ELSEVIER

Agricultural Economics 27 (2002) 201–216

AGRICULTURAL ECONOMICS

www.elsevier.com/locate/agecon

Deforestation and land use change: sparse data environments

Gerald C. Nelson [a,*], Jacqueline Geoghegan [b]

[a] *University of Illinois, Urbana-Champaign, 305 Mumford Hall, 1301 W. Gregory Drive, Urbana, IL 61801, USA*

[b] *Department of Economics, Clark University, 950 Main Street, Worcester, MA 01610-1477, USA*

Abstract

Understanding determinants of land use in developing countries has become a priority for researchers and policy makers with a wide range of interests. For the vast majority of these land use issues, the location of change is as important as its magnitude. This overview paper highlights new economic approaches to modeling land use determinants that combine non-traditional data sources with novel economic models and econometric techniques. A key feature is that location is central to the analysis. All data elements include an explicit location attribute, estimation techniques include the potential for complications from spatial effects, and results are location-specific. The paper reviews the theory underlying these models. Since this paper is intended to provide the potential new researcher with an introduction to the challenges of this analysis, we present an overview of how remotely-sensed data are collected and processed, describe key GIS concepts and identify sources of data for this type of econometric analysis. Finally, selected papers using these techniques are reviewed.

Keywords: Spatial analysis; Land use in developing countries; Spatial econometrics; Remotely-sensed data; Geographic information systems

1. Introduction

Understanding determinants of land use in developing countries has become a priority for researchers and policy makers with a wide range of interests. Concerns about consequences of deforestation for global climate change and biodiversity have received the most attention,[1] but loss of wetlands, declining land productivity, and watershed management are critical problems facing many developing country policy makers striving to enhance economic development while protecting the environment. For the vast majority of these challenges, the location of land use change is as important as its magnitude. For example, the loss of a particular plot of forest containing unique species is more serious than a much larger loss of a forest containing species found in many other places. Deforestation that results in soil erosion above a drinking water supply or major irrigation system has more deleterious effects on water and food availability than elsewhere.

In developing countries, analysis of land use determinants is especially constrained by lack of data. One does not typically find detailed crop or forest surveys, government statistical agencies are often under funded and data collection for agricultural and natural resource statistics can be sporadic.

This paper highlights new economic approaches to modeling land use determinants. These approaches

* Corresponding author. Tel.: +1-217-333-6465; fax: +1-217-333-5538.

E-mail address: g-nelson@staff.uiuc.edu (G.C. Nelson).

[1] Examples of multidisciplinary international efforts to understand land use determinants include the Land Use/Cover Change effort of the International Geosphere–Biosphere Program (http://www.igbp.kva.se/cgi-bin/php/frameset.php), the International Human Dimensions Program (http://www.uni-bonn.de/ihdp/) and the Millennium Ecosystem Assessment (http://www.millenniumassessment.org).

0169-5150/02/$ – see front matter
PII: S0169-5150(02)00080-4

combine non-traditional data sources[2] with novel economic models and econometric techniques. A key feature of this approach is that location is central to the analysis. All data elements include an explicit location attribute, estimation techniques include the potential for complications from spatial effects, and results are location-specific.

This paper has four sections. In the first section, we review the economic theory that is the basis for this type of analysis. In the second section, we provide an overview of how remotely-sensed data are gathered and processed to become useful for econometric analysis. The third section provides a similar overview of GIS concepts and how remotely-sensed data can be integrated with other spatial data. The final section reviews some empirical studies of land use in developing countries.

2. Economic theory underlying determinants of land use[3]

The focus of this approach is an individual parcel of land. The choice of land use on the parcel is made by the "operator", a single person, household, or group of people in the case of common property ownership. Three sets of variables determine this choice. The first set is the location's geophysical characteristics. These might be vegetative (type of forest cover, soil quality), mineral, or even atmospheric (rainfall, evapotranspiration). A second set of characteristics is socioeconomic—location-specific attributes such as prices of inputs and outputs; degree of operator control over the parcel; and household characteristics. Finally, geophysical and socioeconomic variables combine with a set of production technologies that relate inputs and outputs.

This literature typically assumes the operator of the parcel (the person with effective control over the land) uses its resources to increase his or her (or their, in the case of common property) utility. In this theoretical derivation, we equate utility and profit maximization (we address the necessary assumption for this later). The operator chooses a particular land use by comparing the net present value of the returns to all possible land uses. If we assume that a given land use has a single marketed product, the net present value of the return to that land use (h), its rent (R_{hl}) at time T, is given by

$$R_{hlT}=\int_{t=0}^{\infty}(P_{hlT+t}\,y_{hlT+t} - \boldsymbol{w}_{hlT+t}\boldsymbol{x}_{hlT+t})\,\mathrm{e}^{-i_l t}\,\mathrm{d}t \qquad (1)$$

where P is the output price, y the quantity of output, $\boldsymbol{w}$ is a vector of input costs, $\boldsymbol{x}$ is a vector of inputs under operator control and i_l is the location-specific discount rate, all for each land use h at location l at time T. At each parcel, the operator identifies the $\boldsymbol{x}$ to maximize R for each land use and then the operator chooses the land use that has the highest R_{hlT} for the parcel. Note that this formulation assumes that the operator starts tabula rasa; there are no costs of converting from an existing land use to one that has just become the most profitable.

With several restrictive assumptions we arrive at a theoretically-consistent reduced-form estimating equation that includes prices of inputs and output, a vector of geophysical characteristics ($\boldsymbol{c}_l$), parameters of a Cobb–Douglas production function (a_{kh}, input elasticities and b_h, constant productivity shifter) with k inputs, and a location-specific discount rate, i_l. The discount rate is location-specific to capture differences in effectiveness of property rights and cultural values:

$$\begin{aligned} R_{hlT} &= b_h\left[p_{hl}\boldsymbol{c}_l\prod_k w_{khl}^{-a_{kh}}a_{kh}^{a_{kh}}\right]^{1/b_h}\int_{t=0}^{\infty}\mathrm{e}^{-i_l t}\,\mathrm{d}t \\ &= b_h\left[p_{hl}\boldsymbol{c}_l\prod_k w_{khl}^{-a_{kh}}a_{kh}^{a_{kh}}\right]^{1/b_h}\left(\frac{1}{i_l}\right) \end{aligned} \qquad (2)$$

[2] Examples include world-wide datasets such as the Digital Chart of the World and the FAO World Soils Map, regional data sets such as NOAA weather satellite data and the Baltic Sea Region datasets (http://www.grida.no/prog/norbal/baltic/index.htm) and local datasets of land use derived from satellite images.

[3] This section draws heavily from Nelson et al. (2001). The static version of the model was originally developed in Chomitz and Gray (1996) and used to assess the effects of roads on land use in Belize. Nelson and Hellerstein (1997) extended the theoretical model to multiple time periods and used the approach to simulate the land use effects of complete removal of a road network in central Mexico. Numerous authors, including several in this issue, have used variants of this methodology to study determinants of land use in developing countries. Bockstael and associates at the University of Maryland have used a similar methodology to study urban expansion in the Washington, DC/Baltimore, MD region (e.g. Bockstael, 1996; Irwin and Bockstael, 2002). Bell and Irwin (2002) in this issue present this approach in more detail.

Data restrictions impose additional constraints. Prices are rarely available for all potential outputs and data are available for a single period only. Hence, most of the existing literature proxies prices of inputs and outputs using cost of access measures, similar to these developed in Chomitz and Gray (1996):[4]

$$P_{hl} = \exp[\gamma_{0l} + \gamma_{1l} D_l],$$
$$C_{hkl} = \exp[\delta_{0kl} + \delta_{1kl} D_l] \quad (3)$$

where D_l is the cost of access measure from a final destination of output or source of input to location l.[5] Substituting the price proxies and doing additional manipulations gives:

$$\ln R_{hlT} = \eta_{0h} + \sum_m \eta_{1hm} D_{lm} + \sum_r \eta_{2hr} c_{lr} + \eta_{3h} \ln i_l + \varepsilon_{hl} = \boldsymbol{\beta}'_l \boldsymbol{X}_l + \varepsilon_{hl} \quad (4)$$

In Eq. (4), the β_h values are reduced-form coefficients that derive from the production functions and price proxies, $\boldsymbol{X}_l$, is a vector of parcel-specific geophysical and socioeconomic characteristics and ε_{hl} is a stochastic error term. Parcel l is devoted to land use m if $R_{mlT} > R_{hlT}$, $\forall h \neq m$.

What we observe are the actual land use choices rather than the R values. This situation is similar to the discrete choice problem, where maximization of (unobserved) utility leads to an observed choice among discrete alternatives. We can reformulate this problem as finding the probability of choosing land use k at location l:

$$\Pr[\text{choice}\, m] = \Pr[\ln R_{mlT} > \ln R_{hlT}],$$

where $h \in \{1, \ldots, N\}$ a finite set of available choices, and $h \neq m$.

Substituting from Eq. (4) leads to

$$\Pr[\text{choice}\, h] = \Pr[\varepsilon_{hl} - \varepsilon_{1l} < (\boldsymbol{\beta}_h - \boldsymbol{\beta}_1)\boldsymbol{X}_l, \ldots, \varepsilon_{hl} - \varepsilon_{Nl} < (\boldsymbol{\beta}_h - \boldsymbol{\beta}_N)\boldsymbol{X}_l]$$

The choice of estimation techniques depends upon the distribution of the error term. If it is extreme value and the errors are uncorrelated across land uses, McFadden (1973) has shown that

$$\Pr(\text{choice}\, m) = \frac{e^{\boldsymbol{\beta}'_m \boldsymbol{X}_l}}{\sum_{h=1}^{N} e^{\boldsymbol{\beta}'_h \boldsymbol{X}_l}} \quad (5)$$

This is the standard multinomial logit regression.

If we assume instead that the errors are distributed normally, we have the multinomial probit (MNP) model. This does not have the irrelevance of independent alternatives (IIA) problem of the logit model (it allows an unrestricted covariance structure), but the maximum likelihood estimation of the $\boldsymbol{\beta}$ values is computationally challenging for more than four choices.[6]

In estimation, the $\boldsymbol{X}$ vector consists of three sets of explanatory variables: G, site-specific geophysical variables (soil quality, rainfall amounts, slope, elevation, etc.); D, cost-of-access, property rights, and other socioeconomic variables; and S, spatial effects geophysical variables (discussed later). To avoid identification, the β_h values for land use 0 are set to zero (this is usually done by the estimating software). The remaining β_h values can be interpreted as the marginal effects of right-hand side variables on the ln of the ratio of the probability of a land use choice to the zeroth land use. More generally,[7]

$$\frac{\partial \ln(\Pr_i/\Pr_j)}{\partial \boldsymbol{X}_i} = \boldsymbol{\beta}_i - \boldsymbol{\beta}_j. \quad (6)$$

[4] Note that this form assumes that the price proxies for all inputs (C_{khl}) are the same. This assumption does not mean the effect of a change in access cost is the same for all land uses. See Greene (2000) for a theoretical explanation and Nelson and Hellerstein (1997) for an example. This assumption seems reasonable for bulk commodities that are transported in similar size lots, for example in trucks or barges. It is probably less correct if the commodities produced range in variety.

[5] Note that it is possible to have distance be to more than one market. For example, Nelson and Hellerstein (1997) have multiple destinations—nearest village, nearest town, and nearest large population center.

[6] A relatively new estimation approach, called random parameters logit, allows the β_h values to be functions of different exogenous variables and with varying error structures. Its developers state that the various permutations of logit and probit are nested within this technique (see, for example, McFadden and Train, 2000). However, there is no guidance as to the appropriate exogenous variables and error structures. See Nelson et al. (2003) for an example of its use.

[7] One must interpret this effect carefully. An increase in the probability of a land use relative to the base land use (or any other for that matter) may have no significance on its likelihood of being "chosen" when compared to other possible land uses.

We use the estimated β_h values to generate probability predictions for each land use at every location in the area under investigation. For example, with five land use choices, we might find the following land use probabilities at a location: forest type A, 72%; forest type B, 18%; agriculture, 3%; pasture, 6%; urban areas, 1%. The sum of probabilities for all five categories is 100%.

2.1. Assessing predictive power

With a continuous left-hand side variable, we typically assess overall predictive power with an adjusted R^2 measure and contributions of a right-hand side variable with coefficient-specific standard errors. For limited dependent variable models, both of these measures have problems. Since $\hat{y}$ does not exist, standard measures based on $\hat{\varepsilon} = y - \hat{y}$ are not possible.

A number of pseudo R^2 measures have been proposed (see Maddala, 1983). The most common measure currently used is $(1 - \ln L)/\ln L^*$, where L is the likelihood function value for the full model and L^* is the value with only constants on the right-hand side. Greene (2000) points out that this should not be interpreted in the same way as the regular R^2, i.e. a continuous function that is correlated with predictive power.

A standard approach to assessing predictive power for individual observations is based on the predicted probabilities of land use. Probability values range from 0 to 100% for each land use at each location. Typically, these probability values are converted to a point prediction by assigning a location to the land use category with the highest probability. This approach assigns a land use to all locations, but does not distinguish between 'strong' and 'weak' predictions. For example, with five land use categories, the largest probability value can range from 99 to 20.1%. Chomitz and Gray (1996) propose an alternate approach that involves assigning a location to a "natural" land use only if its predicted probability is higher than the actual ratio of that land use to total land area. This approach leaves some locations unassigned. Geoghegan et al. (2001) use another approach that keeps predicted and actual areas the same but allows predicted location to vary. The assignment algorithm allocates locations to the highest probability observations until the actual number has been exhausted.

Once predictions have been made, the most frequently used method of assessing the predictive power is to calculate a "prediction matrix" comparing actual and predicted categories. The matrix rows typically show the number of locations actually in a given category; its columns show the number of locations predicted to be in a given category, where the predicted land use is the one with the highest probability. Diagonal elements are correct predictions.[8] Two types of category-specific ratios can be used to assess the predictive power—number of correctly predicted pixels to number of actual pixels and number of correctly predicted pixels to total predicted pixels. It is quite common to find substantial differences in the ratio values across categories, suggesting differences in predictive power of the right-hand side variables for different land uses. One area of future research is to explore the use of exogenous variables for selected categories only.

The prediction matrix gives no information on the spatial accuracy of the prediction. One approach to this issue is to plot all locations where the predicted and actual land uses differ. Another approach is to use the probability values directly. Two related graphical measures of predictive power are presented in Nelson et al. (2001). The first maps the maximum probability value, $\mathrm{Pr}_{\mathrm{max}}$, at every location. This map gives a spatial representation of the power of the prediction but does not convey any information about prediction accuracy. The second measure maps $\mathrm{Pr}_{\mathrm{diff}} = \mathrm{Pr}_{\mathrm{max}} - \mathrm{Pr}_{\mathrm{actual}}$ (the probability value for the actual land use). If the category with the highest probability value is also the actual category, $\mathrm{Pr}_{\mathrm{diff}} = 0$. Otherwise $0 < \mathrm{Pr}_{\mathrm{diff}} < 1$. A recent paper in the remote sensing literature with some potential in this area is Pontius (2000). He uses the kappa statistic

$$\kappa = \frac{P_{\mathrm{o}} - P_{\mathrm{c}}}{P_{\mathrm{p}} - P_{\mathrm{c}}}$$

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

[8] The prediction matrix is like a "confusion matrix" in the remote sensing literature that compares categories identified by a classification scheme to categories identified by ground observation (Richards, 1993).

to develop measures for overall and location-specific predictive power.

2.2. Theory issues

The theoretical approach described above, and the estimations based on it, can provide powerful results but requires strong simplifying assumptions. The literature on household models (the classic reference is Singh et al., 1986) has clearly demonstrated that household utility maximization and profit maximization are the same only when markets function perfectly, there are no transactions costs, and production can adequately be characterized without regard to temporal effects.[9] For example, the behavior of farmers concerning subsistence crops versus market-oriented crops can differ because of risk aversion or transaction costs. In addition, the decision to enter the market is itself an endogenous choice, leading to possible sample selection bias (Vance and Geoghegan, 2003). Also, as many farmers grow a number of different crops, a portfolio choice model might be an appropriate approach for modeling the entire suite of land use choices.

Since much of this literature is concerned with deforestation, the temporal dimension has received increasing attention. While annual crops can be adequately analyzed using cross-section data, many forestry land uses require multiple years to generate an output. In addition, swidden land uses shift from agriculture to forest and back, potentially confounding any analysis based on a single cross-section (Dvorak, 1992). Once a temporal component is added to production, decision-making under uncertainty becomes even more of an issue. One approach taken in the literature is to convert the dynamic problem to a static problem. For example, Deininger and Minten (1999), Cropper et al. (1999), and Geoghegan et al. (2001) use land use information from two time periods, identify locations where deforestation has taken place and explain deforestation with a set of exogenous right-hand side variables. Mertens and Lambin (2000) identify and explain land cover change trajectories. For example, with two land uses (agriculture and forest) and three periods, there are eight possible trajectories (aaa, aff, afa, aaf, fff, faa, ffa, faf, ffa). Clearly, this approach is not robust. As more land uses or periods are added the number of trajectories expands rapidly. Yet, another approach uses a survival or hazard model methodology, where the linkage through time of the observations is explicitly modelled (see Vance and Geoghegan, 2002 for an implementation example and Bell and Irwin, 2002 for a discussion of the approach).

Second, land use choice may be affected by neighbouring parcels, through spatial externalities or spatial interactions. For example, by incorporating spatially-explicit ecological interdependencies in a forestry management model, Swallow et al. (1997) show that the optimal forestry management scheme can differ substantially from using a non-spatial modeling approach, including different harvesting periods for the different stands and amounts of the total benefits of recreation and forage availability to wildlife. In the urban land use change literature, Irwin and Bockstael (1996) show that a sub-optimal pattern of land uses can occur when individual landowners do not take into account the negative spatial externalities associated with suburbanization. The paper by Anselin in this issue presents the spatial econometric issues in more detail.

Third, the tabula rasa assumption is not appropriate in important situations. Conversion from one land use to another is seldom costless and may be essentially infinite in the case of land uses that are infeasible on a particular parcel (for example, producing paddy rice on a steeply sloping hillside). If there are conversion costs, then there are likely thresholds in land use conversion, so that a larger divergence in the relative prices of two crops is necessary to induce a switch from one crop to another. In addition, the costs are not necessarily symmetric. Conversion from a tree crop to an annual field crop may entail a relatively small cost, but growing new trees can take several years of foregone revenue. Finally, the returns from a particular land use can depend upon the entire history of land uses for the parcel, due to soil quality changes and pest problems.

Fourth, this literature has not dealt systematically with the issue of market structure, in large part because location-specific data on prices have been difficult to obtain. The basic assumption of most research is that distance or cost of access measures are

[9] We would like to thank an anonymous reviewer for emphasizing the relevance of the household literature to this area of research.

acceptable proxies for input and output prices. Once data on actual prices are available, modeling issues including identifying the relevant market, transport costs, and the possibility of price endogeneity.

Finally, any econometric estimation can only capture existing land uses. If the issue at hand is the introduction of a new land use (say, growing a citrus crop for export), it is impossible to estimate coefficients of the effect of right-hand side variables on the probability of that land use.

2.3. Spatial effects and limited dependent variable analysis

The Anselin paper on spatial econometric topics in this issue (Anselin, 2002) describes the potential for bias and inefficiency if spatial effects are not accounted for. Most of the research in this area is for datasets with continuous right-hand side variables. However, the consequences for limited dependent variable analysis, such as land use, are similar. While inefficiency is not usually a problem because of the large datasets typically used, bias in parameter estimates because of spatial autocorrelation is a potential problem. The theory of identifying and correcting for these issues with limited dependent variables is in its infancy. For an examination of these issues see Fleming (in press) and the Anselin paper.

Some authors have combined a regular sampling procedure suggested by Besag (1974) with a simple spatial lag variable included on the right-hand side. The Besag approach is to include only observations separated by sufficient distance in space that the autoregressive effect is absent. Spatial lag variables have included the latitude and longitude values, and average vegetative and soil quality indices in the surrounding locations (e.g. Nelson and Hellerstein, 1997; Nelson et al., 2001, 2003).

Any test for spatial effects requires a measure of errors in prediction with known statistical properties. A recent paper by Kelejian and Prucha (2001) proposes a pseudo-error measure for limited dependent variable estimation, similar to the Moran's I-statistic, derive its statistical properties and develop a measure of spatial correlation. See Munroe et al. (this issue) for an example of how this statistic can be used to assess the value of the ad hoc methods for correcting for spatial autocorrelation mentioned above and De Pinto and Nelson (in press) for more details on use of the statistic to test ad-hoc correction approaches.

2.4. Answering interesting questions

To be of use to policy makers and researchers asking land use questions, this type of analysis must be amenable to simulation of alternatives of policy-relevant variables. With estimates of the β_h values, we can simulate the effect of changes in any of the existing right-hand side variables. The basic approach is to replace the right-hand side variables with new values that reflect a policy or infrastructure change, and recalculate the probabilities and predictions. Comparisons of the old and new land use values indicate where and by how much the exogenous change affects land use. Comparisons can be made using transition matrices[10] and land use change maps. Examples of questions to which this simulation approach has been used include how and where does a new road affect land use (static analysis) and deforestation (dynamic analysis), and how and where do changes in property rights regimes affect land use? We present selected examples at the end of this paper and the papers included in this issue provide more.

3. An introduction to remotely-sensed data

For the researcher new to the area of spatially-explicit land use modeling, we provide a brief introduction to data concepts and sources. Spatially referenced data on land use/cover are needed to estimate the coefficients in the models described above. For locations that are remote or in developing countries, spatially-explicit data collected on the ground are hard to obtain. An alternative that has become increasingly available is remotely-sensed data. This section discusses the basics of how remotely-sensed data are

[10] A transition matrix has one state of nature (e.g. existing land use) along the vertical and a second state of nature along the horizontal (e.g. land use without a reserve). A matrix cell contains the number of members common to both the first and second state. For example, if 9 km^2 of forest in 1987 were converted to agriculture in 1997, the intersection of the 1987 forest column with the 1997 agriculture row would be nine.

collected, highlights different kinds of remotely-sensed data and provides an overview of how these data can be converted into information about vegetative cover and land use.

3.1. Satellites and sensors

Remotely-sensed data include information gathered digitally by aerial photography and satellites.[11] Solar radiation is reflected from the surface of the earth—from soil, water, vegetation and building—to sensors that measure the intensity of different frequencies. Each type of surface reflects or absorbs different frequencies. Hence by a judicious choice of sensor type it is possible to make inferences about what is on the surface of the earth.

3.2. Remotely-sensed data source issues

Fig. 1 illustrates the basics of how a satellite captures information and what the data choices are. Earth observing satellites (as opposed to geosynchronous telecommunications satellites) orbit the earth at a low altitude. The height of the orbit and its inclination determine how often and at what time of day the satellite passes over the same location. The height plus the resolving power and field of view of the sensors determine the width of the swath of the surface observed and how much of the surface is captured by a single sensor. Satellites have been developed that allow resolutions from 1 km (NOAA) to 0.65 m (Quickbird) (and probably smaller for spy satellites). Repeat rates (how often a location is visited) range from every few hours to 16 or more days. The number of frequency values collected range from 4 (MSS) to 100 s (ASTER/MODIS).

It is useful to describe the characteristics of the MSS sensor, carried on the early Landsat satellites, in more detail because it illustrates many of the issues in collecting and using remotely-sensed data.[12] This sensor has an array of six detectors that measure the intensity of light in each of four frequency ranges or bands, from 0.5 to 1.1 μm. For example, band 1 records frequencies of 0.5–0.6 μm (green light) that is reflected by chlorophyll. Band 2 records frequencies of 0.6–0.7 μm (yellow/red light). These frequencies are reflected by chlorophyll. The Landsat satellites operate in a sun-synchronous, near-polar orbit imaging the same 185 km (115 miles) ground swath every 16 days (formerly 18 days on Landsats 1–3).

MSS band 1 can be used to detect green reflectance from healthy vegetation, while MSS band 2 is designed for detecting chlorophyll absorption in vegetation. MSS bands 3 and 4 are ideal for recording near infrared reflectance peaks in healthy green vegetation and for detecting water–land interfaces.

MSS bands 4, 2, and 1 can be combined to make color images (called false-color), where band 4 controls the amount of red in the image, band 2 the amount of green, and band 1 the amount of blue in the composite. This band combination makes vegetation appear as shades of red with brighter reds indicating more vigorously growing vegetation. Soils with no or sparse vegetation will range from white (sands) to greens or browns, depending on moisture and organic matter content. Water bodies appear blue. Deep, clear water appears dark blue to black in color, while sediment-laden or shallow waters appear lighter in color. Urban areas appear blue–gray in color. Clouds and snow appear as bright white; they are usually distinguishable from each other by the shadows associated with the clouds. To see an example of a false-color composite MSS image of Boston harbor, visit http://edcwww.cr.usgs.gov/glis/graphics/guide/landsat/bostonmss.gif.

Two types of satellite images from sensors on the Landsat satellites are widely available and inexpensive (free to US$ 600)—MSS and TM images. AVHRR images are from the weather satellites operated by NOAA. Other sources such as Spot, Space Imaging, and Digital Globe are proprietary and more expensive,

[11] For an excellent on-line introduction to satellite data, see the NASA web site at: http://rst.gsfc.nasa.gov/. For a historical overview of satellite data collection, see Morain (1998). For further detail on the technicalities of the processes involved with developing satellite data, see Mather (1999). For an overview of different approaches to using satellite data to develop land cover maps, see DeFries and Belward (2000). Finally, for sources of satellite data for social science applications beyond those presented in this paper, see Chen (1998).

[12] The MSS discussion is derived from the EROS Data Center web site, http://edcwww.cr.usgs.gov/Webglis/glisbin/guide.pl/glis/hyper/guide/landsat.

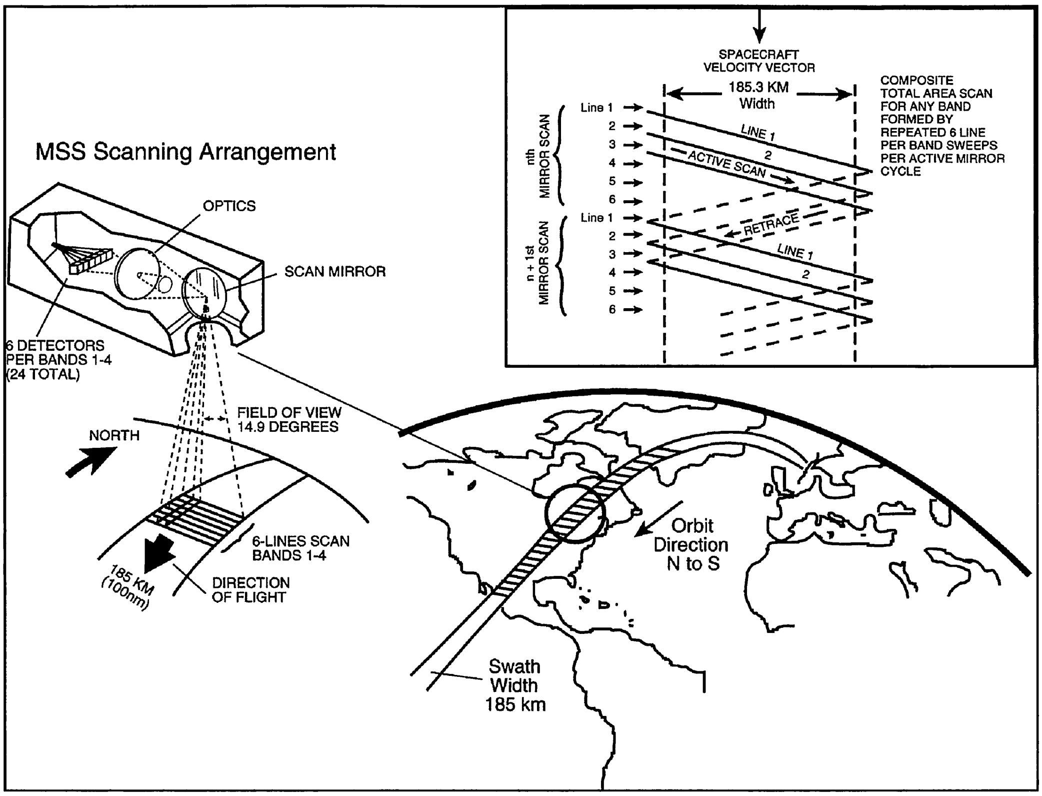

Fig. 1. How satellites collect information. *Source*: EROS data center web site.

but can be purchased with high-quality processing that can significantly accelerate analysis (Table 1).[13]

3.3. Converting reflectance data to useful information

Land uses typically cause distinctive patterns of land cover. 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 use clustering/classification techniques.

3.3.1. Vegetative indices

While there are many ways to combine different spectra to take advantage of the ways in which vegetation absorbs and reflects solar radiation, the most common is the normalized difference vegetative index (NDVI) that uses MSS (or equivalent) bands 2 (0.58–0.68 μm) and 4 (0.725–1.1 μm). The NDVI has a potential range from −1 to 1 but the typical range is between about −0.1 (a not very green area) and 0.6 (for a very green area). In most cases, NDVI is correlated with photosynthesis. Because photosynthesis occurs in the green parts of

[13] More information on image availability, characteristics, and prices can be found at the following web sites: http://edcwww.cr.usgs.gov/, http://www.spaceimaging.com/, http://edcdaac.usgs.gov/landsat7/, http://edcimswww.cr.usgs.gov/pub/imswelcome/, http://www.geocover.com/, http://www.landsat4u.com/Merchant/index.html, http://members.aol.com/landsatcd/MOREHTML/shuttle.html, http://www.nasm.edu/ceps/homepage.html, http://eol.jsc.nasa.gov/sseop/, http://www.spot.com/ and http://www.rsi.ca/.

Table 1
Selected satellites and their characteristics

Satellite/sensor	Repeat rate	Area of image	Pixel dimension	Frequencies	Dates available
Landsat/MSS	16–18 days	150 km × 150 km	80 m	4—green, red, infrared	Early 1970s–early 1990s
Landsat/TM	16–18 days	150 km × 150 km	30/15 m	7—blue, green, red, 1 near infrared, 2 mid infrared, 1 thermal	Mid 1970s–today
AVHRR	Two times per day	800 km × 800 km	1.1 km	4/5—green, red, infrared, lower frequencies	Mid 1970s–today
IKONOS (Space Imaging)	1–3 days	Variable	1–4 m	Same as MSS	2000–today
Quickbird (Digital Globe)	1–3 days	16 km × 16 km	0.6 m	4—blue, green, red, near infrared	2002–today
SPOT (Spot Image)	3–6 days	60 km × 60 km	10–20 m	3—green, red, near-infrared	1986–today

Note: For more details on the types of satellites, see http://atlas.esrin.esa.it:8000/lib/faq-1.html.

plant material the NDVI is normally used to estimate green vegetation. However, a variety of complications make the NDVI and other vegetative indices at best imperfect estimates of the amount of vegetation:

$$\mathrm{NDVI}_{i,j} = \frac{B4_{i,j} - B2_{i,j}}{B4_{i,j} + B2_{i,j}} \tag{7}$$

where $Bn_{i,j}$ is the intensity value of MSS band n at relative points i and j.

3.3.2. Clustering/classification techniques

A land use category is a qualitative label given to areas with similar operational characteristics (e.g. forest, agriculture). There are often (but not always) discontinuities in the characteristics of reflected light as the field of view moves across different land uses. Clustering techniques operate by assuming that pixels with similar spectral characteristics have the same land use. 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. One example is the histogram peak approach (Fig. 2).

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 maximize, then assume a normal distribution and use a maximum likelihood estimator to assign remaining pixels to clusters (Fig. 3).

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.

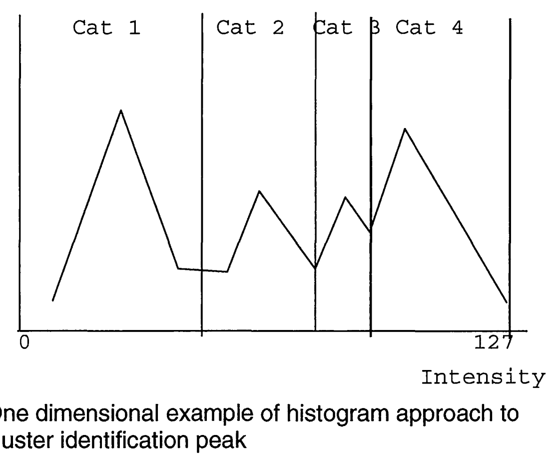

Fig. 2. Histogram peak approach to classification.

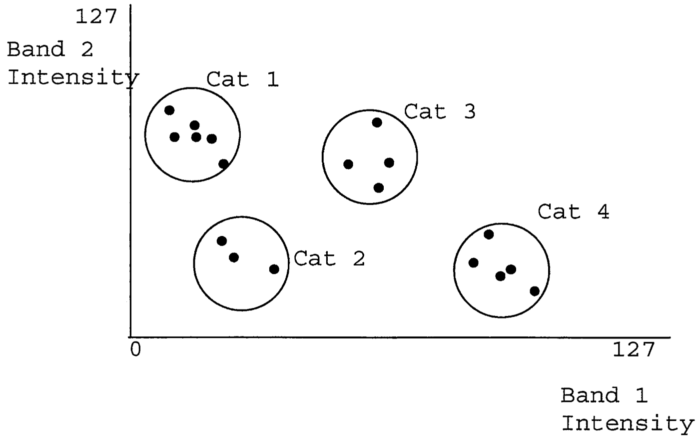

Fig. 3. Distance-based classification algorithms.

4. Combining (remotely-sensed) land use data with other geographic data

4.1. GIS theory

In this section, we present a brief introduction to some of the most important concepts in GIS theory. The first is the difference between raster and vector representation of space.

4.1.1. Data representation—raster and vector

Raster data record spatial information in a regular grid. Each cell within this grid contains a number representing a particular geographic feature, such as soil type, elevation or slope, land use, and price of an input or output. Raster data are commonly used to store information about geographic features that vary continuously over a surface, such as elevation, reflectance, groundwater depths. Socioeconomic variables such as transportation cost are also sometimes stored as raster data. Image data are a form of raster data in which each cell (also called a pixel, short for picture element) stores a value measuring the intensity of light of a given frequency range arriving at a satellite or aerial camera.

With vector data, spatial information is stored as x, y coordinates in a rectangular (planar) coordinate system. Point features are recorded as single x, y locations. Line features, including the outlines of polygons, are recorded as an ordered series of x, y coordinates. Each vector feature (point, line, or polygon) has an attribute table that describes the attributes of the feature. Vector data are used for recording the location of discrete geographic features with precise locations like streets, parcel boundaries, counties, and telephone poles.

4.1.2. Georeferencing

A second key concept is georeferencing. The earth is a sphere (although not a perfect one) and we typically work with data taken from the surface of that sphere and projected into two dimensions. Map projections are attempts to portray the surface of the earth or a portion of the earth on a flat surface. Some distortions always result from this process. Some projections minimize distortions in some of the map features at the expense of maximizing errors in others. Others moderately distort all map features.

Geodetic datum define the size and shape of the earth and the origin and orientation of the coordinate systems used to map the earth. A datum is a set of parameters defining a coordinate system, and a set of control points whose geometric relationships are known, either through measurement or calculation. A datum is defined by a spheroid, which approximates the shape of the earth, and the spheroid's position relative to the center of the earth. There are many spheroids in use, and many more datum based upon them. A local datum aligns its spheroid to closely fit the earth's surface in a particular area and its 'origin point' is located on the surface of the earth.

4.1.3. Map algebra

A third set of key map concepts is map algebra, where manipulations of one or more spatially referenced datasets create new datasets. Examples include calculating slope from an elevation dataset; calculating travel time or transportation cost data from a friction or impedance surface (a map that gives the cost of moving across each map element) and calculating the distances between different locations.

4.2. Extracting data from remotely-sensed and GIS sources

Unfortunately for economic modeling purposes, there are few standards for data structures. Every GIS, remote sensing, and econometric software package has its own approach to storing data. To move data from one package to another it is necessary to know something about file formats. To provide some insights on how to proceed in this area, we present a brief description of the file formats for selected software packages.

4.2.1. Raster data formats

Raster data are unique values for locations arranged in a regular pattern. The data representation requires knowledge of the value at each point and the location of one of the points. The regular structure makes it possible to infer the location of all the other points.

Data formats include:

- byte binary—1 (8 bit) byte represents one pixel; values from 0 to 256 (unsigned) or −127 to 127 (signed); common for satellite image data;
- two byte binary—2 (8 bit) bytes represent one pixel; values from 0 to 32,768 if unsigned, −16,384 to +16,384 if signed; common for digital elevation data and newer satellite images; byte order and sign bit vary by computer system;
- real single precision binary—4 (8 bit) bytes represent one pixel; value range infinite; used for continuous variables;
- character—each byte is an ASCII character. The value 135 would take 3 bytes.

```
    16 (feature id)
  -27943.236328     2219055.250000
  -27179.146484     2219286.000000
END
    26 (feature id)
  -63867.117188     2217780.250000
  -62606.257812     2218345.000000
  -61211.492188     2219213.750000
  -60352.003906     2220168.500000
  -60021.976562     2220086.000000
  -59469.496094     2219540.000000
END
```

Fig. 4. An example of Arc Info's Ungen vector data format.

Since this type of data is often voluminous a compression algorithm is typically used. Unfortunately there are no standard approaches to compression. Run length encoded (RLE) is a common compression approach. Data are stored in two byte pairs. The first byte is how many repetitions; second byte is what is repeated. A variant on this is to include some kind of additional information on whether a row is compressed.

4.2.2. Vector data formats

The most easily transferable vector data are in Arc Info's Ungen format. All features are represented by ASCII data describing the location of its nodes. A point would have just one x, y pair. A line would have two x, y pairs. A road would have multiple x, y pairs but the first and last would not be the same. A polygon would have the same first and last x, y pairs. In addition, each feature has a unique id value. For extensive datasets, this Ungen format can create very large file sizes. The price of easy transferability is an inefficient data storage method. An alternative that is growing in popularity is ESRI's shape file. Although this is a binary file format, the structure has been published by ESRI and most GIS packages can read and write data in this format (Fig. 4).

4.3. Secondary sources of GIS data

In this section, we describe a few of the spatial datasets with regional or world coverage that are readily available, often via the Internet. See Bell and

Irwin (2002), for additional locations of (primarily US) datasets useful for regional and local analysis.

4.3.1. VMap—formerly Digital Chart of the World (DCW)

VMap, a revised version of the Digital Chart of the World (DCW) is a 1:1,000,000 scale map based on the operational navigation charts (ONCs) used by aircraft pilots. It was developed by the US National Imagery and Mapping Agency. VMap has data on coastlines, international boundaries, cities, airports, elevations, roads, railroads, water features, cultural landmarks, and much more. It is the most detailed global database available that provides consistent treatment of geographic information worldwide, and is the only source of spatial data for many areas of the globe. The database totals 1.7GB in size and comes on four CD-ROMs. Parts of the original DCW dataset can be downloaded from http://www.maproom.psu.edu/dcw/. More than 200 attributes are organized into 17 thematic layers with text annotation for cities, mountains, lakes, and other geographic features.

4.3.2. FAO World Soils Map

The Digital Soil Map of the World (version 3) was released in May 1994 and version 3.5 is now available. The database is derived from the FAO/UNESCO Soil Map of the World at the original scale of 1:5,000,000. The database is available on CD and can also be downloaded from http://www.fao.org/sd/eidirect/gis/eigis000.htm.

4.3.3. FAO Africa Rainfall

Since 1988, the FAO has been operating the Africa Real Time Environmental Monitoring Information System (ARTEMIS, http://metart.fao.org/). The system acquires and processes hourly estimates of rainfall and in near-real-time vegetation index (NDVI) images, using Meteosat and NOAA data. The system covers the whole of Africa and the products are produced on a ten-day and monthly basis for use in early warning for food security and desert locust control.

4.3.4. Earthsat TM

The US National Aeronautics and Space Administration (NASA) funded a project to acquire TM images of the entire globe for 1990 (or the nearest year with cloud free cover) and process them with a standard set of protocols. These images are being transferred to the EROS Data Center's Earth Observing System Data Gateway at http://edcimswww.cr.usgs.gov/pub/imswelcome/ and can be ordered for US$ 60 per scene. A similar dataset is being constructed for 2000.

4.3.5. TRFIC

The Tropical Rain Forest Information Center at Michigan State University provides a valuable archival and distribution site for public domain images. They purchase images for their own use and make them available at reduced cost. They also store and disseminate public domain datasets purchased by others. The web site for data access and purchase is http://www.bsrsi.msu.edu/trfic/index.html.

4.3.6. CIESIN

CIESIN (http://www.ciesin.org/) has links to a variety of georeferenced socioeconomic and environmental data sources. We describe two here.

China Dimensions (http://sedac.ciesin.org/china/) has a variety of socioeconomic data, including geographic information system (GIS) databases that cover the administrative regions of China, at a scale of 1:1,000,000. These databases may be integrated with agricultural, land use, environmental, and socioeconomic data to track China's economic growth, population increases, and environmental change.

A spatial dataset describing Central American vegetation, land cover, and conservation status is now available for downloading via file transfer protocol (ftp) at: ftp://ftp.ciesin.org/pub/data/conservation/PROARCA/. The dataset was developed by Proyecto Ambiental Regional de Centroamerica/Central America Protected Areas Systems (PROARCA/CAPAS) and is being distributed on behalf of the Nature Conservancy (http://www.tnc.org/). The dataset is in ArcView 3.0 format.

4.3.7. USGS national land cover data

The United States Geological Survey (http://landcover.usgs.gov/mrlcreg.html) is in the process of developing a national land cover dataset from Landsat–TM images that contain over twenty land cover classes for each state. Currently, approximately twenty eastern states are available in the final format, while the rest of the country is in preliminary form, as accuracy assessment has not been completed.

4.4. Primary sources of georeferenced data

Another approach for developing spatial datasets, albeit more expensive, is to collect primary data. There are a few large, interdisciplinary research projects underway that link spatially-explicit socio-demographic data with satellite and other GIS data in an agricultural development context. Some examples of these include case studies in Mexico (Geoghegan et al., 2001; Vance and Geoghegan, 2002), the Amazon (Wood and Skole, 1998; Moran and Brondizio, 1998), Thailand (Rindfuss et al., 2001), Cameroon (Mertens et al., 2000), Vietnam (Müller and Zeller, this issue) and Honduras (Munroe et al., this issue). In this approach, enumerators interview households with a standard survey instrument to collect assorted socio-demographic data and then these household locations and their associated agricultural plots are located in space using global positioning systems (GPS) technology, which gives the precise location on the earth's surface.

5. Modeling determinants of deforestation in developing countries, examples

In this section, we review four papers that use a spatially-explicit modeling approach to examine issues of land use and deforestation in developing countries. Kaimowitz and Angelsen (1998) provide an excellent review of deforestation models and other models of forest use in developing countries through the mid 1980s. A recent issue of *Land Economics* (Vol. 77, No. 2, 2001) presents several papers on this topic as well.

5.1. Roads, land, markets and deforestation: a spatial model of land use in Belize

This path breaking paper by Chomitz and Gray (1996) develops the theoretical model widely used in this literature. The model is used to identify the determinants of forest loss in southern Belize, an area experiencing rapid expansion of both subsistence and commercial agriculture. The paper uses geographic data to distinguish the effects of roads from other determinants of forests and forest loss. One of the challenges in this literature is how to deal with the possibility that road location is endogenous; that is, that roads are built to access favorable areas. Chomitz and Gray address this problem by using an instrumental variables approach, calculating an accessibility measure as if there were no roads. A second issue is how to deal with the potential for spatial autocorrelation. The authors report the use of a bootstrapping procedure to estimate the standard errors of the coefficients.

The authors find that market distance, land quality and tenure have strong interaction effects on the likelihood and type of cultivation. In a region with geophysical characteristics favorable for commercial agriculture, a location near a market has a 34% chance of being converted to commercial agriculture but only a 1.4% chance of being in semi-subsistence agriculture. In a different location, with geophysical variables more favorable to semi-subsistence agriculture, a location near a market has a 45% probability of being in semi-subsistence agriculture but only a 5% probability of being in commercial agriculture. As distance to the market increases, the probability of being in either semi-subsistence or commercial agriculture drops off, but it drops more rapidly for commercial agriculture.

5.2. Sustainable development in Panama's Darien Province: modeling land use change with spatial econometric analysis

This paper (Nelson et al., 1999) reports results from spatial econometric analysis undertaken for a sustainable development project preparation at the Inter-American Development Bank. A major element of the project is paving the Pan American highway which runs roughly north south through the province to a point about 70 km from the Colombian border. The highway, which was originally completed in 1983, passes near a reserve for indigenous populations, and stops just north of the Darien National Park, a UNESCO Biosphere Reserve and World Heritage Site. Concerns arose that paving the road would encourage more use of the park and bring pressure from immigrants on the reserve.

The study used a methodology similar to that developed by Chomitz and Gray to predict land use. Then using the estimated coefficients, the effect of paving the road on land use was simulated by re-computing the cost of access, and calculating new probability values and land use predictions. The results suggested that, except for one type of forest (cativo) with small area, land use would not change significantly, especially in those areas of concern. The original construction of the road in 1983 led to significant reductions in access cost and therefore land use change, especially in the northern part of the province. The paving would have relatively little effect, especially because the province is well served with navigable rivers.

5.3. Predicting the location of deforestation: the role of roads and protected areas in North Thailand

Cropper et al. (1999) examine the question of what factors affect the location of deforestation in northern Thailand. The authors use the model to predict where deforestation is likely to occur and to examine the effects of two government policies—road building and establishment of protected areas—on this likelihood. By 1986, 10% of Thailand lay within protected areas, of which 52% was in national parks and 42% in wildlife sanctuaries. The authors use a spatially-explicit model to assess whether official protection has reduced the probability of deforestation in these protected areas. The authors use binary probit (deforestation/no deforestation) (in contrast to the two papers reviewed above which use a logit model with multiple land use categories), which does not suffer from the irrelevance of independent alternatives (IIA) problem (see Greene, 2000).

Their results on the effects of roads are qualitatively similar to those of other studies. The effect of roads on land use is conditioned by geographic and socioeconomic variables. For example, steeper slopes and higher elevations reduce the probability that a location has been cleared. The question of endogeneity of reserves is dealt with by using a watershed proxy. They find, after correcting for this endogeneity, that wildlife sanctuaries have a much larger protective effect than do protected areas.

The authors use their results to simulate the effect on protected areas of increased road building. They find that bringing a paved road 1 km closer to each point in their sample increases the number of areas with high probability of clearing. The locations of these points are often near points predicted to be cleared even before the simulated road building.

5.4. Agricultural expansion and deforestation: linking satellite and survey data in southern Mexico

Geoghegan et al. (2001) compare two separate econometric models of deforestation in the southern Yucatan peninsula of Mexico. This region is part of the largest continuous expanse of tropical forests remaining in Central America and Mexico, and has been identified as a "hot spot" of forest and biotic diversity loss. Two complementary datasets, one from household survey data on agricultural practices including information on socio-economic factors and the second from satellite imagery linked with aggregate government census data, are used in two econometric modeling approaches. The first econometric model uses the satellite data, other spatial environmental variables, and *aggregate* socioeconomic data (e.g. census data), in a similar manner to the papers described above, using a logit model to estimate the probability that any particular *pixel* in the landscape will be deforested. The explanatory variables are also similar to previous studies: slope, elevation, soil type, distances to road, village, market, and nearest agricultural land use; and variables available from the census, such as population density, where the census data are measured at the village level, so that all pixels in a particular village have the same value for these variables. The results are also similar to previous studies: the higher the elevation, the smaller the probability of deforestation; the further a pixel is from the road, the less likelihood of deforestation; the closer a pixel is to a market or a village, the greater the probability of deforestation.

The second regression uses the survey data in an OLS model to ask questions about the amount of deforestation (a continuous variable) associated with *each* individual farmer from the household survey and to explain these choices as a function of demographic, market, environmental, and geographic variables. The geophysical variables are the same as in the logit model—elevation; slope; soil type. The other variables come from the survey work and include distance from house to agricultural plot; distance from plot to major road; road distance to nearest major market; measures of household human and physical capital; and household population. The estimated coefficients for the geophysical variables, as well as the household population and human capital measures, are statistically significant and the same sign as in the logit model. However, none of the location variables are statistically significant in this model. One tentative conclusion of the paper is that while location affects the overall probability of deforestation, it does not appear to explain the total amount of deforestation on a given location by an individual farmer.

6. Conclusions

Analysis of land use and land use change in developing countries is difficult to imagine without the use of georeferenced data. The only way to make location-specific predictions is with spatial data. In addition, remotely-sensed, georeferenced data are often the only information available about an area of interest, or are much less expensive than relying exclusively on primary data collection. Combining remotely-sensed data with selected primary data collection has great promise for addressing many of the questions raised in Section 2.2.

Working with spatial data in economic analysis also provides novel challenges. The volume of data is often much larger than that found in most economic analyses. Instead of potential observations in the 100s or 1000s, the number might be in the millions. For these larger datasets the modeling challenges grow. The data are seldom available in a form that is convenient for analysis so unfamiliar data processing techniques must be learned or the services of the relevant expert found. Econometric analyses that take seconds or minutes for small datasets take hours for extremely large file sizes. Moving data between GIS software and econometric software requires special attention be paid to file structures. Finally, as indicated in the Anselin paper in this issue, the development of formal theory for dealing with spatial dependencies with a limited dependent variable such as land use and its implementation in software packages is still in its infancy.

Finally, these techniques make it possible to perform simulations of the consequences of a variety of policy changes—from infrastructure investments such as roads and harbors to agricultural price and macroeconomic policies that affect relative prices facing land use decision makers. With such simulations it is possible to pinpoint the location of expected changes, desirable and undesirable, reducing the cost of mitigation efforts. It is this ability to make location-specific simulations, still in its early stages, that has perhaps the most potential for relevance to a wide range of policy makers.

Acknowledgements

The authors would like to thank participants in the IAAE and AAEA learning workshops for valuable feedback on the overview material presented here. This material is based upon research supported in part by the Cooperative State Research, Education and Extension Service, US Department of Agriculture, under Project no. ILLU 05-0361 and by the National Aeronautics and Space Administration New Investigator Program in Earth Sciences, NAG5-8559. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors. The authors are solely responsible for the analysis and conclusions presented here.

References

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

Bell, K., Irwin, E.G., 2002. Spatially-explicit micro-level modeling of land use change at the rural-urban interface. Agric. Econ. 27, 217–232.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc., Ser. B 36, 192–236.

Bockstael, N.E., 1996. Modeling economics and ecology: the importance of a spatial perspective. Am. J. Agric. Econ. 78, 1168–1180.

Chen, R.S., 1998. An annotated guide to earth remote sensing data and information resources for social science applications. In: Liverman, D., Moran, E., Rindfuss, R., Stern, P. (Eds.), People and Pixels: Linking Remote Sensing and Social Science. National Academy of Science Press, Washington, DC.

Chomitz, K.M., Gray, D.A., 1996. Roads, land use, and deforestation: a spatial model applied to Belize. World Bank Econ. Rev. 10, 487–512.

Cropper, M., Griffiths, C., Mani, M., 1999. Roads, population pressures, and deforestation in Thailand, 1976–1989. Land Econ. 75, 58–73.

De Pinto, A., Nelson, G.C., 2002. Abstract of correcting for spatial effects in limited dependent variable regression: assessing the value of "ad hoc" techniques. American Association of Agricultural Economics Annual Meetings, Long Beach. Am. J. Agric. Econ. (in press).

DeFries, R.S., Belward, A.S., 2000. Global and regional land cover characterization from satellite data: an introduction to the special issue. Int. J. Remote Sens. 26, 1083–1092.

Deininger, K.W., Minten, B., 1999. Poverty, policies, and deforestation: the case of Mexico. Econ. Dev. Cultural Change 47, 313–344.

Dvorak, K.A., 1992. Resource management by West African farmers and the economics of shifting cultivation. Am. J. Agric. Econ. 74, 809–815.

Fleming, M., 2002. Techniques for estimating spatially dependent discrete choice models. In: Florax, R. (Ed.), New Directions in Spatial Econometrics, vol. 2 (in press).

Geoghegan, J., Cortina Villar, S., Klepeis, P., Macario Mendoza, P., Ogneva-Himmelberger, Y., Roy Chowdhury, R., Turner II, B.L., Vance, C., 2001. Modeling tropical deforestation in the southern Yucatan peninsular region: comparing survey and satellite data. Modell. Agric. Ecosyst. Environ. 85, 25–46.

Greene, W.H., 2000. Econometric Analysis, 4th ed. Prentice-Hall, Upper Saddle River, NJ.

Irwin, E.G., Bockstael, N.E., 2002. Interacting agents, spatial externalities, and the endogenous evolution of residential land use pattern. J. Econ. Geogr. 2, 31–54.

Kaimowitz, D., Angelsen, A., 1998. Economic Models of Tropical Deforestation: A Review. Center for International Forestry Research, Bogor.

Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran's *I*-test statistic with applications. J. Econom. 104, 40.

Maddala, G.S., 1983. Limited-dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge, 401 pp.

Mather, P.W., 1999. Computer Processing of Remotely-Sensed Images: An Introduction, 2nd ed. Wiley, UK.

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

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. J. Appl. Econom. 15, 447–470.

Mertens, B., Lambin, E.F., 2000. Land-cover-change trajectories in southern Cameroon. Ann. Assoc. Am. Geogr. 90, 467–494.

Mertens, B., Sunderlin, W.D., Ndoye, O., Lambin, E.F., 2000. Impact of macroeconomic change on deforestation in southern Cameroon: integration of household survey and remotely-sensed data. World Dev. 28, 983–999.

Morain, S.A., 1998. A brief history of remote sensing applications, with emphasis on Landsat. In: Stern, P. (Ed.), People and Pixels: Linking Remote Sensing and Social Science. Committee on the Human Dimensions of Global Environmental Change, National Research Council. National Academy of Science Press, Washington, DC.

Moran, E.F., Brondizio, E., 1998. Land use change after deforestation in Amazonia. In: Stern, P. (Ed.), People and Pixels: Linking Remote Sensing and Social Science. National Academy of Science Press, Washington, DC.

Müller, D., Zeller, M., 2002. Land use dynamics in the central highlands of Vietnam: a spatial model combining village survey data with satellite imagery interpretation. Agric. Econ. 27, 333–354.

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

Nelson, G.C., Hellerstein, D., 1997. Do roads cause deforestation? Using satellite images in econometric analysis of land use. Am. J. Agric. Econ. 79, 80–88.

Nelson, G.C., Harris, V., Stone, S.W., 1999. Sustainable Development in Panama's Darien Province: Modeling Land Use Change with Spatial Econometric Analysis. Sustainable Development Department, Environment Division, Inter-American Development Bank, Washington, DC, 27 pp.

Nelson, G.C., Harris, V., Stone, S.W., 2001. Deforestation, land use, and property rights: empirical evidence from Darien, Panama. Land Econ. 77, 187–205.

Nelson, G.C., Stone, S.W., De Pinto, A., 2003. Land use and road improvements: a spatial econometric analysis. Int. Regional Sci. Rev. (in press).

Pontius, R.G.J., 2000. Quantification error versus location error in comparison of categorical maps. Photogr. Eng. Remote Sens. 66, 1011–1016.

Richards, J.A., 1993. Remote Sensing Digital Image Analysis: An Introduction, 2nd ed. Springer, Berlin, 340 pp.

Rindfuss, R.R., Entwisle, R., Walsh, S.J., Prasartkul, P., Sawangdee, Y., Crawford, T.W., Reade, J., 2001. Continuous and discrete: where they have met in Nang Rong, Thailand. In: Crews-Meyer, K.A. (Ed.), Remote Sensing and GIS Applications for Linking People, Place and Policy. Kluwer Academic Publishers, Boston.

Singh, I., Squire, L., Strauss, J. (Eds.), 1986. Agricultural Household Models: Extensions, Applications, and Policy. Johns Hopkins University Press, Baltimore, 335 pp.

Swallow, S.K., Talukdar, P., Wear, D.N., 1997. Spatial and temporal specialization in forest ecosystem management under sole ownership. Am. J. Agric. Econ. 79, 311–326.

Vance, C., Geoghegan, J., 2002. Temporal and spatial modeling of tropical deforestation: a survival analysis linking satellite and household survey data. Agric. Econ. 27, 317–332.

Vance, C., Geoghegan, J., 2003. Modeling the determinants of semi-subsistent and commercial land-uses in an agricultural frontier of southern Mexico: a switching regression approach. Int. Regional Sci. Rev. (in press).

Wood, C.H., Skole, D., 1998. Linking satellite, census and survey data to study deforestation in the Brazilian Amazon. In: Stern, P. (Ed.), People and Pixels: Linking Remote Sensing and Social Science. National Academy of Science Press, Washington, DC.