fit for each sampling window is estimated as 1 minus the proportion of cells that would have to be changed to make the sampling windows each have the same number of cells in each category, regardless of their spatial arrangement:

$$F_w = \frac{\sum_{s=1}^{t_w} \left[1 - \frac{\sum_{i=1}^{p} |a_{1i} - a_{2i}|}{2w^2} \right]_s}{t_w} \qquad (4.3.1)$$

F_w is the fit for sampling window size w, w the dimension of one side of the (square) sampling window, a_{ki} the number of cells of category i in scene k in the sampling window, p the number of different categories (i.e. land use types) in the sampling windows, s the sampling window of dimension w by w which slides through the scene one cell at a time and t_w the total number of sampling windows in the scene for window size w.

To determine an overall degree of fit between two maps the information in the plot of window sizes versus fit (Figure 4.3) must be summarized. A weighted average of the fits at different

window sizes is a possible way of summarizing the overall fit that allows more weight to be given to smaller window sizes while not totally ignoring the large window sizes. For this purpose the following formula can be used:

$$F_t = \frac{\sum_{w=1}^{n} F_w e^{-k(w-1)}}{\sum_{w=1}^{n} e^{-k(w-1)}} \qquad (4.3.2)$$

F_t is a weighted average of the fits over all window sizes, F_w the fit for sampling windows of linear dimension w, k a constant, and w the dimension of one side of the (square) sampling window. The value of k determines how much weight is to be given to small versus large sampling windows. A default value for $k = 0.1$.

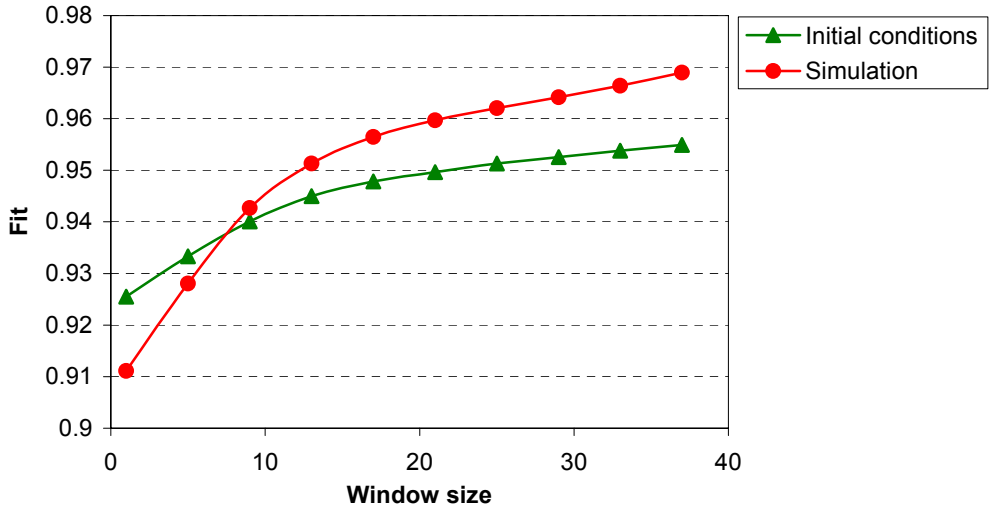

Figure 4.3. Example plots of window sizes versus fit for two models

The calculated fit of the model is only a relative value that can be used to compare the results of different models. To get an impression of the absolute performance of the model the fit should also be compared with a certain standard or 'null model'.

4.3.2 Kappa characteristic

The Kappa statistic is a measure of accuracy that ranges between 0 (completely inaccurate) and 1 (completely accurate) and measures the observed agreement between the classification and the reference data and the agreement that might be attained solely by chance matching (Munroe et al. 2002). The Kappa statistic (Pontius 2002) is:

$$\kappa = \frac{P_o - P_c}{P_p - P_c} \qquad (4.3.3)$$

where P_o is the observed proportion correct, P_c is the expected proportion correct due to chance and P_p is the proportion correct with perfect classification. In addition to the standard Kappa index of agreement, Pontius (2000, 2002) defines three variations: Kappa for no information (K_{no}), Kappa for location (K_{loc}), and Kappa for quantity (K_{quan}). K_{no} is an overall index of agreement, K_{loc} is an index that measures the agreement in terms of location only and K_{quan}

measures the agreement in terms of quantity. According to Pontius (2000) a Kappa value higher than 0.5 can be considered as satisfactory for land use change modelling. Landis and Koch (1977) characterize agreement as follows: values > 0.75 are very good to excellent, values between 0.4 and 0.75 are fair to good and values of 0.4 or less indicate poor agreement.

4.3.3 ROC

Besides its use as a measure for goodness of fit, as described in Chapter 3.2.2, the ROC can also be used to assess the predictive power of a regression model. The ROC procedure offers a way of identifying an optimum probability threshold by simply reading the point on the curve at which the sum of sensitivity and specificity is maximized. One can assess whether increasing or reducing probability thresholds for accepting presence, as optimized from the ROC curve, has any effects on the predicted frequency of occurrence during real model applications. Logistic models can be calibrated for predicting presence of a certain land use using part of the data and applying them to the other part of the data. One can use geographic samples, or random samples of the complete data set. This procedure is recommended for testing any presence-absence model on fully independent data (Manel et al. 1999).

In ecological studies frequent use has been made of the ROC in applications aiming at prediction. For example, Manel et al. (2001) used this method to predict the occurrence of aquatic invertebrates in 180 Himalayan streams. The model was calibrated for the five westernmost regions and applied to the geographically distinct eastern regions. In LUCC studies Pontius and Batchu (2003) have used the ROC to assess the predictive power of a land use model. They describe a methodology to quantify the certainty in predicting the location of change for a given quantity of change. The methodology converts a map of relative propensity for disturbance to a map of probability of future disturbance, based on a quantifiable validation of a map's predictive ability. They applied the methodology for the Western Ghats in India. The probability of forest disturbance was determined for the period 1920–1990 and extrapolated to the future by validation with the ROC and an independent prediction of the quantity of post-1990 disturbance.

4.3.4 Other validation techniques

Besides the above-mentioned validation techniques some other less frequently used methods for validation of spatial patterns exist. Herold et al. (2003) describe a validation based on spatial metrics that summarize landscape patterns. Spatial metrics can be defined as quantitative and aggregate measurements derived from digital analysis of thematic-categorical maps showing spatial heterogeneity at a specific scale and resolution. Six different spatial metrics were used to validate the modelled urban extent: class area (sum of all urban patches); number of patches, largest patch index (area of the largest patch divided by the total area covered by urban); Euclidian mean nearest neighbour distance (the distance mean value over all urban patches to the nearest neighbouring urban patch); area-weighted mean patch fractal dimension (area-weighted mean value of the fractal dimension values of all urban patches); and contagion (measures the overall probability that a cell of a patch type is adjacent to cells of the same type). This last index is an overall measure of the landscape heterogeneity and provides a subtle characterization of the spatial arrangement of vacant/undeveloped and urban land (Herold et al. 2003).

Another technique for validation of maps is the fuzzy set statistic. The main purpose of the fuzzy set approach is to take into account that there are grades of similarity between pairs of cells in two maps. Fuzziness means a level of uncertainty and vagueness of a map. The approach therefore is fundamentally different from its counterpart, the cell-by-cell map comparison, which considers pairs of cells to be either equal or unequal. The fuzzy set approach expresses similarity

of each cell in a value between 0 (distinct) and 1 (identical). In order to distinguish minor differences from major differences, the fuzzy set approach takes two types of fuzziness into account: fuzziness of categories and fuzziness of location. Fuzziness of category means that some categories in the map are more similar to each other than others. Fuzziness of location means that the spatial specification found in a categorical map is not always as precise as it appears. The two fuzziness parameters can be combined into one overall similarity measure, the Kappa statistic (K_{Fuzzy}):

$$K_{Fuzzy} = \frac{P_o - P_e}{1 - P_e} \tag{4.3.4}$$

where P_o is the observed percentage of agreement (i.e. average similarity) and P_e is the expected similarity (Hagen 2003). The fuzzy set approach and some other map comparison methods are incorporated in the software package Map Comparison Kit (Visser 2004). This program with documentation can be downloaded at http://www.mnp.nl/ (search for Map Comparison Kit)

4.4 Scale dependency

Scale refers to the spatial or temporal dimension used to measure and study a phenomenon. All scales have extent and resolution: extent refers to the size of a dimension, e.g. the size of the study area or the duration of time under consideration, whereas resolution refers to the precision used in measurement, i.e. grain size (Turner et al. 1989). The grain of an observation is the finest distinction made between isolated datum values. It determines the smallest entities that can be seen in the study. In contrast to the grain, the extent determines the largest entities that can be detected in the data. The scale of the study is an interaction of grain and extent. If the extent is large, the sampling protocol will be expensive unless the grain is relatively coarse (Allen and Hoekstra 1991).

Each analysis of spatial pattern incorporates scale explicitly or implicitly into the process of identifying research objects: the very act of identifying a particular pattern means that scale, extent and resolution have been used. These choices over scale, extent and resolution critically affect the type of pattern that will be observed: patterns that appear at one level of resolution may be lost at lower or higher levels; patterns that occur over one extent of a dimension may disappear if the extent is increased or decreased (Allen and Hoekstra 1991). Figure 4.4 illustrates the effect of grain size for the analysis of a classic von Thünen land use pattern. When this type of land use pattern is studied at a detailed resolution (resolution 1) urban land use and horticulture are negatively correlated. However, when analysed at a coarser resolution (resolution 2) urban land use and horticulture fall within the same unit of observation and therefore are positively correlated. Thus, observations and theories derived at one scale may not apply at another. Figure 4.4 also illustrates that the scales of organization of natural and human systems are different from the scales of observation, which often are determined by the measurement technique (Turner II et al. 1995).

Different studies have shown that the grain and extent of analysis influence the identified statistical relationships between land use and its predictors (Veldkamp and Fresco 1997; Walsh et al. 2001; Verburg and Chen 2000). Possible explanations for the influence of the grain of analysis are:

- Reduction of spatial variability: coarse grain sizes obscure variability whereas fine grain sizes obscure general trends. Shifts in grain size may produce more than averages or constants; they may make homogeneity out of heterogeneity and vice versa (Kolasa and Rollo 1991).

- Emergent properties: changes in grain size are frequently associated with new or emergent properties. In complex, constitutive hierarchies, characteristics of larger units are not simple combinations of attributes of smaller units.
- Some factors can have influence over a considerable distance. At coarse grains these factors fall within the same unit of analysis and therefore cause a change in correlation structure (Figure 4.4).
- Stronger overlap among variables: aggregation reduces intraclass variance and the size of the sample population, smoothing distributions and reducing the number of outlier values identified within each class. This can create strong overlap among variables, greatly reducing the potential value of such variables for distinguishing classes.

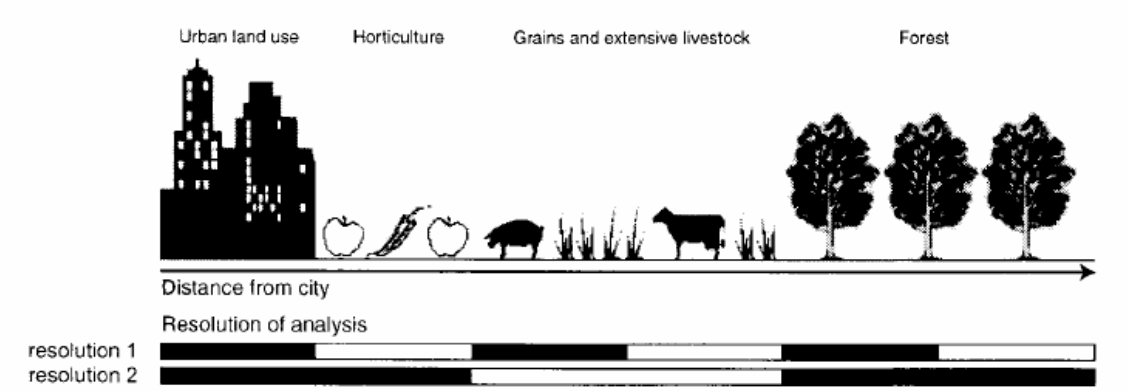

Figure 4.4. Land use pattern according to the von Thünen theory and schematic representation of grain size at two different resolutions (Verburg and Chen 2000)

The influence of the extent of analysis can be explained by the decreasing importance of local situations with an increasing extent of analysis. A smaller extent allows the introduction of specific variables that are important for the area under analysis. Therefore, a smaller extent offers better insight into the specific situation of the region whereas a larger extent allows for identification of general patterns (Verburg and Chen 2000).

Another scaling issue concerns thematic aggregation, e.g. the aggregation of individual crops and land use systems into land use types. Cultivated land includes a diversity of crops and cropping intensities. All these individual crops and land use systems have their own spatial distribution and explanatory factors. At the same time, the distribution of cultivated land has properties that cannot be derived from the distribution of all the different crops separately.

Methods
Basically three methods exist to explore scale dependencies in empirical relations based on land use patterns. First, scale dependence can be captured by increasing artificially the resolution of the model and comparing the results. This method is described in several studies, e.g. Veldkamp and Fresco (1996), Verburg and Chen (2000) and Walsh et al. (2001). A second method is to include regional or spatially lagged parameters in the model, e.g. distance to the main city or distance to market. The third method is to use multilevel statistics, as described in Chapter 3.9.

Example
Walsh et al. (2001) assessed the statistical relationships between plant biomass levels and selected social, biophysical and geographical variables at nine different cell resolutions, ranging from 30 to 1050 metres. The basic intent was to examine the scale dependence of population and environment relationships in a study area in north-east Thailand. This region has experienced pronounced land use changes associated with deforestation and agricultural extensification and intensification to support lowland rice and upland cash crops. The variation in plant biomass was

assessed through a satellite-based measure, the normalized difference vegetation index (NDVI). For each variable and at successively coarser resolutions, cell values were calculated by hierarchical aggregation of the original 30-metre grid. The beta values of each variable were assessed and the R^2 values of each multiple regression model were tracked over the nine spatial resolutions. The results indicated that population factors were more important at finer scales and biophysical factors at coarser scales for explaining variation in plant biomass levels.

5 Challenges for empirical analysis in LUCC: Beyond regression?

LUCC researchers face the challenge of developing adequate explanations for the occurrence, timing and nature of major environmental phenomena. The previous chapters have provided an overview of statistical and empirical techniques for the analysis of spatial patterns of LUCC. The application of these techniques in various case studies has resulted in a better understanding of the driving factors of LUCC. However, the use of these techniques has a number of drawbacks and limitations. Therefore, statistical methods by themselves are not sufficient to fully understand and predict land use change. This final chapter will shortly discuss the challenges faced by LUCC researchers, which include the lack of a full-fledged theory, the issue of causality versus empirical evidence, the use of different perspectives, scaling issues and the use of case studies.

Lack of full-fledged theory

A theory of land use change should conceptualize the relationships between the driving and conditioning forces and land use change, relationships among the driving forces, and human behaviour and organization underlying these relationships. Different disciplinary theories can help to analyse aspects of land use change in specific situations. The synthesis of these theories is essential, but 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 (Overmars and Verburg 2005). So far researchers have not yet succeeded in integrating all disciplines and complex elements of the land use system into an all-compassing theory of land use change. Conclusions drawn from disciplinary LUCC studies can vary substantially between disciplines, which implies that the complexity of the land use system as a whole is not completely understood (Lambin et al. 2001).

Statistical techniques as presented in this report can help to explore data sets, identify associations between LUCC and its drivers and contribute to theory building and testing. However, the absence of an all-compassing theory is no reason to limit ourselves to inductive approaches. Theories from multiple disciplines, such as economics (e.g. optimal resource allocation), geography, ecology (e.g. complex systems theory) and anthropology (e.g. human behaviour), can contribute to the explanation of LUCC. Furthermore, there are also a number of theories available that focus on (specific) land use conversions, e.g. neo-Malthusian theory speaking about poverty traps, neo-Boserupian theory about the positive effects of population on land use sustainability, the induced intensification thesis (Turner II and Ali 1996), neo-Thünen theory about moving frontiers and urban markets (Walker and Solecki 2004) and the theories of Fujita and Krugman about urban development (Krugman 1999; Fujita et al. 1999). Most current theories cannot adequately explain the complexity of land use change. Assumed agent behaviours in the formal bid-rent paradigm are limited, as well focused as they are on the maximization of rents, profits and utility. This may be reasonable for explaining land use under the implicit institutional environment of the bid-rent paradigm, where rents accrue to landowners whose property rights are never disputed, and the economy is free from catastrophic shocks. Of course, many situations are vastly different from the utopia of von Thünen. Individuals may seek to minimize risks or take them, as the case may be (Rabin 1998). Poorly defined property rights are not conducive to the competitive bidding process that leads to the equilibrium rent profile, so essential to both urban and agricultural models.

Empirical methods can be used to test theories in specific case studies. Such a theory-based approach is important to explore for several reasons. It structures the model around the critical human-environment relationships identified within the theory, and focuses attention on the data required to explore those relationships. Furthermore, this approach may improve the theoretical foundations of the macro relations on which LUCC models are based, e.g. logit functions or

cellular automata have to be further specified to improve the explaining and predicting capacity of those models. The relations currently used are often difficult to connect to human behaviour at lower levels of analysis, which makes many policy measures difficult to implement in models, since these will affect individual behaviour and not the aggregated effect that is normally modelled.
The most effective way to reap the benefits of more deductive work does not seem to be to rigidly 'go deductive' and stay there. Such a 'process-led approach' may blind the analyst to alternative processes at work. Rather, the message should be that researchers will profit most from developing a consciousness of the whole spectrum between the inductive and deductive extremes, and an awareness of the advantages of deductive approaches versus the currently dominant inductive research routines, and then seeking the most fertile sequences and interactions between inductive and deductive work. Ultimately, this will contribute to theory development in the field of land use change.

Causality versus empirical evidence
A major drawback of the empirical quantification of relations between land use and its supposed drivers is the induced uncertainty with respect to the causality of the supposed relations: judgements about the results based on their interpretability can be dangerously close to circular reasoning. The greatest danger of all is of leaping directly from the exploratory stage, or even from statistical tests based on descriptive models, to conclusions about causes (James and McCulloch 1990). This asks for validation of the causality of empirically derived relations. Statistical approaches do not allow for describing causal relationships. Besides, most causal explanations are valid at the scale of study, mostly the individual agent of LUCC, and therefore subject to upscaling problems. A combination of methods, including multiscale analysis and multi-agent models, can be used to integrate the empirical and individual behaviour approaches. An example of such a combined approach is a study of Overmars in the Philippines (Overmars and Verburg 2005; Overmars et al. 2005). They applied an approach that explores the results of statistical models based on geographic data in combination with a household-level analysis of decision making in order to incorporate the theories about human decision making in spatially explicit models.

Different perspectives and scales
Different perspectives and scales of analysis are needed to obtain a full understanding of the process of LUCC. The three generic approaches to study land use change, which are the empirical, the narrative and the modelling approaches, are representative of the different perspectives (Lambin et al. 2003). Integration of the results of these approaches should lead to a better understanding of LUCC. Techniques for the empirical approach have been extensively discussed in this report. Further implementation and sophistication of these techniques at multiple scales should lead to a better understanding of LUCC. The narrative and modelling approach will now be discussed in some more detail.

The narrative perspective seeks depth of understanding through historical detail and interpretation. It tells the LUCC story, providing an empirical and interpretative baseline by which to assess the validity and accuracy of the other visions. It is especially beneficial in identifying stochastic and random events that significantly affect LUCC but might be missed in approaches employing less expansive time horizons or temporal sampling procedures (Briassoulis 2000). The narrative approach is mostly valid at the level of individual agents of LUCC, while most of the interesting features of LUCC occur at more aggregate levels of analysis. Coleman (1990) developed a framework that describes the interaction between micro and macro level for social systems, which can be applied to land use change research as well. Land use change assessments are often made using remote sensing and GIS at the regional (macro) level, while at the same time trying to explain these macro-level developments by specifying a micro-level mechanism. Figure 5.1, based on the work of Coleman (1990), depicts the relations between the

macro and micro levels. Macro-level analyses (pathway A) of land use are normally based on empirical techniques, e.g. the analysis of spatial patterns of land use derived from remote sensing. Pathway B explains the underlying processes that lead to the different land use patterns, e.g. the individual decisions in response to land use policies. Together, these individual decisions lead to the changes in land use pattern. Following this trajectory one can explain why differences in macro conditions lead to different land use patterns. A better understanding of individual behaviour regarding land use change and its spatial impact makes it possible to better address stakeholders, which leads to much more efficient policy support and interventions.

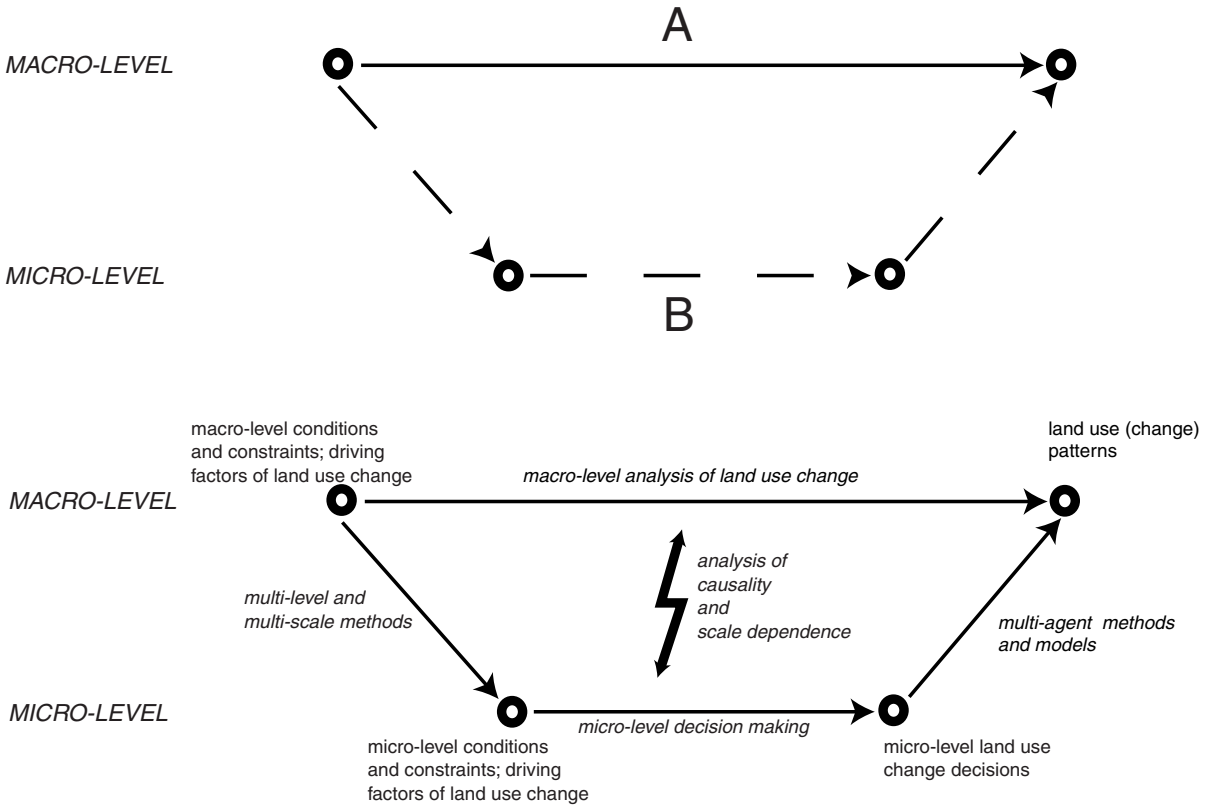

Figure 5.1. Representation of the linkage between micro-level and macro-level research in land use change

For the modelling approach many types of models are available focusing on different processes, e.g. deforestation or agricultural intensification, and based on a wide range of simulation techniques, e.g. regression models, expert models and cellular automata models (Irwin and Geoghegan 2001; Briassoulis 2000; Verburg, Schot et al. 2004). A relatively new type of model is multi-agent systems. Multi-agent models simulate decision making by individual agents of land use change, explicitly addressing interactions among individuals. The explicit attention to interactions between agents makes it possible for this type of model to simulate emergent properties of systems. These are properties at the macro scale that are not predictable from observing the micro units in isolation. If the decision rules of the agents are set such that they sufficiently look like human decision making they can simulate behaviour at the meso level of social organization, i.e. the behaviour of heterogeneous groups of actors. Multi-agent-based models of LUCC are particularly well suited to representing complex spatial interactions under heterogeneous conditions and modelling decentralized, autonomous decision making (Parker et al. 2003; Bousquet and LePage 2004). Multi-agent systems are able to formalize decision-forming behaviour of individual stakeholders, based on a theoretical argumentation. Most multi-agent models focus on either hypothetical or simplified situations to explore interactions between agents and between agents and the environment, rather than simulating landscape change at the regional level.

Case studies

Globally valid explanations of what factors drive land use change remain largely incomplete (NRC 1999). Various hypotheses have produced rich arguments, but empirical evidence on the causes of land use change, e.g. deforestation, continues to be largely based on cross-national statistical analyses (Rudel and Roper 1997). Common understanding of the causes of LUCC is dominated by simplifications that, in turn, underlie many environment development policies. Case study evidence supports the conclusion that the simple answers found in population growth, poverty and infrastructure rarely provide an adequate understanding of LUCC. Rather, individual and social responses follow from changing economic conditions, mediated by institutional factors. Opportunities and constraints for new land uses are created by markets and policies, increasingly influenced by global factors (Lambin et al. 2001).

Geist and Lambin (2002, 2004) and McConnell and Keys (2005) aimed to generate a general understanding of the proximate causes and underlying driving forces of, respectively, tropical deforestation, desertification and agricultural intensification from local-scale case studies, while preserving the descriptive richness of these studies. Proximate causes are human activities or immediate actions at the local level, such as agricultural expansion, that originate from intended land use and directly impact forest cover. The findings suggest that no universal link between cause and effect exists; instead LUCC is determined by different combinations of various proximate causes and underlying driving forces in varying geographical and historical contexts. The challenge for comparative land use analysis is to develop generalizations about land cover consequences not only for individual operations but for their application in sequences, including the role of natural cycles and landscape patterns (Turner II et al. 1995). It is a challenge to comprise such sets of region-specific drivers with empirical techniques and to analyse comparative case studies to reveal the general patterns in LUCC.

Final remark

This section has provided some alternatives to the statistical methods described in this report. Most of these alternatives address issues that cannot adequately be addressed by statistical methods. By no means should this give the impression that statistical methods are not very useful in land use change analysis. Statistical methods have been used as a powerful and flexible tool in many studies. The alternative methods each have their specific strengths, typical applications and limitations. For the analysis of land use and land cover change there is no single, perfect method or approach. The selected method should fit the research questions, available data and resources. For a full understanding of the land use change processes at work a combination of methods and approaches might provide most insight, applied to a cohesive theoretical framework. Comparisons and contradictory results should be used to signal weaknesses and demand a re-evaluation of the data and methods of analysis, and perhaps the underlying theory. Therefore, it is essential to make best use of the diversity of approaches to understand the complex dynamics underlying the changes in our environment. Statistical methods, as described in this report, can make an important contribution to this type of analysis.

6 References

Agarwal D.K., Silander J.J.A., Gelfand A.E., Dewar R.E. and Mickelson J.J.G. 2005. Tropical deforestation in Madagascar: Analysis using hierarchical, spatially explicit, Bayesian regression models. *Ecological Modelling* 185(1):105–131.

Agresti A. and Finlay B. 1997. *Statistical methods for the social sciences*. 3rd edition. Prentice Hall, Upper Saddle River, New Jersey, USA.

Allen T.F.H. and Hoekstra T.W. 1991. Role of heterogeneity in scaling of ecological systems under analysis. In: Kolasa J. and Pickett S.T.A. (eds), *Ecological heterogeneity*. Springer-Verlag, New York, USA.

Alonso W. 1964. *Location and land use*. Harvard University Press, Cambridge, USA.

Anselin L. 1988. *Spatial econometrics: Methods and models*. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Anselin L. 1992. *SpaceStat tutorial*. West Virginia University, Morgantown, USA.

Anselin L. 2002. Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27:247–267.

Anselin L., Dodson R. F. and Hudak S. 1993. Linking GIS and spatial data analysis in practice. *Geographic Systems* 1:3–23.

Anselin L. and Griffith A.D. 1988. Do spatial effects really matter in regression analysis? *Papers Regional Science Association* 65:11–34.

Aspinall R. 2004. Modelling land use change with generalized linear models: A multi-model analysis of change between 1860 and 2000 in Gallatin Valley, Montana. *Journal of Environmental Management* 72(1–2):91–104.

Augustin N.H., Mugglestone M.A. and Buckland S.T. 1996. An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology* 33(2):339–347.

Baidu-Forson J. 1999. Factors influencing adoption of land-enhancing technology in the Sahel: Lessons from a case study in Niger. *Agricultural Economics* 20:231–239.

Baltenweck I., van de Steeg J. and Staal S.J. 2004. Farming systems characterisation in the Kenyan Highlands: Use of alternative methodologies. Working document. ILRI, Nairobi, Kenya.

Besag J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society* Series B 36:192–236.

Bousquet F. and LePage C. 2004. Multi-agent simulations and ecosystem management: A review. *Ecological Modelling* 176:313–334.

Boyce M.S., Vernier P.R., Nielsen S.E. and Schmiegelow F.K.A. 2002. Evaluating resource selection functions. *Ecological Modelling* 157(2–3):281–300.

Briassoulis H. 2000. Analysis of land use change: Theoretical and modeling approaches. In: Loveridge S. (ed), *The web book of regional science*. West Virginia University, Morgantown, USA.

Brown D.G., Goovaerts P., Burnicki A. and Li M.Y. 2002. Stochastic simulation of land-cover change using geostatistics and generalized additive models. *Photogrammetric Engineering and Remote Sensing* 68(10):1051–1061.

Burnsilver S.B., Boone R.B. and Galvin K.A. 2003. Linking pastoralists to a heterogeneous landscape: The case of four Maasai group ranches in Kajiado District, Kenya. In: Fox J., Rindfuss R.R., Walsh S.J. and Mishra V. (eds), *People and the environment: Approaches for linking household and community surveys to remote sensing and GIS*. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Cardille J.A. and Foley J.A. 2003. Agricultural land-use change in Brazilian Amazônia between 1980 and 1995: Evidence from integrated satellite and census data. *Remote Sensing of Environment* 87:551–562.

Chen K. 2002. An approach to linking remotely sensed data and areal census data. *International Journal of Remote Sensing* 23(1):37–48.

Chomitz K.M. and Gray D.A. 1996. Roads, land use and deforestation: A spatial model applied to Belize. *World Bank Economic Review* 103:487–512.

Chomitz K.M. and Thomas T.S. 2003. Determinants of land use in Amazônia: A fine-scale spatial analysis. *American Journal of Agricultural Economics* 85(4):1016–1028.

Cliff A.D. and Ord J.K. 1981. *Spatial processes: Models and applications*. Pion, London, UK.

Coleman J.S. 1990. *Foundations of social theory*. The Belknap Press of Harvard University Press, Cambridge, USA.

Comrey A.L. and Lee H.B. 1992. *A first course in factor analysis*. Lawrence Erlbaum Associates Publishers, Hillsdale, USA.

Costanza R. 1989. Model goodness of fit: A multiple resolution procedure. *Ecological Modelling* 47:199–215.

Davis J.C. 1986. *Statistics and data analysis in geology*. 2nd edition. John Wiley and Sons, New York, USA.

de Almeida C.M., Batty M., Vieira Monteiro A.M., Câmara G., Soares-Filho B.S., Coutinho Cerqueira G. and Lopes Pennachin C. 2003. Stochastic cellular automata modeling of urban land use dynamics: Empirical development and estimation. *Computers, Environment and Urban Systems* 27:481–509.

de Leeuw J., Waweru M.N., Okello O.O., Maloba M., Nguru P., Said M.Y., Aligula H.M., Heitkonig I.M.A. and Reid R.S. 2001. Distribution and diversity of wildlife in northern Kenya in relation to livestock and permanent water points. *Biological Conservation* 100(3):297–306.

de Wolff T., Staal S., Kruska R., Ouma E., Thornton P. and Thorpe W. 2000. *Improving GIS derived measures of farm market access: An application to milk markets in the East African highlands.* Paper presented at the Fifth Seminar on GIS and Developing Countries (GISDECO 2000), 'GIS Tools for Rural Development', 2–3 November 2000, IRRI, Los Banos, Philippines.

Dohoo I.R., Ducrot C., Fourichon C., Donald A. and Hurnik D. 1997. An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. *Preventive Veterinary Medicine* 29(3):221–239.

Duchateau L., Kruska R.L. and Perry B.D. 1997. Reducing a spatial database to its effective dimensionality for logistic-regression analysis of incidence of livestock disease. *Preventive Veterinary Medicine* 32(3–4):207–218.

Efron B. 1982. *The jackknife, the bootstrap and other resampling plans*. CBMS-NSF Regional Conference Series in Applied Mathematics Monograph 38. SIAM, Philadelphia, USA.

Escobar G. and Berdegué J. 1990. *Tipificación de sistemas de producción agrícola*. Red Internacional de Metodologías de Investigación en Sistemas de Producción, Santiago, Chile.

FAO (Food and Agriculture Organization of the United Nations). 1997. *Africover land cover classification*. FAO, Rome, Italy.

Fujita M., Krugman P. and Mori T. 1999. On the evolution of hierarchical urban systems. *European Economic Review* 43:209–251.

Geda A., de Jong N., Mwabu G. and Kimenyi M.S. 2001. Determinants of poverty in Kenya: A household level analysis. Working paper. Institute of Social Science, The Hague, the Netherlands.

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

Geist H.J. and Lambin E.F. 2004. Dynamic causal patterns of desertification. *Bioscience* 54(9):817–829.

Geoghegan J., Cortina Villar S., Klepeis P., Macario Mendoza P., Ogneva-Himmelberger Y., Chowdhury R.R., Turner II B.L. and Vance C. 2001. Modeling tropical deforestation in the southern Yucatán peninsular region: Comparing survey and satellite data. *Agriculture, Ecosystems and Environment* 85:25–46.

Gibson C.C., Ostrom E. and Anh T.K. 2000. The concept of scale and the human dimensions of global change: A survey. *Ecological Economics* 32:217–239.

Gobin A., Camping P. and Feyen J. 2002. Logistic modelling to derive agricultural land use determinants: A case study from southeastern Nigeria. *Agriculture, Ecosystems and Environment* 89:213–228.

Goldstein H. 1995. *Multi-level statistical methods*. Halsted Press, New York, USA.

Greene W.H. 2000. *Econometric analysis*. 4th edition. Prentice Hall International Inc, Upper Saddle River, USA.

Hagen A.E. 2003. Fuzzy set approach to assessing similarity of categorical maps. *International Journal of Geographic Information Systems* 173:235–249.

Hary I., Schwartz H.J., Pielert V.H.C. and Mosler C. 1996. Land degradation in African pastoral systems and the destocking controversy. *Ecological Modelling* 86(2–3):227–233.

Herold M., Goldstein N.C. and Clarke K.C. 2003. The spatiotemporal form of urban growth: Measurement, analysis and modeling. *Remote Sensing of Environment* 86:286–302.

Hietel E., Waldhardt R. and Otte A. 2004. Analysing land-cover changes in relation to environmental variables in Hesse, Germany. *Landscape Ecology* 19:473–489.

Hilferink M. and Rietveld P. 1999. Land use scanner: An integrated GIS based model for long term projections of land use in urban and rural areas. *Journal of Geographical Systems* 1:155–177.

Holloway G., Nicholson C., Delgado C., Staal S. and Ehui S. 2004. A revised Tobit procedure for mitigating bias in the presence of non-zero censoring with an application to milk-market participation in the Ethiopian highlands. *Agricultural Economics* 31(1):97–106.

Holloway G., Shankar B. and Rahman S. 2002. Bayesian spatial probit estimation: A primer and an application to HYV rice adoption. *Agricultural Economics* 27:383–402.

Hoshino S. 1996. Statistical analysis of land use change and driving forces in the Kansai District, Japan. Working paper WP-96-120. IIASA, Laxenburg, Austria.

Hoshino S. 2001. Multilevel modeling on farmland distribution in Japan. *Land Use Policy* 18:75–90.

Hosmer D.W. and Lemeshow S. 2000. *Applied logistic regression*. Wiley and Sons, New York, USA.

Hox J. 1995. *Applied multi-level analysis*. TT-Publikaties, Amsterdam, the Netherlands.

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

James F.C. and McCulloch C.E. 1990. Multivariate analysis in ecology and systematics: Panacea or Pandora's box? *Annual Reviews of Ecology and Systematics* 21:129–166.

Jobson J.D. 1992. *Applied multivariate data analysis*. Springer, New York, USA.

Kaluzny S.P., Vega S.C., Cardoso T.P. and Shelly A.A. 1997. *S-Plus spatial stats: User's manual for Windows and Unix*. Springer, New York, USA.

Kavzoglu T. and Mather P.M. 2003. The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing* 24(23):4907–4938.

Kim J. 1970. Factor analysis. In: Nie N.H., Hull C.H., Jenkins J.G., Steinberger K. and Bent D.H. (eds), *Statistical package for the social sciences*. McGraw Hill, New York, USA.

Köbrich C., Rehman T. and Khan M. 2003. Typification of farming systems for constructing representative farm models: Two illustrations of the application of multi-variate analyses in Chile and Pakistan. *Agricultural Systems* 76:141–157.

Kolasa J. and Rollo C.D. 1991. Introduction: The heterogeneity of heterogeneity: A glossary. In: Kolasa J. and Pickett S.T.A. (eds), *Ecological heterogeneity*. Ecological studies 86. Springer-Verlag, New York, USA.

Krugman P. 1999. The role of geography in development. *International Regional Science Review* 22(2):142–161.

Kruska R.L, Reid R.S., Thornton P.K., Henninger N. and Kristjanson P.M. 2003. Mapping livestock-oriented agricultural production systems for the developing world. *Agricultural Systems* 77:39–63.

Lambin E.F. 2003. Linking socio-economic and remote sensing data at the community or at the household level: Two case studies from Africa. In: Fox J., Rindfuss R.R., Walsh S.J. and

Mishra V. (eds), *People and the environment: Approaches for linking household and community surveys to remote sensing and GIS*. Kluwer Academic Publishers, Norwell, USA.

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

Lambin E.F., Turner II B.L., Geist H.J., Agbola S.B., Angelsen A., Bruce J.W., Coomes O.T., Dirzo R., Fischer G., Folke C., George P.S., Homewood K., Imbernon J., Leemans R., Li X., Moran E.F., Mortimore M., Ramakrishnan P.S., Richards J.F., Skånes H., Steffen W., Stone G.D., Svedin U., Veldkamp A., Vogel C. and Xu J. 2001. The causes of land-use and land-cover change: Moving beyond myths. *Global Environmental Change* 11:261–269.

Landis J.R. and Koch G.G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159–174.

Lawley D.N. and Maxwell A.E. 1971. The scope of factor analysis. In: *Factor analysis as a statistical method*. Butterworths, London, UK.

Legendre P. and Legendre L. 1998. *Numerical ecology: Developments in environmental modelling 20*. Elsevier, Amsterdam, the Netherlands.

LeSage J.P. 1999. *The theory and practice of spatial econometrics*. University of Toledo, Toledo, USA.

Li X. and Yeh A.G-O. 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographic Information Systems* 16(4):323–343.

Long D.S. 1998. Spatial autoregression modeling of site-specific wheat yield. *Geoderma* 85:181–197.

López E., Bocco G., Mendoza M. and Duhau E. 2001. Predicting land-cover and land-use change in the urban fringe: A case in Morelia city, Mexico. *Landscape and Urban Planning* 55:271–285.

Lynne G.D., Shonkwiler J.S. and Rola L.R. 1988. Attitudes and farmer conservation behaviour. *American Journal of Agricultural Economics* 70:12–19.

Manel S., Dias J.M. and Ormerod S.J. 1999. Comparing discriminant analysis, neural networks and logistic regression for predicting species' distributions: A case study with a Himalayan river bird. *Ecological Modelling* 120:337–347.

Manel S., Williams H.C. and Ormerod S.J. 2001. Evaluating presence-absence models in ecology: The need to account for prevalence. *Journal of Applied Ecology* 38:921–931.

McConnell W.J. and Keys E. 2005. Meta-analysis of agricultural change. In: Moran E.F. and Ostrom E. (eds), *Seeing the forest and the trees: Human-environment interactions in forest ecosystems*. In press.

McConnell W.J. and Moran E.F. 2001. *Meeting in the middle: The challenges of meso-level integration*. International workshop, October 17–20, 2000, Ispra, Italy. LUCC Report Series No. 5. Indiana University, Bloomington, USA.

McDonald J.F. and Moffit R.A. 1980. The uses of Tobit analysis. *Review of Economic and Statistics* 62:318–321.

McFadden D. 1973. Conditional logit analysis of qualitative choice behaviour. In: Zarembka P. (ed), *Frontiers in Econometrics*. Academic Press, New York, USA.

Meisel J.E. and Turner M.G. 1998. Scale detection in real and artificial landscapes using semivariance analysis. *Landscape Ecology* 13:347–62.

Menard S. 2001. *Applied logistic regression analysis*. Sage University Papers, series on quantitative applications in the social sciences, 07-106. Sage, Thousand Oaks, USA.

Mertens B., Poccard-Chapuis R., Piketty M.G., Lacquies A.E. and Venturieri A. 2002. Crossing spatial analyses and livestock economics to understand deforestation processes in the Brazilian Amazon: The case of São Félix do Xingú in South Pará. *Agricultural Economics* 27(3):269–294.

Mertens B., Sunderlin W.D., Ndoye O. and Lambin E.F. 2000. Impact of macroeconomic change on deforestation in south Cameroon: Integration of household survey and remotely sensed data. *World Development* 28(6):983–999.

Moody A. and Woodcock C.E. 1994. Scale dependent errors in the estimation of land cover proportions: Implications for global land cover data sets. *Photogrammetric Engineering and Remote Sensing* 60:585–594.

Müller D. and 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.

Munroe D.K., Southworth J. and Tucker C.M. 2002. The dynamics of land-cover change in western Honduras: Exploring spatial and temporal complexity. *Agricultural Economics* 27:355–369.

Nelson G.C. 2002. Introduction to the special issue on spatial analysis for agricultural economists. *Agricultural Economics* 27:197–200.

Nelson G.C. and Geoghegan J. 2002. Deforestation and land use change: Sparse data environments. *Agricultural Economics* 27:201–216.

Nelson G.C., Harris V., Stone S. and De Pinto A. 2004. Land use and road improvements: A spatial econometric analysis. *International Regional Science Review* 27:297–325.

NRC (National Research Council), Board on Sustainable Development, Policy Division, Committee on Global Change Research. 1999. *Global environmental change: Research pathways for the next decade*. National Academy Press, Washington, D.C., USA.

Osgood D.W. and Smith G.L. 1995. Applying hierarchical linear modeling to extended longitudinal evaluations: The Boys Town follow-up study. *Evaluation Reviews* 19(1):3–38.

Overmars K.P., de Groot W.T. and Huigen M.G.A. 2005. Comparing inductive and deductive modeling of land use decisions: Principles, a model and an illustration from the Philippines. Submitted for publication.

Overmars K.P., de Koning G.H.J. and Veldkamp A. 2003. Spatial autocorrelation in multi-scale land use models. *Ecological Modelling* 164:257–270.

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 Geographic Information Systems* 19(2):125–152.

Overmars K.P. and Verburg P.H. In press. Multi-level modelling of land use from field to village level. *Agricultural Systems*, forthcoming.

Pan W.K.Y. and Bilsborrow R.E. 2005. The use of a multi-level statistical model to analyze factors influencing land use: A study of the Ecuadorian Amazon. *Global and Planetary Change*, in press.

Pan W.K.Y., Walsh S.J., Bilsborrow R.E., Frizelle B.G., Erlien C.M. and Baquero F. 2004. Farm-level models of spatial patterns of land use and land cover dynamics in the Ecuadorian Amazon. *Agriculture, Ecosystems and Environment* 101:117–134.

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

Peppler-Lisbach C. 2003. Predictive modelling of historical and recent land-use patterns. *Phytocoenologia* 33(4):565–590.

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

Pijanowski B.C., Pithadia S., Shellito B.A. and Alexandridis K. 2005. Calibrating a neural network-based urban change model for two metropolitan areas of the upper Midwest of the United States. *International Journal of Geographical Information Science* 19(2):197–216.

Polsky C. 2004. Putting space and time in Ricardian climate change impact studies: The case of agriculture in the US Great Plains. *Annals of the Association of American Geographers* 94(3):549–564.

Polsky C. and Easterling III 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* 85:133–144.

Pontius Jr. R.G. 2000. Quantification error versus location error in comparison of categorical maps. *Photogrammetric Engineering and Remote Sensing* 66(8):1011–1016.

Pontius Jr. R.G. 2002. Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering and Remote Sensing* 68(10):1041–1049.

Pontius Jr. R.G. and Batchu K. 2003. Using the relative operating characteristic to quantify certainty in prediction of location of land cover change in India. *Transactions in GIS* 7(4):467–484.

Pontius Jr. R.G., Gilmore R., Huffaker D. and Denman K. 2004. Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling* 179(4):445–461.

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

Rabin M. 1998. Psychology and economics. *Journal of Economic Literature* 36(1):1–46.

Radeloff V.C., Hagen A.E., Voss P.R., Field D.R. and Mladenoff D.J. 2000. Exploring the spatial relationship between census and land-cover data. *Society and Natural Resources* 13:599–609.

Rasul G., Thapa G.B. and Zoebisch M.A. 2004. Determinants of land-use changes in the Chittagong hill tracts of Bangladesh. *Applied Geography* 24:217–240.

Richards J.A. 1986. *Remote sensing digital image analysis*. Springer-Verlag, New York, USA.

Rindfuss R.R., Prasartkul P., Walsh S.J., Entwisle B., Sawangdee Y. and Vogler J.B. 2003. Household-parcel linkage in Nang Rong, Thailand: Challenges of large samples. In: Fox J., Rindfuss R.R., Walsh S.J. and Mishra V. (eds), *People and the environment: Approaches for linking household and community surveys to remote sensing and GIS*. Kluwer Academic Publishers, Norwell, USA.

Rindfuss R.R., Walsh S.J., Mishra V., Fox J. and Dolcemascolo G.P. 2003. Linking household and remotely sensed data, methodological and practical problems. In: Fox J., Rindfuss R.R., Walsh S.J. and Mishra V. (eds), *People and the environment: Approaches for linking household and community surveys to remote sensing and GIS*. Kluwer Academic Publishers, Norwell, USA.

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

Rosenblatt F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65:386–408.

Rounsevell M.D.A., Ewert F., Reginster I., Leemans R. and Carter T.R. 2005. Future scenarios of European agricultural land use: II. Projecting changes in cropland and grassland. *Agriculture, Ecosystems and Environment* 107(2-3):101–116.

Rudel T. and Roper J. 1997. The paths to rain forest destruction: Crossnational patterns of tropical deforestation, 1975–1990. *World Development* 25:53–65.

Rumelhart D., Hinton G. and Williams R. 1986. Learning internal representations by error propagation. In: Rumelhart D.E. and McClelland J.L. (eds), *Parallel distributed processing: Explorations in the microstructures of cognition*. MIT Press, Cambridge, USA.

Schabenberger O. and Pierce F.J. 2002. *Contemporary statistical models for the plant and soil sciences*. CRC Press, Boca Raton, USA.

Schneider L.C. and Pontius Jr. R.G. 2001. Modeling land use change in the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems and Environment* 85:83–94.

Seré C. and Steinfeld H. 1996. *World livestock production systems: Current status, issues and trends*. FAO Animal Production and Health Paper 127. FAO, Rome.

Serneels S. and Lambin E.F. 2001. Proximate causes of land use change in Narok District, Kenya: A spatial statistical model. *Agriculture, Ecosystems and Environment* 85:65–81.

Skapura D. 1996. *Building neural networks*. ACM Press, New York, USA.

Snijders T.A.B. and Bosker R.J. 1999. *Multi-level analysis: An introduction to basic and advanced multi-level modeling*. Sage, New York, USA.

Sokal R.R. 1977. Clustering and classification: Background and current directions. In: van Ryzin J. (ed), *Classification and clustering: Proceedings of an advanced seminar conducted by the Mathematics Research Centre*. University of Wisconsin, Madison, USA.

Sonneveld B.G.J.S. 2002. Formalizing the use of expert judgements for land degradation assessment: A case study for Ethiopia. Working paper WP 02–11. Centre for World Food Studies, Amsterdam, the Netherlands.

Sonneveld B.G.J.S. 2003. Formalizing expert judgements in land degradation assessment: A case study for Ethiopia. *Land Degradation and Development* 14:347–361.

Speybroeck N., Berkvens D., Mfoukou-Ntsakala A., Aerts M., Hens N., van Huylenbroeck G. and Thys E. 2004. Classification trees versus multinomial models in the analysis of urban farming systems in Central Africa. *Agricultural Systems* 80(2):133–149.

SPSS (Statistical Producers for Social Science). 2000. SPSS software and manual. Marketing Dept, SPSS Inc, Chicago, Illinois, USA.

Staal S.J., Baltenweck I., Waithaka M.M., de Wolff T. and Njoroge L. 2002. Location and uptake: Integrated household and GIS analysis of technology adoption and land use, with application to smallholder dairy farms in Kenya. *Agricultural Economics* 27:295–315.

Staal S. J., Kruska R., Baltenweck I., de Wolff T., Muriuki H., Thornton P. and Thorpe W. 2002. Integrated household and GIS analysis of smallholder systems: Market access, prices, and technology uptake on Kenyan dairy farms. Working document. ILRI, Nairobi.

StatSoft. 2003. *Electronic statistics textbook*. http://www.statsoft.com/textbook/stathome.html. Tulsa, USA.

Thompson D.M., Serneels S. and Lambin E.F. 2002. Land use strategies in the Mara ecosystem: A spatial analysis linking socio-economic data with landscape variables. In: Walsh S.J. and Crews-Meyer K.A. (eds), *Linking people, place and policy: A GIScience approach*. Kluwer Academic Publishers, Norwell, USA.

Thornton P.K., Kruska R.L., Henninger N., Kristjanson P.M., Reid R.S. and Robinson T.P. 2003. Locating poor livestock keepers at the global level for research and development targeting. *Land Use Policy* 20(4):311–322.

Tobin J. 1958. Estimation of relationships for limited dependent variables. *Econometrics* 26:24–36.

Turner II B.L. and Ali A.M.S. 1996. Induced intensification: Agricultural change in Bangladesh with implications for Malthus and Boserup. *Proceedings of the National Academy of Sciences* 93:14984–14991.

Turner II B.L., Kasperson R.E., Meyer W.B., Dow K.M., Golding D., Kasperson J.X., Mitchel R.C. and Ratick S.J. 1990. Two types of global environmental change: Definitional and spatial-scale issues in their human dimensions. *Global Environmental Change* 1:14–22.

Turner II B.L., Ross R.H. and Skole D.L. 1993. *Relating land use and global land cover change*. IGBP Report No. 24; HDP Report No. 5.

Turner II B.L., Skole D., Sanderson S., Fischer G., Fresco L.O. and Leemans R. 1995. *Land-use and land-cover change science/research plan*. IGBP Report No. 35; HDP Report No. 7. IGBP, Stockholm, Sweden.

Turner M.G., Dale V.H. and Gardner R.H. 1989. Predicting across scales: Theory development and testing. *Landscape Ecology* 3:245–252.

Veldkamp A. and Fresco L.O. 1996. CLUE-CR: An integrated multi-scale model to simulate land use change scenarios in Costa Rica. *Ecological Modelling* 91:231–248.

Veldkamp A. and Fresco L.O. 1997. Reconstructing land use drivers and their spatial scale dependence for Costa Rica (1973 and 1984). *Agricultural Systems* 55(1):19–43.

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

Verburg P.H. and Chen Y. 2000. Multiscale characterization of land-use patterns in China. *Ecosystems* 3:369–385

Verburg P.H., de Groot W.T. and Veldkamp A. 2003. Methodology for multi-scale land-use change modelling: Concepts and challenges. In: Dolman A.J., Verhagen A. and Rovers

C.A. (eds), *Global environmental change and land use*. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Verburg P.H., de Koning G.H.J., Kok K., Veldkamp A. and Bouma J. 1999. A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. *Ecological Modelling* 116:45–61.

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

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

Verburg P.H., Schot P., Dijst M. and Veldkamp A. 2004. Land use change modelling: Current practice and research priorities. *Geojournal* 61(4):309–324.

Verburg P.H., Soepboer W., Limpiada R. and Espaldon V. 2002. Modeling the spatial dynamics of regional land use: The CLUE-S model. *Environmental Management* 30(3):391–405.

Verburg P.H. and van Keulen H. 1999. Exploring changes in the spatial distribution of livestock in China. *Agricultural Systems* 62:51–67.

Visser H. 2004. *The map comparison kit: Methods, software and applications*. Report 550002005/2004. RIVM, Bilthoven, the Netherlands.

von Thünen J.H. 1966. Der isolierte staat in beziehung der landwirtschaft und nationalökonomie. In: Hall P. (ed), *Von Thünen's isolated state*. Pergamon Press, Oxford, UK.

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

Walker R. and Solecki W.D. 2004. Theorizing land-cover and land-use change: The case of the Florida Everglades and its degradation. *Annals of the Association of American Geographers* 94(2):311–238.

Walsh S.J., Bilsborrow R.E., McGregor S.J., Frizelle B.G., Messina J.P., Pan W.K.T., Crews-Meyer K.A., Taff G.N. and Baquero F. 2003. Integration of longitudinal surveys, remote sensing time series, and spatial analyses. In: Fox J., Rindfuss R.R., Walsh S.J. and Mishra V. (eds), *People and the environment: Approaches for linking household and community surveys to remote sensing and GIS*. Kluwer Academic Publishers, Norwell, USA.

Walsh S.J., Crawford T.W., Welsh W.F. and Crews-Meyer K.A. 2001. A multiscale analysis of LULC and NDVI variation in Nang Rong District, northeast Thailand. *Agriculture, Ecosystems and Environment* 85:47–64.

Watson M.K. 1978. The scale problem in human geography. *Geografiska Annaler* 60B:36–47.

Weiss E., Marsh S.E. and Pfirman E.S. 2001. Application of NOAA-AVHRR NDVI time-series data to assess changes in Saudi Arabia's rangelands. *International Journal of Remote Sensing* 22(6):1005–1027.

Wester-Herber, M. 2004. Underlying concerns in land-use conflicts--the role of place-identity in risk perception. *Environmental Science and Policy* **7**(2): 109-116.

Whittle P. 1970. *Probability*. Penguin Books, Harmondsworth, UK.

Wood C.H. and Skole D. 1998. Linking satellite, census, and survey data to study deforestation in the Brazilian Amazon. In: Liverman D., Moran E.F., Rindfuss R.R. and Stern P.C. (eds), *People and pixels: Linking remote sensing and social science*. National Academy Press, Washington, D.C., USA.

Glossary

Autocorrelation: Autocorrelation is the correlation (relationship) between members of a time series of observations and the same values at a fixed time interval later.

Bayesian statistics: Bayesian analysis is an approach to statistical analysis that is based on Bayes Law, which states that the posterior probability of a parameter *p* is proportional to the prior probability of parameter *p* multiplied by the likelihood of *p* derived from the data collected.

Bias: Bias refers to how far the average statistic lies from the parameter it is estimating, that is, the error that arises when estimating a quantity. Errors from chance will cancel each other out in the long run, those from bias will not.

Canonical correlation: Canonical correlation investigates the relationship between two sets of variables (it is used as either a hypothesis-testing or exploratory method).

Categorical data: A set of data is said to be categorical if the values or observations belonging to it can be sorted according to category. Each value is chosen from a set of non-overlapping categories, e.g. colour or land use types.

Censored observations: Observations are referred to as censored when the dependent variable of interest represents the time to a terminal event, and the duration of the study is limited in time.

Classification: The ordering or arrangement of objects into groups or sets on the basis of their relationships.

Discrete data: A set of data is said to be discrete if the values or observations belonging to it are distinct and separate, i.e. they can be counted (1, 2, 3, ...).

Endogenous variable: An endogenous variable is a variable that appears as a dependent variable in at least one equation in a structural model.

Exogenous variable: An exogenous variable is a variable that never appears as a dependent variable in any equation in a structural model.

Extent: Extent is the size of the spatial, temporal, quantitative or analytical dimensions of a scale.

Factor analysis: Factor analysis is a technique (i) to reduce the number of variables; and (ii) to detect structure in the relationships between variables, that is to classify variables.

Goodness of fit: Goodness of fit is the degree of agreement between an empirically observed distribution and a mathematical or theoretical distribution.

Heteroscedasticity: Heteroscedasticity describes a data sample or data-generating process in which the errors are drawn from different distributions for different values of the independent variables.

Hierarchy: A hierarchy is a conceptually or causally linked system of grouping objects or processes along an analytical scale.

Homoscedasticity: Homoscedasticity describes a statistical model in which the errors are drawn from the same distribution for all values of the independent variables.

Land cover: The observed physical cover, as seen on the ground or through remote sensing, including the vegetation (natural or planted) and human constructions (buildings, etc.), that cover the earth's surface. Water, ice, bare land, and salt flats or similar non-vegetated surfaces are included in land cover.

Land use: A series of operations and associated inputs on land, carried out by humans, with the intention of obtaining products and/or benefits through using land resources.

Least squares: The method of least squares is a criterion for fitting a specified model to observed data. For example, it is the most commonly used method of defining a straight line through a set of points on a scatterplot.

Level: Level is the unit of analysis that is located at the same position on a scale.

Logistic regression: Logistic regression is a regression model for binary (dichotomous) outcomes, for which the data are assumed to follow binomial distributions with probabilities that depend on the independent variables.

Multicollinearity: A term used to describe the condition when one or more variables from which the respective matrix was computed are linear functions of other variables.

Multinomial regression: Multinomial logit regression models are extensions of the standard logit regression models to the case where the dependent variable has more than two categories.

Neural networks: Neural networks are analytic techniques modelled after the (hypothesized) processes of learning in the cognitive system and the neurological functions of the brain and capable of predicting new observations (on specific variables) from other observations (on the same or other variables) after executing a process of so-called learning from existing data.

Principal component analysis: A linear dimensionality reduction technique, which identifies orthogonal directions of maximum variance in the original data, and projects the data into a lower-dimensionality space formed of a subset of the highest-variance components.

Probability: A probability provides a quantitative description of the likely occurrence of a particular event. Probability is conventionally expressed on a scale from 0 to 1; a rare event has a probability close to 0, a very common event has a probability close to 1.

Regression equation: A regression equation expresses the relationship between two (or more) variables algebraically. It indicates the nature of the relationship between variables. In particular, it indicates the extent to which you can predict some variables by knowing others, or the extent to which some are associated with others.

Scale: Scale is the spatial, temporal, quantitative or analytical dimensions used to measure and study any phenomenon.

Spatial analysis: The study of spatial relationships between geographic features by using the processes of modelling, examination and interpreting, for the purpose of evaluating, estimating, predicting and understanding these relationships.

Spatial autocorrelation: The property of random variables to take values over distance that are more similar or less similar than expected from randomly associated pairs of observations due to geographic proximity (spatially dependent).

Stepwise regression: A model-building technique which finds subsets of predictor variables that most adequately predict responses on a dependent variable by linear or non-linear regression, given the specified criteria for adequacy of model fit.

Validation: Validation is the comparison of a conceptual model to the real system.

Published by:

International Livestock Research Institute
Nairobi
Kenya

and

LUCC Focus 3 Office
Wageningen University
The Netherlands