Available online at www.sciencedirect.com
ScienceDirect
Land Use Policy 24 (2007) 521–530
Land Use Policy
www.elsevier.com/locate/landusepol

Issues in spatially explicit statistical land-use/cover change (LUCC) models: Examples from western Honduras and the Central Highlands of Vietnam

Darla K. Munroe[a,b,*], Daniel Müller[c]

[a]Department of Geography, The Ohio State University, 1036 Derby Hall, 154 North Oval Mall, Columbus, OH 43210 1361, USA
[b]Department of Agricultural, Environmental and Development Economics, The Ohio State University, 1036 Derby Hall, 154 North Oval Mall, Columbus, OH 43210 1361, USA
[c]Institute of Agricultural Economics and Social Sciences, Humboldt University Berlin, Luisenstr. 56, 10099 Berlin, Germany

Received 8 March 2005; received in revised form 31 August 2005; accepted 9 September 2005

Abstract

Land-use/cover change (LUCC) results from the complex interaction of social, ecological and geophysical processes. Land users make decisions about their environment that are governed and influenced by political and institutional constraints at local, regional, national and international levels. Statistical analysis of LUCC phenomena is one powerful tool due to its ability to test theoretical assumptions, rank relative factors, and yield rigorous hypotheses test. However, due to the complex nature of coupled human–environment systems, LUCC statistical modeling presents conceptual as well as technical challenges. Careful consideration of these challenges, as well as implementing approaches to deal with them, is necessary in order for such models to inform policy and practice. As examples, we present illustrations of two statistical models of land use in the mountains of western Honduras and the Central Highlands of Vietnam.

Keywords: Land-use/cover change; Statistical modeling; Sampling; Scale; Geosimulation; Categorical data analysis

Introduction

Land-use/cover change (LUCC) results from the complex interaction of social, ecological and geophysical processes. Land users make decisions about their environment that are governed and influenced by political and institutional constraints at local, regional, national and international levels. Despite much work studying LUCC and LUCC processes, a definitive understanding of LUCC is elusive, and there is no universal set of guidelines for policy that could mitigate LUCC across the board; the local situation and context are critical (Geist and Lambin, 2002).

LUCC researchers are increasingly realizing that an integrated set of techniques is required to understand these phenomena. In addition to process-based modeling and the use of case studies in examining LUCC (Veldkamp and Lambin, 2001), statistical analysis is a powerful tool due to its ability to test theoretical assumptions, rank relative factors, and yield rigorous hypotheses tests. Statistical analyses of LUCC present the opportunity to link the general and the specific, in that one can effectively identify the relative contribution of broader factors such as institutional and market forces, while controlling for factors particular to the location. However, modeling LUCC processes requires integrating data across space, time and level of analysis, and there remain barriers to broader implementation of best-practice techniques (Rindfuss et al., 2004).

The underlying objective of this paper is to examine statistical tools and techniques most useful to studying LUCC, and to assess current best practice. The structure of

*Corresponding author. Department of Geography, The Ohio State University, 1036 Derby Hall, 154 North Oval Mall, Columbus, OH 43210 1361, USA. Tel.: +1 614 247 8382; fax: +1 614 292 6213.
E-mail addresses: munroe.9@osu.edu (D.K. Munroe), d.mueller@rz.hu-berlin.de (D. Müller).

0264-8377/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.landusepol.2005.09.007

this paper is as follows. We begin with examining conceptual challenges researchers face in the implementation of LUCC statistical models and the myriad technical difficulties that arise as a result (Rindfuss et al., 2004). We then explore the means available to overcome these problems, with reference to key LUCC statistical models from the current literature. Finally, we present two empirical examples, from study areas in western Honduras and the Central Highlands of Vietnam that make use of some of the issues. We end with a final discussion of the main impediments and challenges to broader implementation of best practice techniques.

Conceptual challenges

This section examines the tradeoffs that are implicit in the empirical representation of LUCC processes and the implications for statistical analysis. LUCC is necessarily conceptualized as comprising a dynamic, coupled human–natural system. Land-use change, such as urbanization or deforestation, can have profound influences on ecosystems, but ultimately, ecosystem responses influence further land-use change. For example, areas proximate to a lake may be prime real estate, spurring development, but if water quality in the lake declines below some threshold (due to pollution and erosion processes), further development would be affected. These dynamics tend to be complex and nonlinear. Effectively studying one part of the system (e.g., drivers of land market activity) can yield insights, but precludes incorporation of feedback effects. Therefore, regardless of analytical methodology, it is important to conceptualize LUCC as a coupled system. As such, careful thought must be given to how studies of various components of the system can be integrated.

LUCC as a discrete process

LUCC is most often represented as a categorical phenomenon, but categorical formulations can obscure the true processes at work for both land use and land cover. Observed land-use patterns are generally theorized to reflect the land user's choice, i.e., that of a particular land-use alternative, selected over all others. Thus, land use is a discrete process representing the presence or absence of a particular land-use category. The underlying decision process behind this choice; i.e., that agents allocate their production factors to maximize their utility or profit given prevailing constraints and preferences, is unobserved. Beyond a certain threshold this decision process is transformed into a discrete phenomenon. Following this logic, the land-use choice becomes a latent variable, and much of the information regarding this choice is not directly measurable (Anselin, 2002). For example, when the opportunity cost of forestland changes due to an increase in the world coffee price, a land user may be motivated to clear forest, but the exact decision calculus is often not directly observed.

Empirically, the discrete representation of land use is often proxied by a discrete representation of land cover. The threshold to obtain the discrete representation of land-cover from remote-sensing data is imposed by a supervised or unsupervised classification algorithm or by an image interpreter. The remote-sensing community has long noted the difficulties due to loss of information in the classification process (Southworth et al., 2004), and this loss of information complicates modeling as well. Furthermore, land-cover data may be a poor proxy to capture real-world land use, e.g., in the case of grassland cover used as intensive pasture or rangeland.

Social versus physical space

Land-use change is fundamentally a spatial process, resulting from the collective outcome of myriad processes: socioeconomic, institutional, biophysical and ecological. One conceptual challenge in integrating these processes empirically is the accurate representation of the spatially related factors and processes in question, including their variation. Space and spatial relationships are relevant to most economic phenomena, but can be conceptually challenging to specify in practice (van der Veen and Otter, 2001). Social scientists traditionally utilize units like households, villages or administrative regions as their basic research object. Natural scientists, on the other hand, are concerned with units representing the ecological processes under question and often look at finer spatial variation, and here frequently continuous spatial surfaces, to study processes of their interest (Irwin and Geoghegan, 2001).

Scale considerations

Land-use processes also occur at a variety of different levels, from the local to global. The resolution at which spatial interactions occur may be specific to the process. Moreover the spatial extent, or boundaries, may vary according to the process. For example, land users often interact strategically in land-use decisions. Residential conversion in one location may make further conversion more likely, or may repel conversion if congestion becomes a problem. Forest clearing at a frontier may follow an entropy process; i.e., spreading out from an edge. Land-use change is a fundamentally local process, but it is nested in a structure of hierarchical decision-making (Moran et al., 2002). At the local level, one can consider such factors as changes in population density, technical innovation, or changes in agricultural production systems (e.g., switching from a subsistence crop to a cash crop). Incentives and regulations imposed at an administrative level, such as policy changes, are translated and realized by human decision-making at the local level. Likewise, changes at a local level have the potential to influence changes at a regional level. For example, rapid land-use change in one locality could lead to change in regional policy. Therefore, successful statistical modeling of LUCC must carefully

consider how the formulation and representation of the processes to be studied will qualitatively affect the modeling process, and one must note that inferences made are generally framework-specific.

Empirical challenges

In addition to conceptual challenges, LUCC statistical models generally suffer from a variety of integration and specification problems. Initially, multiple regression was the primary means to link hypothesized driving forces of land use to observed land-cover patterns. More recently, researchers have attempted to employ statistical techniques as complementary tools to integrated, mixed method approaches (Brown et al., 2000; Dutcher et al., 2004; Parker et al., 2003). Multiple regression is most useful for empirically testing theoretical relationships among a set of variables, conducting rigorous hypothesis testing, and ranking the relative contribution of factors. It is much less useful for forecasting purposes, or otherwise making out of sample predictions. Careful specification testing and correction must be made in statistical analyses when there are substantial patterns of dependence or heterogeneity across space and/or time. Lastly, spatial processes and data are likely to yield scale-dependent results; that is, interactions among a set of variables may appear different depending on the level of analysis at which they are modeled. In this section, we review some of the common pitfalls and challenges of statistical analyses, with special regard to tools and techniques to deal with such problems.

Statistical analysis of LUCC processes presents significant data challenges, reviewed in Nelson and Geoghegan (2002). A key feature of spatially explicit econometric estimations is that all data elements contain locational attributes, i.e., are georeferenced. Spatially explicit models require special techniques to estimate regression parameters and—depending on data resolution—to yield location-specific results. The majority of the economic models on land-use change link survey or census data to spatially explicit environmental data, usually generated using geographical information systems (GIS), in order to estimate the effects of exogenous and predetermined factors on land-cover change. Ideally, land-use change modeling should yield insights, and help to confirm or reject theoretical hypotheses, on the most relevant drivers of change (Veldkamp and Lambin, 2001).

Spatial data integration

The integration of disparate land-use processes is no trivial matter (Nelson and Geoghegan, 2002). Land use is the result of a set of coupled social and ecological systems. Policy analysis often requires working at coarse scales because the base unit of analysis is the administrative area in question, which may preclude the consideration of ecological units as such. A prime example of this problem is mismatch in managing watershed-level processes through local residential districts (Bockstael and Bell, 1998). The data model most frequently used in LUCC models is a raster, or grid model. Remotely sensed data are generally captured in this format as satellite images are comprised of pixels. Furthermore, gridded data is the most efficient way to store and manipulate large amounts of spatially continuous data. A disadvantage of grid modeling is the artificial imposition of the rectangular boundaries of the grid cells; the grid cell or the pixel in remote sensing relates to the minimum mapping unit of the sensor, and does not relate to a behavioral decision-making unit nor to landscape elements in any straightforward way (Fisher, 1997; Rindfuss et al., 2004).

Cross-process data integration can be particularly difficult, as the units of analysis, spatial extent, resolution and relevant time frame may not coincide (Bockstael, 1996; Verburg et al., 2002). Walsh et al. (2001), Crawford (2002) and Müller and Munroe (2005) all develop techniques to integrate discrete socioeconomic and continuous spatial data at the village level. Relevant processes may not be stationary (i.e., independent) across space, time, or level of the system. Thus, there can arise significant cross-scale, spatial and temporal heterogeneity or dependence. These problems require the use of sampling strategies that filter such dependence (Elhorst, 2001), or other more explicit correction for spatial autocorrelation (Rindfuss et al., 2004), including the use of spatial regression techniques (Anselin, 2002). Irwin and Geoghegan (2001) state that spatial data must be employed creatively, and their use should at best capture the power of the underlying spatial process.

In practice, any number of tradeoffs must be made. Household-level statistical analysis of farm practices in developing countries have long suffered from having to use proxy variables to represent spatially related factors, because more precise measures have been unavailable, or only available at great cost. Geophysical variables like altitude and slope are often averaged per plot, farm or village. Agroecological zones are generally only roughly differentiated. Sampled points of a continuous phenomenon (such as rainfall) must be interpolated to represent a surface, or aggregated to correspond to administrative or economic units as entities. This manipulation induces measurement error, often left unmodeled (Anselin, 2002). All other spatially varying socioeconomic effects are usually captured by dummy variables or distance measures for the location of objects without a consideration of spatial effects. Unfortunately, the use of these proxies has limited the ability of researchers to quantify the directions and magnitude of the effects resulting from spatially explicit factors on farmer's choice of technology and on their allocation of resources. Examples of studies that have creatively dealt with these issues with varying degrees of success include, among others, Southgate et al. (1991), Elnagheeb and Bromley (1994), Godoy et al. (1997), and Bergeron and Pender (1999). Ultimately, a successful study is one in which the authors have a good qualitative sense of

how these technical challenges have influenced model results.

Scale dependence

Scale dependence, or that relationships among a set of variables appear to change depending on the level of analysis at which they are observed, is ubiquitous in LUCC analyses (Walsh et al., 1999; Veldkamp and Lambin, 2001). Scale effects often arise in LUCC studies due to the complex nature of the linked socioecological systems under study. In addition, problems that arise simply from the manipulation of spatial data can also lead to the appearance of scale effects. Scale effects can arise from unit of analysis mismatch, the use of aggregated data, differing extents across social vs. ecological processes, the modifiable areal unit problem (MAUP), ecological fallacy, and lastly, self-organizing or complex systems (Atkinson and Tate, 2000; Cressie, 1996; Jelinski and Wu, 1996). It can get even messier: cross-scale interaction or hierarchical dependence creates locally spatially dependent processes (Veldkamp et al., 2001). Ideally, theory should be a guide for interpreting the factors and processes guiding multilevel processes. However, to date, there is no test or set of tests to distinguish how and why scale effects arise in a given dataset; it remains an empirical issue. Researchers must take care to recognize such processes, if present, and attempt to assess their qualitative effect on model results. Robustness tests for various modeling approaches and sensitivity analysis across scales allow assessing empirical results and model strength (Evans and Kelley, 2004). Explicitly integrating various scales in the analysis of econometric land-use models can control for influences on and yield insights into relations on various levels of decision processes (Walsh et al., 1999, 2001). Hoshino (2001) and Polsky and Easterling III (2001) develop multilevel models that explicitly take cross-level interactions into account. Müller and Munroe (2005) control for fixed effects on village level by accounting for the lack of independence across villages in a pixel-level econometric set-up.

New developments

There have also been many promising developments in statistical methodology that have greatly improved on prior limitations. New techniques in Bayesian hierarchical modeling have enabled complex, multilevel models that do not otherwise have a straightforward analytical solution; process-based models can be linked via their observed associations. Bayesian hierarchical modeling approaches can potentially overcome the problems of spatial/temporal dependence, differing data resolutions and multiplicative error resulting from uncertainty from various data sources. In the Bayesian literature, assimilation of spatially or temporally misaligned data is known as melding, which can overcome a host of potential measurement biases, and missing data (Poole and Raftery, 2000; Sneddon, 2000).

Expansions to the traditional categorical representations include mixed logit models or spatial probit models for explicitly modeling the structure of spatial dependence. Some researchers have borrowed techniques from other disciplines, such as the social interactions literature in sociology (Irwin and Bockstael, 2004) to exploit their rich possibilities for representing interactions. Finally, geocomputation and geosimulation techniques have greatly expanded the analytical and inferential properties of spatial data. Possibilities range from bootstrapping of standard error estimates, to permutation techniques for spatial analysis, to pure inference (Fotheringham et al., 2001).

Model evaluation

Finally, there must be a standard by which one can determine whether a statistical LUCC model is a good one, which may not be as straightforward in a spatially explicit context. Veldkamp and Lambin (2001) state that a LUCC model is useful if it can predict the future or explain the past. Statistical models are limited in their ability to make out-of-sample predictions, though forecasting techniques could potentially be used to study LUCC processes. There are two distinct, yet related, criteria one might consider in evaluating statistical LUCC models: (1) the overall explained variation of the regression; and (2) the significance of regression parameters, particularly when the model is used to test theoretical precepts. It may be that a model has high overall explanatory value, even if the amount of explained variation is low. For example, Munroe et al. (2004) discuss how all other land-use incentives are influenced by elevation. An examination of relative altitude within their study area quickly determines which areas are likely to be forested, even though more precise knowledge on the locally based drivers of clearing is needed at low elevations in order to predict clearing on a pixel-by-pixel basis.

There are a few ways to measure the overall fit of a regression when the dependent variable is categorical (binomial or multinomial). Traditionally, limited dependent variables can be assessed with a *Pseudo-*R^2, which compares the ratio of the explained variation with the specified regressors (independent variables), as compared to a null model with a simple constant term (Maddala, 1983). This approach has been criticized because the log-likelihood value is sensitive to the number of regressors included; spurious regressors may increase the apparent explained variation. An increasingly common statistical measure for model comparison is the Akaike's information criterion (AIC). Aspinall (2004) employs multimodel inference using AIC to rank the plausibility of the models over multiple time periods. Different rankings of model performance over time allows for the assessment of land-use drivers changing in strength or influence over time.

Cross-tabulations of fitted vs. observed land cover, also called confusion matrix or prediction matrix, provide a tabulated overview of the 'correctly' estimated observations. In the binary case most researchers adopt a probability threshold of 50% as the cut-off point while in a multinomial setting the threshold is much less apparent as the distribution of probability values may be centered around a single value for a number of locations. Researchers usually use a maximum probability assignment rule, although other assignments for predicted values are possible (Nelson and Geoghegan, 2002). An important criticism of simple cross-tabulation statistics has been made by Pontius (Pontius, 2002; Pontius and Schneider, 2001; Pontius et al., 2004). A model may perform well in replicating aspects of the landscape, such as amount, degree or even pattern of land use, but still not predict specific pixels correctly. Pontius suggests separately considering quantity (i.e., proportion of the landscape allocated in each land-use type) vs. location (i.e., predicting the land-use outcome of a particular pixel). Therefore, two models may be equally bad at predicting the exact location of a scarce land-use type, but one model may be closer to predicting the overall level of that land use within the landscape. Pontius suggests employing the relative operating characteristic (ROC), derived from a series of contingency tables comparing actual and fitted changes, or the distinct tabulation of errors of quantity, and given quantity, errors of location.

Predicted probabilities can be spatially examined by mapping predicted probabilities. If a sample was used in estimation, predicted probabilities are obtained by applying the estimated coefficients on the entire dataset. The resulting prediction maps yield important insights into the locational accuracy of the model. Prediction maps facilitate geovisualization and help to assess the accuracy of the predictions for the entire research area. In-depth investigation of fitted land-use values further help to clarify the prediction potential of variables, processes not captured with the data at hand, and locations where the model does not perform appropriately.

Analysis

In the next sections, we demonstrate techniques that could potentially extend the inferential capabilities of a multinomial logit using two empirical examples. The first uses repeated spatial sampling in order to examine whether estimated areas at risk are robust. The second explores what additional information can be gleaned by examining estimated probabilities in greater detail. The analysis builds on prior work on Western Honduras (Munroe et al., 2002, 2004; Southworth et al., 2004) and Vietnam (Müller and Zeller, 2002; Müller, 2003; Müller and Munroe, 2005). The empirical model posits that observed land cover is a function of underlying land-use incentives, which include geophysical factors that influence agroclimatic suitability for particular land uses (such as slope, elevation, soil quality, etc.), as well as the relative influence of market accessibility (i.e., effective distance to roads and centers of exchange).

We assume that land is devoted to its highest valued use, or the use that brings the greatest rent (i.e., profit or utility) to the user. Chomitz and Gray (1996) demonstrated that relative land rent is a function of the collective impacts of distance to markets and geophysical variation as follows:

$$\ln R_{ik} = \alpha_{0k} + \alpha_{1k} D_i + \alpha_{2k} G_i, \quad (1)$$

where R represents rent derived from land use k at location i, D represents the cost of access between location i and relevant centers of exchange, and G represents the geophysical characteristics of location i. We econometrically estimate Eq. (1) by specifying it as follows:

$$\ln R_{ik} = \alpha_{0k} + \alpha_{1k} D_i + \alpha_{2k} G_i + \ldots + u_{jk} \equiv \beta_i X_k + u_{ik}, \quad (2)$$

where X is the set of independent variables, α the vector of associated parameters, and u the random Weibull-distributed disturbance term. Eq. (2) can be expressed as a multinomial logit as follows:

$$PROB(i_k) = \frac{\exp(\beta_i X_k)}{\sum_j \exp(\beta_i X_j)}, \quad (3)$$

where the probability of land-cover category k at location i is estimated relative to the probability of all possible land covers j for j not equal to k. Details and full results can be found in Müller and Zeller (2002) and Munroe et al. (2004).

Geosimulation of fitted values

Econometric models of land cover and land-cover change are of most use for policy purposes when they are able either to predict changes likely to occur, or explain changes that happened in the past (Veldkamp and Lambin, 2001). In particular, it may often be of interest for policy makers to determine where future land-cover changes may be likely, and to convey information about the direction and strength of major factors influencing land cover that can potentially be altered by policy interventions.

However, the fitted values of any particular regression analysis depend on the observations selected to be included in the analysis. For the western Honduras study region, the source data consist of roughly 1,270,000 pixels. In order to estimate an econometric model, it is generally necessary to begin with some sample of the total number of pixels to reduce computational time. Moreover, spatial sampling can be very helpful when there is a substantial amount of spatial autocorrelation in the underlying data (Munroe et al., 2002; Müller and Zeller, 2002; Müller and Munroe, 2005). However, the estimated coefficients may thus depend on those pixels that are sampled, and it is reasonable to expect that outliers or extreme observations may have an impact on the overall analysis. In addition, a multinomial logit model is nonlinear and

requires maximum likelihood estimation, so even when the same data are used multiple times, there may be some variation in the results. To lend additional credence to a model that is designed to yield insights regarding where future land-cover changes are likely, additional computational verification of the validity of model results is helpful.

For the purpose of an illustrative demonstration regarding how simulation approaches may be of use in LUCC research, the following analysis was conducted. A multinomial logit model was estimated for the western Honduras dataset. The dependent variable was land-cover-change trajectories from the period 1987–1996, forming eight unique classes of land-cover-change trajectories (Mertens and Lambin, 2000; Munroe et al., 2004). The independent variables included slope, elevation and distance to local and regional markets. Fitted values were generated by assigning each pixel to the land-cover-change trajectory for which the predicted probability value was the highest. The full dataset was read into Matlab, and a spatially stratified sample of 5000 observations was chosen 1000 times. The sampling scheme was as follows: the middle pixel in a 25×25 window was selected, but the starting point for the initial window varied via a number between 1 and 5, to begin the sampling somewhere in the first few rows and columns. The multinomial logit was run 1000 times with each of the subset of pixels. Each time, those observations for which the observed land cover was forest in the last period (1996), but whose fitted values were nonforest, were identified as areas possibly at risk for clearing in future periods. The number of times that a particular forested pixel had a fitted value of nonforest was tabulated. Fig. 1 displays the results of this analysis.

This technique may serve as a guide for particular regions within the study area that might warrant further investigation regarding their risk of future clearings. First, there were several forested pixels in the analysis whose fitted values were nonforest multiple times, so these individual pixels indicate they might be at risk. In addition, there appear to be specific regions within the study area whose forested pixels pop up as nonforest. There is one isolated incidence of a fitted value of nonforest in the central east region of the map, but no other such values in this general region; therefore, this one instance of a fitted nonforest value may not be highly significant. Conversely, the area in the northwest portion of the study appears to have a highly significant cluster of forested pixels fitted as nonforest. This particular region of the study area represents an area that is suitable for coffee production, and is close to roads and settlements, so its significance is

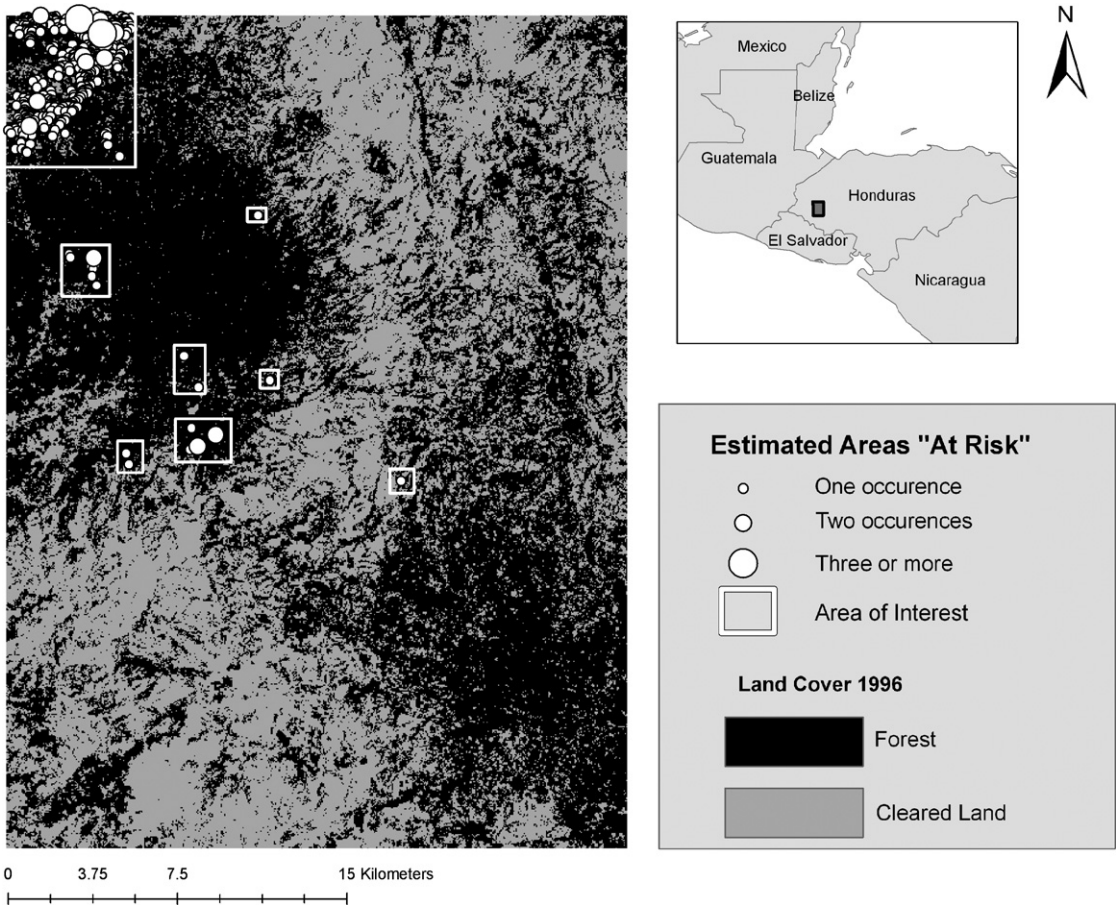

Fig. 1. Estimated areas at risk for forest clearing (areas currently forested, but predicted as nonforest), resulting from 1000 simulations of a multinomial logit model.

confirmed by this area's suitability for clearing according to the theoretical model.

An additional step from this analysis might be to evaluate point locations of forested pixels with fitted values as nonforest via point pattern analysis (e.g., quadrat analysis or nearest-neighbor statistics) to determine the extent of clustering, or whether these fitted values are spatially random. This initial experiment demonstrates great potential for communication to policy makers and planners regarding areas "at risk" that may warrant further examination, and potentially greater forest protection.

Exploring predicted probabilities

A second issue explored in this paper relates to the broader issue of analyzing statistical model results. Generally speaking, when dealing with categorical LUCC, categorical regression models are most often used. However, underlying outputs of econometric estimation techniques such as probit or logit do not translate directly into fitted values. Instead, these models yield predicted probabilities for each category (i.e., land-cover class) specified. Most commonly, an assignment rule is used to translate those probability values into a fitted value, such as maximum probability. Using maximum probability, each pixel would be assigned to that class for which the predicted probability is the highest. This approach, while useful, can be problematic. First, in assigning each pixel to one land-cover outcome, information is lost regarding how likely the other possible outcomes might have been. It may be that one particular outcome was only higher than the other possibilities by some small amount. Secondly, predicted probabilities for each observation are point estimates, and do not contain any information about the underlying uncertainty associated with each coefficient. Generating results for a range of outcomes including the information contained in standard errors of each β coefficient would help with this problem. Lastly, for all pixels that are assigned to a particular categorical outcome, there is no particular way of knowing which of those pixels have a high probability of that outcome vs. those whose relative probability value was much lower. Particularly, as the results of these models are used to inform policy regarding past and future land-use changes, the use of categorical maps of outcomes are, at best, incomplete, or even misleading. Some notable studies that have attempted to move beyond this limitation are Serneels and Lambin (2001), who in addition to generating maps of fitted values for each land-cover outcome, also generated maps of probability surfaces as a means to identify areas at risk. Their interpretation was, for example, that if a particular pixel had forest as the most probable land cover, but this probability was relatively low, it may be more at risk for future clearing. Müller (2003) summarized trends in probability values by land-use outcomes as a further means of assessing model fit.

The work presented here demonstrates some additional information that can be gleaned from categorical regression analysis using data from Dak Lak province in the Central Highlands of Vietnam (Müller and Zeller, 2002; Müller and Munroe, 2005). In this analysis, employing the same estimation techniques described in Eqs. (1) to (3), the dependent variable (categorical outcome) was specified as the probability of observing rice paddy (intensive cultivation), mixed agriculture (annual and perennial mixed cultivation) or natural and semi-natural landscapes (including forest, bush and grasslands) in the study region. The range of estimated probability values associated with each class are displayed. One can spatially examine areas by their estimated probability values to see whether particular areas have a land-use class whose predicted value is clearly distinct from the other two categories vs. areas where there is greater indeterminacy. For example, in Müller and Zeller (2002) it was noted that edge areas (areas on the border of two distinct, yet homogeneous patches) are more likely to have low predicted probabilities across classes, which could either indicate simple image interpretation error due to pixel-edge effects, or the higher likelihood of land-use conversions at the frontiers of land-use categories. More systematically, areas where predicted probabilities are low for all classes could warrant further explanation. These could be areas at risk for future changes (as Serneels and Lambin, 2001), or simply areas where some other factor, not included in the model, explains what is truly going on.

Table 1 presents summary statistics for the probability values of all classes, and Fig. 2 displays these values graphically. According to the maps, there are some regions where rice paddy and mixed agriculture are both fairly likely, and a maximum probability assignment rule (which in this case would be the land-use class for which the probability value is at least greater than 0.33), could result in much confusion pixel-by-pixel, though the average suitability by region may be fairly accurately measured by the independent variables. This finding may indicate

Table 1
Trends in estimated probability values, multinomial logit model of land use in Vietnam

$N = 558{,}227$	Probability values		
	Mixed agriculture	Rice paddy	Natural land
Mean	0.24	0.08	0.68
Median	0.04	0.00	0.96
Std. error	0.30	0.18	0.40
Minimum	0.00	0.00	0.00
Maximum	0.98	0.91	1.00
Fitted values (max probability rule)	139,394	41,080	376,701
Number of upper outliers*	49,964 (9%)	122,856 (22%)	0

*Outliers defined using a $\pm 1.5 \times$ IQR rule.

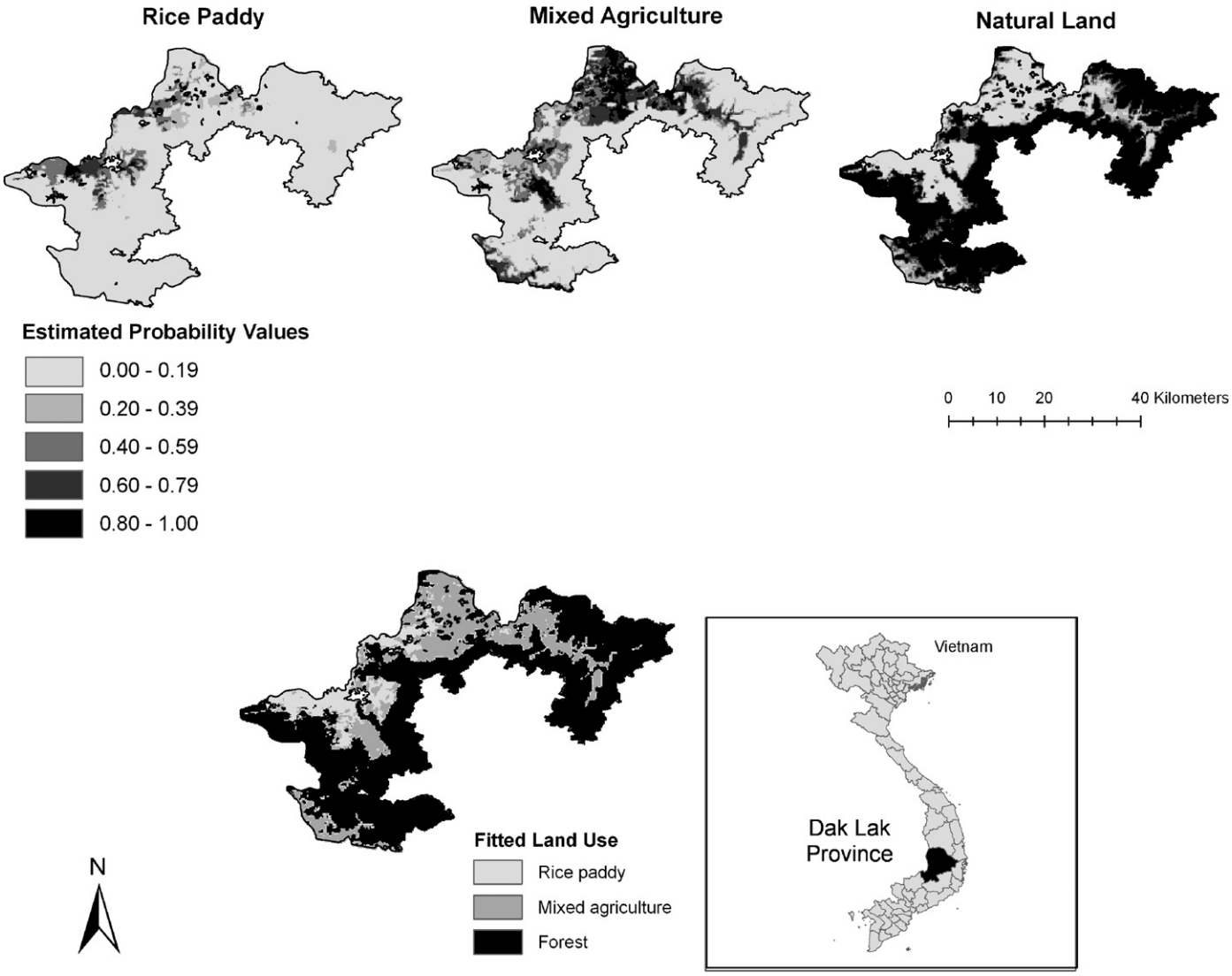

Fig. 2. Probability maps for a multinomial logit model in Dak Lak Province, Vietnam.

that either key variables promoting paddy over mixed agriculture are missing, or the decision to convert land into rice paddy follows an interaction effect that is unmodeled. For natural land, the underlying suitability corresponding to the independent variables is much clearer. Overall, it appears that the multinomial logit picked up a clear separation in the patterns within the covariates relating to natural land pixels, vs. to mixed agriculture and rice paddy. The minimum values across all classes are 0, but the mean and median for natural land pixels are well above 0.5.

Discussion

The analysis presented here indicates that statistical models, while representing powerful tools for exploring patterns between land-use/cover processes and their hypothesized covariates, are subject to challenges and pitfalls in their use. Particularly if such models are used to develop policy prescriptions and priorities, it is important to be able to assess qualitatively how these pitfalls may lead to misleading inferences about the nature of these local processes. Serneels and Lambin (2001) provide a nice discussion of outcomes from statistical LUCC models; statistical models can yield insights on location (*where* is change likely to happen?) or explanation (*why* is change happening?), and both these products may be of use to policy makers, but likely in different ways.

A key research priority should be to bring the ever-increasing set of tools, collectively known as geocomputational approaches, into mainstream statistical analysis. Some of the most limiting and confounding aspects of statistical analyses (i.e., the hampered ability to represent complex, nonlinear processes) can potentially be overcome by incorporating some of these approaches. At the very least, we can increase the inferential abilities of these models in a spatial domain by relaxing rigid assumptions (Goodchild et al., 2000). Fotheringham et al. (2001) discuss how geospatial problems, computing power, and statistical inference can be leveraged for more rigorous analysis. In the LUCC arena, geocomputation could help develop

multiple scenarios from a given dataset, for example. Standard errors in a spatial framework that violate independence could be bootstrapped. Estimated parameter values and the associated standard errors could be used to generate a range of outcomes, instead of one single prediction. The implementation of geocomputational techniques in LUCC analysis thus could yield more nuanced insights as to the nature of recent and future change. Most commercial and open source statistical software packages have some functionality for computational approaches, but the spatially explicit extensions of these approaches remain limited.

Statistical analysis of LUCC will also be particularly plagued by the problems relating to spatial and temporal dependence and heterogeneity. It is often possible to identify and correct for these confounding effects, but there may not always be a "one-size fits all" approach. Given the myriad possible corrections for spatial autocorrelation, for example, theory should guide the identification and estimation of the most appropriate model for a particular question and the associated set of parameters to be estimated. In practice, however, it may not always be evident. Aspinall (2004) represents good use of multi-model comparison. He specified a series of straightforward and relatively simple models, and examined through the use of information criteria their degree of fit through time and space. His suggestion is to rely on inference from a set of models, rather than just one idealized model (Aspinall, 2004). More research in this direction is warranted, particularly to identify means of communicating multi-model comparisons to policy makers and other individuals in positions to require decision support.

Regarding policy, there is a tradeoff that is often discussed in LUCC regarding how to generalize from the local to the regional or global, and there are very real constraints regarding the resolution that is possible at higher levels of analysis. However, effective policy must acknowledge the impacts of local-level variation, particularly if the basis for policy is local (Geist and Lambin, 2002). There is often a discussion in LUCC that the agents of change are these local-level actors. Being able to model their behavior, and directly linking this observed behavior to theoretical drivers, provides scope for scaling up the results of their actions. In that way, aggregations can result in accurate regional descriptions of actual and potential influences on changes of land use. Statistical tools and methods are likely to remain an important component of LUCC analyses. If properly used, they remain powerful tools for rigorous hypothesis testing and ranking the relative influence among a set of factors, which can thus be translated into real policy objectives.

Acknowledgements

We gratefully acknowledge the support of the National Science Foundation (NSF) (SBR-9521918) as part of the ongoing research at the Center for the Study of Institutions, Population, and Environmental Change (CIPEC) at Indiana University; and the Deutsche Forschungsgemeinschaft (DFG) under the Emmy-Noether Program. We would also like to thank Catherine Calder for thoughtful discussion on Bayesian modeling, and Elena Irwin for useful comments. Lastly, we are grateful for the suggestions of two anonymous reviewers and to the editors of the special issue.

References

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

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, 91–103.

Atkinson, P.M., Tate, N.J., 2000. Spatial scale problems and geostatistical solutions. Professional Geographer 52 (4), 607–623.

Bergeron, G., Pender, J.L. 1999. Determinants of land use change: evidence from a community study in Honduras. Environment and Production Technology Division (EPTD), Discussion Paper No. 46, International Food Policy Research Institute (IFPRI). Washington, DC.

Bockstael, N.E., 1996. Modeling economics and ecology: the importance of a spatial perspective. American Journal of Agricultural Economics 78, 1168–1180.

Bockstael, N., Bell, K., 1998. Land-use patterns and water quality: the effect of differential land management controls. In: Just, R., Netanyahu, S. (Eds.), Conflict and Cooperation on Trans-Boundary Water Resources. Kluwer Academic Publishers, Boston.

Brown, D.G., Pijanowski, B.C., Duh, J.D., 2000. Modeling the relationships between land use and land cover on private lands in the Upper Midwest. Journal of Environmental Management 59, 247–263.

Chomitz, K.M., Gray, D., 1996. Roads, lands use, and deforestation: a spatial model applied to Belize. World Bank Economic Review 10, 487–512.

Crawford, T.W., 2002. Spatial modelling of village functional territories to support population-environment linkages. In: Walsh, S.J., Crews-Meyer, K.A. (Eds.), Linking People, Place, and Policy: A GIscience Approach. Kluwer Academic Publishers, Dordrecht, pp. 91–111.

Cressie, N., 1996. Change of support and the modifiable areal unit problem. Geographical Systems 3, 159–180.

Dutcher, D.D., Finely, J.C., Luloff, A.E., Johnson, J., 2004. Landowner perceptions of protecting and establishing riparian forests: a qualitative analysis. Society and Natural Resources 17 (4), 329–342.

Elhorst, J.P., 2001. Dynamic models in space and time. Geographical Analysis 33 (2), 119–140.

Elnagheeb, A.H., Bromley, D.W., 1994. Extensification of agriculture and deforestation: empirical evidence from Sudan. Agricultural Economics 10 (2), 193–200.

Evans, T.P., Kelley, H., 2004. Multi-scale analysis of a household level agent-based model of landcover change. Journal of Environmental Management 72 (1–2), 57–72.

Fisher, P., 1997. The pixel: a snare and a delusion. International Journal of Remote Sensing. 18 (3), 679–685.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2001. Quantitative Geography. Sage Publications, Thousand Oaks, CA.

Geist, H.J., Lambin, E.F., 2002. Proximate causes and underlying driving forces of tropical deforestation. Bioscience 52, 143–150.

Godoy, R., O'neill, K., Groff, S., Kostishack, P., Cubas, A., et al., 1997. Household determinants of deforestation by Amerindians in Honduras. World Development 25, 977–987.

Goodchild, M.F., Anselin, L., Appelbaum, R.P., Harthorn, B.H., 2000. Toward spatially integrated social science. International Regional Science Review 23, 139–159.

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

Irwin, E.G., Bockstael, N.E., 2004. Land use externalities, growth management policies, and urban sprawl. Regional Science and Urban Economics 34 (6), 705–725.

Irwin, E.G., Geoghegan, J., 2001. Theory, data, methods: developing spatially explicit economic models of land use change. Agriculture, Ecosystems and Environment 85, 7–23.

Jelinski, E.D., Wu, J., 1996. The modifiable areal unit problem and implications for landscape ecology. Landscape Ecology 11 (3), 129–140.

Maddala, G.S., 1983. Limited-dependent and qualitative variables in econometrics. Cambridge University Press, New York.

Mertens, B., Lambin, E.F., 2000. Land-cover change trajectories in southern Cameroon. Annals of the Association of American Geographers 99, 467–494.

Moran, Emilio F., Elinor Ostrom, J. C. Randolph. 2002. Ecological systems and multi-tier human organization. In: Douglas Kiel, L. (Ed.), Knowledge Management, Organizational Intelligence and Learning, and Complexity. Encyclopedia of Life Support Systems (EOLSS), Oxford, UK (Developed under the auspices of the UNESCO, EOLSS Publishers. Online publication: 〈http://www.eolss.net〉 (subscription required). Academic colleagues may request a hard copy from cipec@indiana.edu).

Müller, D., 2003. Land-use change in the Central Highlands of Vietnam: a spatial econometric model combining satellite imagery and village survey data. Doctoral Dissertation, Georg-August University Göttingen. Available at 〈http://webdoc.sub.gwdg.de/diss/2003/mueller/〉.

Müller, D., Munroe, D.K., 2005. Tradeoffs between rural development policies and forest protection: spatially-explicit modeling in the Central Highlands of Vietnam. Land Economics 81 (3), 412–425.

Müller, D., Zeller, M., 2002. Land use dynamics in the Central Highlands of Vietnam: a spatial model combining village survey data and satellite imagery interpretation. Agricultural Economics 27, 333–354.

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

Munroe, D., Southworth, J., Tucker, C.M., 2004. Modeling spatially and temporally complex land cover change: the case of western Honduras. The Professional Geographer 56 (4), 544–559.

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

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

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

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., 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.

Pontius Jr., R.G., Shusas, E., McEachern, M., 2004. Detecting important categorical land changes while accounting for persistence. Agriculture, Ecosystems and Environment 101 (2–3), 251–268.

Poole, D.J., Raftery, A.E., 2000. Inference for deterministic simulation models: the Bayesian melding approach. Journal of the American Statistical Association 95, 1244–1255.

Rindfuss, R.R., Walsh, S.J., Turner, I.B.L., Fox, J., Mishra, V., 2004. Developing a science of land change: challenges and methodological issues. Proceedings of the National Academy of Sciences 101, 13976–13981.

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

Sneddon, G., 2000. A statistical perspective on data assimilation in numerical models. Studies in the Atmospheric Sciences, Lecture Notes in Statistics, vol. 144, Springer, New York, pp.7–21.

Southgate, D., Sierra, R., Brown, L., 1991. The causes of tropical deforestation in Ecuador: a statistical analysis. World Development 19, 1145–1151.

Southworth, J., Munroe, D., Nagendra, H., 2004. Land cover change and landscape fragmentation—comparing the utility of continuous and discrete analyses for a western Honduras region. Agriculture, Ecosystems and Environment 101, 185–205.

Van der Veen, A., Otter, H.S., 2001. Land use changes in regional economic theory. Environmental Modeling and Assessment 6, 145–150.

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

Veldkamp, A., Verburg, P.H., Kok, K., de Koning, G.H.J., Priess, J., Bergsma, A.R., 2001. The need for scale sensitive approaches in spatially explicit land use change modeling. Environmental Modeling and Assessment 6, 111–121.

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

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

Walsh, S.J., Crawford, T.W., Crews-Meyer, K.A., Welsh, W.F., 2001. A multiscale analysis of land use land cover change and NDVI variation in Nang Rong district, Northeast Thailand. Agriculture, Ecosystems and Environment 85, 47–64.