ELSEVIER

Agriculture, Ecosystems and Environment 85 (2001) 25–46

Agriculture Ecosystems & Environment

www.elsevier.com/locate/agee

# Modeling tropical deforestation in the southern Yucatán peninsular region: comparing survey and satellite data

Jacqueline Geoghegan[a,*], Sergio Cortina Villar[b], Peter Klepeis[c], Pedro Macario Mendoza[b], Yelena Ogneva-Himmelberger[d], Rinku Roy Chowdhury[d], B.L. Turner II[d], Colin Vance[e]

[a] *Department of Economics, Marsh Institute, Clark University, 950 Main Street, Worcester, MA 01610, USA*
[b] *El Colegio de la Frontera Sur, San Cristobal, Chiapas, and Chetumal, Quintana Roo, Mexico*
[c] *Department of Geography, Colgate University, NY, USA*
[d] *Graduate School of Geography, Marsh Institute, Clark University, Worcester, MA 01610, USA*
[e] *National Center for Environmental Economics, US Environmental Protection Agency, Washington, DC, USA*

**Abstract**

This paper presents some initial modeling results from a large, interdisciplinary research project underway in the southern Yucatán peninsular region. The aims of the project are: to understand, through individual household survey work, the behavioral and structural dynamics that influence land managers' decisions to deforest and intensify land use; model these dynamics and link their outcomes directly to satellite imagery; model from the imagery itself; and, determine the robustness of modeling to and from the satellite imagery. 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. Both models test hypotheses concerning deforestation during different time periods in the recent past in the region. The first uses the satellite data, other spatial environmental variables, and *aggregate* socio-economic data (e.g., census data) in a discrete-choice (logit) model to estimate the probability that any particular *pixel* in the landscape will be deforested, as a function of explanatory variables. The second model uses the survey data in a cross-sectional regression (OLS) model to ask questions about the *amount* of deforestation associated with *each* individual farmer and to explain these choices as a function of *individual* socio-demographic, market, environmental, and geographic variables. In both cases, however, the choices of explanatory variables are informed by social science theory as to what are hypothesized to affect the deforestation decision (e.g., in a von Thünen model, accessibility is hypothesized to affect choice; in a Ricardian model, land quality; in a Chayanovian model, consumer–labor ratio). The models ask different questions using different data, but several broad comparisons seem useful. While most variables are statistically significant in the discrete choice model, none of the location variables are statistically significant in the continuous model. Therefore, 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. 

*Keywords:* Land-use/cover change; Econometric models; Survey and satellite data; Tropical deforestation; Mexico

* Corresponding author. Tel.: +1-508-793-7709; fax: +1-508-793-8849.
*E-mail address:* jgeoghegan@clarku.edu (J. Geoghegan).

0167-8809/01/$ – see front matter 
PII: S0167-8809(01)00201-8

## 1. Introduction

Be it potential climate warming, loss of biotic diversity and ecosystem services, or improvement of human well being, land-use and land-cover change is a critical element and issue. Its role and implications are so large that the International Geosphere–Biosphere Programme (IGBP) and the International Human Dimension Programme (IHDP) joined to create a shared international program of study on Land-Use/Cover Change (LUCC) (IGBP Report 35–IHDP Report 7, 1995; IGBP Report 48–IHDP Report 10, 1999). The LUCC effort identifies the need to improve understanding, models, and projections of land dynamics, especially at sub-global or regional levels and focusing particularly on the spatial explicitness of processes and outcomes. Tropical deforestation is identified by LUCC and other communities as one of several foci of study because of its current pace and magnitude globally and because of its significance to environmental and human issues noted above (Kremen et al., 2000).

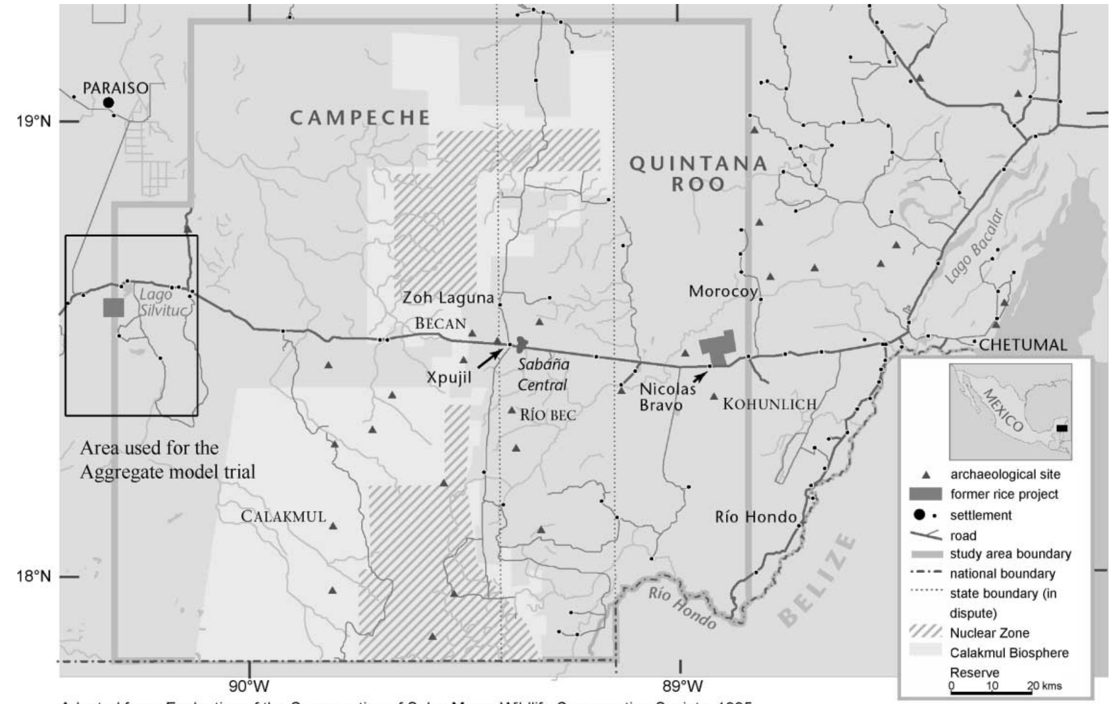
The southern Yucatán peninsular region is exemplary. As part of the largest continuous expanse of tropical forests remaining in Central America and Mexico, it has been identified as a "hot spot" of forest and biotic diversity loss (Achard et al., 1998), despite the establishment of the Calakmul Biosphere Reserve in the center of the region and the large-scale Mundo Maya (Maya World) program that seeks to make large portions of the region an archaeological-ecotourism center (Primack et al., 1998). The veracity of these efforts notwithstanding, the region is occupied by groups and individuals whose livelihood is drawn largely from the land, mostly by cultivation but also by livestock, and various forces seek "to develop" the region further for land-consuming activities.

This paper presents some initial modeling results from a large, interdisciplinary, research effort, the Southern Yucatán Peninsular Region project (SYPR) begun in 1997 and designed to address the kinds of problems identified in the LUCC science and implementation plans (above). These include an integrative understanding of land-change dynamics, drawing from historical analysis, social-science field studies, remote sensing science, and ecology. Part of project aims to: understand, through individual household survey work, the behavioral and structural dynamics that influence land managers' decisions to deforest and intensify land use; model these individual choices and link their outcomes directly to satellite imagery through geographic information systems (GIS); model land-use change using information strictly available from the remotely sensed imagery, with the addition of other GIS and readily available aggregate data; and, then compare these modeling approaches. Of particular methodological interest is to investigate the value added of using GIS and global positioning systems (GPS) in survey research to investigate how the individual chooses land-use practices and how these *explicitly* vary over space and time.

To demonstrate an application of the approach taken in this project, this paper focuses on the land-use change from forestry to agricultural uses. While there are many types of land-use changes occurring in the region, there is a large body of existing land-use/land-cover research that focuses on modeling tropical deforestation using many different modeling techniques (see review, Kaimowitz and Angelsen, 1998). By limiting this analysis to this one type of land-use/land-cover change here, comparisons can be drawn to the existing literature. Many research projects throughout the social sciences test hypotheses concerning land manager's decision-making, especially for small holders, and other researchers use remotely sensed data to test hypotheses econometrically (Chomitz and Gray, 1996; Nelson and Hellerstein, 1997; Pfaff, 1999). Few research projects, however, link survey and satellite data, develop behavioral models, and test the derived hypotheses of land-use choice using these data. [1] Using the survey data and the satellite data, two different econometric models are developed here to test hypotheses concerning the causes of deforestation in the region.

Current deforestation in the southern Yucatán peninsular region follows primarily from agricultural expansion. The four principle agricultural land uses represented currently within the region are: *milpa* (subsistence-oriented swidden agriculture, based on maize [*Zea mays* L.], squash [*Cucurbita* spp.] and legumes [*Phaseolus* spp.]), market-based chili (*Capsicum* spp.) cultivation; failed large-scale rice projects (several of which have been converted to pasture for livestock); and incipient, small-scale orchards. In the recent past, most farmers were small

[1] For other exceptions, see the chapters in Liverman et al. (1998).




Adapted from: Evaluation of the Conservation of Selva Maya. Wildlife Conservation Society. 1995

**Note: The thick gray line represents the boundaries of the general region of study, comprising about 22,005.6 km$^2$.**

Fig. 1. The southern Yucatán peninsular region.

holders or *ejidatarios* primarily engaged in subsistence production;[2] farmers remain largely *ejidatarios* today, but recently an increasing proportion of them engage the market, primarily through chili cultivation. The region today is best considered one dominated by semi-subsistence small holders experimenting with chili and other vegetables, fruits, non-timber forest products, livestock, bee-keeping, and, where possible, engaged in modest off-farm employment.

A more complete understanding of the processes affecting agricultural land-use decisions is necessary in order to understand deforestation. Another part of the project uses the survey data to model and test hypotheses concerning small holder behavior and the amount of land used for agricultural production for each land manager, focusing on the theoretical implications concerning each land manager's integration into the market (Vance, 2000; Vance and Geoghegan, 2000). The models developed in this paper can then be considered as reduced form versions of structural models of deforestation.

## 2. The study region

The southern Yucatán peninsular region is located in the southern portions of the Mexican states of Campeche and Quintana Roo and encompasses approximately 22,000 km$^2$, of which 18,700 km$^2$ of *ejido* lands constitutes the main area of study ("*ejido* assessment region") (see Fig. 1). A semi-evergreen tropical forest dominates this karstic region.[3] Northern Belize, El Petén, Guatemala, and the state of Chiapas, Mexico, combine with the study region to form the largest continuous expanse of tropical moist forest remaining in Central America. Over the course of

[2] An *ejidatario* is a member of an *ejido*, a unit of land originally assigned by the federal government to a group of people for use in usufruct predicated on communal control.

[3] Karst refers to lands dominated by limestone and typified by erosional process of solution, paucity of surface water (e.g., streams), and abundant sinks and cave systems.

human occupation, the region has experienced one wave of boom-bust occupation and land use, starting with the nearly complete deforestation of the region 1000 years ago during the Classic Period of Maya domination (A.D. 100–900). After the collapse of this civilization and the resulting extreme depopulation of the region, the forests returned (Whitmore et al., 1990). A second cycle began in the 1880s with the entry of national and international forest extraction industries in search of logwood (*Haematoxylon campechianum* L.), chicle (*Manilkara zapota* L. Van Royen, formerly *Achras sapota* L.) the tree resin used as a base for chewing gum, and tropical hardwoods primarily, mahogany (*Swietenia macrophyla* King) and cedar (*Cedrela odorata* L.), a prelude to significant immigrant-led agriculture from the 1960s to the present as well as assorted development projects. This second wave of occupation and land-use is in an incipient stage, at least in terms of population.[4]

Forestry dominated land-use through most of the 20th century, but a combination of factors led to a focus on agricultural production rather than forestry, beginning in the late 1960s: a bust in chicle on the international market after World War II; a devastating hurricane cutting across the region in 1955, highlighting the isolation of the area to the federal government; depletion of precious hardwood reserves by the 1960s; statehood for Quintana Roo, which required a minimum population of 80,000; and the identification of the region as a release valve for land stress elsewhere in Mexico with a potential for significant market cultivation. These and other factors led to investment in infrastructure, opening the southern Yucatán peninsular region to rapid development. Government-directed colonization schemes and various mechanized cattle and rice projects led to forest clearing and rapid population growth after 1970. In 1989, both national and international interests led to the creation of the Calakmul Biosphere Reserve, encompassing one-third of the study area (Primack et al., 1998). Since 1993, the creation and improvement of roads as well as other investment in infrastructure and services seek to expand El Mundo Maya, making the region part of one of the largest ecological and archeological tourist schemes in the world, stretching from Mexico to El Salvador.

As the SYPR project is particularly interested in the deforestation–agricultural linkage, further focus is on the factors that led to an increase of agricultural use in the region. The completion of highway 186 between 1967 and 1970 instigated the first influx of large-scale agricultural settlement to the region. Linking the coastal highways circumventing the peninsula as well as the two state capitals and the rest of Mexico, the highway was part of an extensive road-building project across the peninsula that was intended to integrate the peasant economy with local urban centers (Ewell and Merrill-Sands, 1987): in this case, Escarcega, Campeche, and Chetumal, Quintana Roo, both of which grew rapidly after the completion of the road. Whatever the expectations may have been with regard to the region's ability to complement Mexico's urban/industrial sector as a source of food products, the initial infusion of settlers in the 1970s was mostly comprised of subsistence farmers. Their settlement along the highway may have had less to do with linkages to markets than with subsidized access to farmland given a lack of opportunities elsewhere. While the government extended usufruct land grants in the form of collectively managed *ejidos*, as sanctioned by Article 27 of the Mexican constitution, it initially did little to ensure that *ejidatarios* had access to the capital necessary to produce beyond subsistence needs. With land cheap and possessing no apparent comparative advantages to food producing areas elsewhere in Mexico, most of the agriculture in the region was practiced on an extensive basis, using traditional slash-and-burn techniques, locally known as *milpa*.

The major exceptions to these land-uses were government sponsored, large-scale rice projects developed on the eastern and western edges of the region within large *bajos* or seasonal wetlands created by poljes (sediment-filled depressions with hard clay lenses that inhibit subsurface drainage). The aim was to use the annual rise and fall of water within the *bajos* coupled with mechanized cultivation to grow rice for the market. Each of these schemes failed, and with Mexico's debt crisis of the early 1980s, were no longer supported by government funds. Much of these deforested *bajos* were converted to pasture for cattle, as were smaller, upland plots of land elsewhere. In addition to the rice

[4] At the zenith of Maya occupancy, the southern Yucatán peninsular region maintained population densities that approached or exceeded 100 people/km$^2$ (Whitmore et al., 1990). Today the region averages just over 2 people/km$^2$, although population growth rates approach 4% per annum (Ericson et al., 1999).

projects, there were also government-sponsored cattle initiatives in the *bajos* of select *ejidos*, however, these also failed due to inadequate water control, disease, distance to market, and debt (Klepeis, 2000).

Although the *ejidos* of Mexico were established under the same federal act that declared their community structure (no private holders), considerable variation exists among *ejidos* in terms of the rules of access to land. *Ejidos* in the region are relatively new, compared to the rest of Mexico, and most of them operate under an usufruct system in which individual *ejidatarios* maintain more-or-less permanent use rights to the parcels that they cultivate. This system translates into a practice in which any given parcel is controlled by one farmer or household over an extended period of time (e.g., decades), such that household decision-making is predicated on the assurance that they control certain parcels. This kind of household control is evident even in those few *ejidos* in which "informal" recognition of parcel control is absent. Thus, previous and subsequent to recent changes in the Mexican Constitution concerning land tenure, the farmers of the region have held relatively strong usufruct rights to their parcels. This tie between household and parcels permits land-use decisions to be linked to the geographical locations in which they have impact.

Households in the region are not of the autarkic type characterized in the classic Chaynovian model: allocating family labor solely to on-farm tasks with no reliance on hired labor. To the contrary, the overwhelming majority of households in the SYPR survey sample (described below) participate in the market as sellers and/or buyers of labor, with only 6% households completely self-sufficient and slightly over half both selling and hiring labor at some times in the year. Among the 80% of households selling labor off the farm, the majority participated as day laborers on the farms of neighbors in the village. But it is also common for family members to leave the village temporarily to take advantage of employment opportunities, which are often afforded by efforts of the government to develop the region as a tourist zone (e.g., restoration work in the local ruins or in road and hotel construction). Less than 10% of the sample households have members working full-time off the farm throughout the year.

By the mid-1980s, a radical revision of economic policies toward greater liberalization was underway that would be bolstered by legal reforms beginning in the following decade. In 1986, Mexico entered into the General Agreement on Tariffs and Trade (GATT), the impact of which reached the agricultural sector by 1990, when tariffs on most products were dropped or drastically lowered, subsidies on inputs were withdrawn or sharply reduced, and the guarantee price was eliminated for all crops but maize and beans (Foley, 1995). The continuation of these reforms was secured under the terms of the North American Free Trade Agreement (NAFTA) effective in 1994, obligating Mexico to liberalize fully its agriculture, including maize and beans, within a 15-year period. On the legal front, Article 27 of the constitution, which had served as the embodiment of the government's commitment to the rural poor since the end of the Mexican Revolution in 1917, was amended in 1992 to (1) permit lands formerly held in usufruct under the *ejido* system to be bought and sold, (2) open the possibility for joint ventures between *ejidos* and private interests, and (3) terminate the continued distribution of land to peasant communities (Goldring, 1995). It was anticipated that these neoliberal reforms would, in the words of President Salinas, both "capitalize the countryside and open productive options" by establishing a legal framework guaranteeing property ownership (cited in Foley, 1995, p. 62). The impacts of these changes on land access in the southern Yucatán area are not yet clear, aside from initiating comprehensive survey and mapping of *ejido* boundaries (Klepeis and Vance, 2000).[5]

[5] The impacts of recent reforms on the *ejidatarios* of the region are not fully understood and are the subject of on-going study. As elsewhere in Mexico, *ejidatarios* may receive certificates specifying their rights to the use of *ejido* land, rights which can be transferred to another family member or sold to another member of the same *ejido* with *ejido* assembly permission (Stephen, 1998). Many, if not most, of the *ejidos* in the study region, grant this land by designating specific parcels to the *ejidatario*, who must use these lands and no others. "Parcelization" is an *ejido* function, not part of the Mexican certification process. If given assembly permission, an *ejidatario* may receive title to the said parcels, in which case the property ceases to belong to the ejido, and can be bought and sold outside the *ejido* structure. To our knowledge, no privatization by *ejidatarios* has taken place within the southern Yucatán region, although many, if not most, *ejidatarios* have obtained certificates to their *ejido* lands. It is also noteworthy that *ejidatarios* are technically restricted from cutting forest growth of more than 20 years, even if it exists on their land.

## 3. Data methodology

### *3.1. The survey design and questionnaire*

The survey data used in this study were collected by the project over an 11-month period beginning in October of 1997. To accommodate the limitations of available sampling frames and control data collection costs, selection of respondents proceeded according to a stratified, two-stage cluster sample, with the first stage unit being *ejidos* (approximately 125 in the region) and the second stage unit being *ejidatarios* (Warwick and Lininger, 1975; Deaton, 1997). Completion of the first stage selections involved compiling an enumeration of *ejidos* in the region with the aid of 1990 cartographic sources and census records, both published by the Mexican government's National Institute of Statistics and Geography (INEGI). Using these maps, the region was partitioned into 11 geographic divisions — or *strata* — and randomly selected one *ejido* from each of the strata. Each *ejido* was assigned a probability of selection equal to the ratio of its population to the population of the stratum. The strata were drawn to ensure that *ejidos* from across the region were represented in the sample, thereby capturing variability in both ecological conditions and in the influence of market and road proximity. Ten out of the 11 *ejidos* selected have assigned parcels, in which *ejidatarios* have usufruct access to a specific area of land. Settlers from some 23 states in Mexico have colonized the Southern Yucatán Peninsular Region and the sample captures this regional and ethnic diversity.

In the second stage, the survey respondents themselves were randomly selected from each *ejido* after an inventory of households was enumerated. This inventory was secured after establishing contact with the *comisariado* (community leader) of each *ejido*, who, along with the *ejido* assembly, granted permission to proceed with the surveying. The number of households surveyed from each *ejido* was approximated such that the corresponding stratum was represented in roughly the same proportion as its share of the total population of the study zone. After eliminating 12 records collected during a pilot phase of the survey, the final sample size was 188 observations.

A standardized questionnaire was used to elicit the socio-economic and land use data. The questionnaire was administered to the household head and was comprised of two sections. The first section elicited information on migration history, farm production and inputs, off-farm employment participation, land acquisition, and the demographic composition of the household.[6] Completion of the second section involved a guided tour of the respondent's agricultural plot. Using a GPS, the interviewer created a geo-referenced sketch map detailing the distinct fields within the plot as the farmer-respondent provided an interpretation of the use of these fields. By visiting the respondent's parcels to draw the sketch map, a more open dynamic between researcher and respondent was established. Via a process akin to "participant observation", the confidence of the respondent in the researcher was heightened and data from the in-house survey was revised and new information on land-use dynamics was obtained. The sketch map not only documented the spatial configuration of contemporary land uses, including forested areas, but also the land use transition histories of the principal areas of activity. This information was later used to calculate various indices of land use change, such as the area deforested since acquisition of the plot, as well as to inform the classification and interpretation of the satellite imagery.

### *3.2. Satellite data*

Landsat TM images from several dates between 1988 and 1997 were used for this study.[7] Four scenes have been processed for the project for different dates.[8] The individual images were first

[6] As the region was, until recently, a forested frontier, much of the population growth has been generated by migrants largely appended to the development of *ejidos*. The application for a new *ejido* and the resulting migration follows from push and pull factors — land pressures elsewhere in Mexico and minimal non-farm economic opportunities for the rural impoverished as well as land give-a-ways in the region of study and, in the past, the lure of government sponsored projects.

[7] Some of the steps used by the SYPR project are somewhat novel to Landsat TM analysis and are the subject of various publications under preparation. Details of the steps used can be found on the SYPR project web site: <earth.clarku.edu/lcluc/>.

[8] The scenes and dates are: path20/row 47 (4/1/1987, 4/27/1988, 5/8/1992, 10/29/1994, 5/22/1995, 2/5/1996); path 19/row 47 (1/14/1985, 1/4/1987, 11/7/1994, 1/31/1997); path 20/row 48 (12/7/1988, 12/13/1993, 2/5/1996); path 19/row 48 (11/11/1984, 2/21/1993, 1/31/1997).

georeferenced to a latitude–longitude projection (Datum NAD27, Mexico) to under 0.5 RMS positional error. Following image registration, the TM scenes were subjected to haze removal using ERDAS Imagine's Tasselled Cap transformation, reducing the haze-related atmospheric noise in each of the dates processed. Using the de-hazed red and infrared bands, NDVI images were produced for each date for subsequent analysis. The de-hazed images were then processed using principal components analysis (PCA) to complete noise removal and reduce data redundancy. PCA was run separately on the visible and near infrared bands, and the first three principal components from the separate analysis were retained for further processing. The higher order components captured random as well as systematic noise in the data, such as striping, and were dropped from further analysis.

The next steps consisted of performing texture analysis on the 3-band principal components image. By generating measures of spectral variance in immediate spatial neighborhoods of individual pixels, texture analysis produced three additional bands of information based on the original 3-band PCA image. The three texture bands were stacked with the three PCA bands to produce a 6-band image. Next, the NDVI image produced from the originally de-hazed red and infrared bands was added to the PCA and texture bands to generate a final 7-band image for signature development and classification. Signature development involved the creation of training sites for land cover classes of the target classification scheme. Ground truth data were derived from a variety of sources: GPS assisted field visits, topographic maps, vegetation and land-use maps, and detailed sketch maps on recent land-use history collected during field research. After the development of training sites, the corresponding land-cover signatures were evaluated using accepted measures of separability such as Euclidean distance, divergence, transformed divergence, and Jefferies–Matusita distance, and further refined. The final signatures were then used in a maximum likelihood supervised classification to produce categorical land-use/cover maps for change detection and modeling.

The final classification scheme has 11 classes: water; savanna; herbaceous wetland vegetation ("tular"); seasonally inundated, semi-evergreen lowland forest (*selva baja*); semi-evergreen upland forest (*selva mediana*); herbaceous secondary vegetation (approximately 1–3 years regrowth); shrub dominated secondary vegetation (approximately 4–9 years regrowth); tree-dominated secondary vegetation (over 10 years regrowth); cropland; pasture (clear, well-maintained); and *Pteridium aquilinum* (L.) Kuhn (a bracken fern that invades apparently after overcropping *milpas* and after which farmers tend to abandon the invaded plot). [9]

### 3.3. Other data sources

Other spatial data collected or created for use in the satellite imagery model include: elevation and slope from a digital elevation model; soil types digitized from a 1:250,000 Mexican government (INEGI) map; digitized road network from INEGI 1:50,000 topographic maps; and socio-demographic data from Mexican government censuses.

## 4. Modeling

Two econometric modeling approaches are used in this paper. Both models test hypotheses concerning deforestation during different time periods in the recent past in the southern Yucatán peninsular region. The first uses the satellite data, other spatial environmental variables, and *aggregate* socio-economic data (e.g., census data) in a discrete-choice (logit) model to estimate the probability that any particular *pixel* in the landscape will be deforested, as a function of explanatory variables. The second model uses the survey data in a cross-sectional regression (OLS) model to ask questions about the *amount* of deforestation associated with *each* individual farmer and to explain these choices as a function of *individual* socio-demographic, market, environmental, and geographic variables. The

[9] Subsequent to development of the modeling exercises for this paper, the SYPR project has concluded that for region-wide assessments seven classes will be used: seasonally inundated, semi-evergreen lowland forest; semi-evergreen upland forest; secondary vegetation (4–15 years); agricultural use (cropland, pasture, and 1–3 years secondary growth), bracken fern, inundated/semi-inundated Savannas; water. For future household assessment linked to the sketch maps, however, the classes used in this exercise will be followed.

two geographic regions covered by the two models differ, as the satellite data model is based on an initial exploratory "window" dataset derived before the final boundaries of the current study regions were defined (see Fig. 1). In both cases, however, the choices of explanatory variables are informed by social science theory as to what are hypothesized to affect the deforestation decision (e.g., in a von Thünen model, accessibility is hypothesized to affect choice; in a Ricardian model, land quality; in a Chayanovian model, consumer–labor ratio). Finally, the period of assessment (below) permits us to disregard the possibility of land cleared for large-scale rice and cattle projects; such deforestation took place earlier than the period under consideration in this study. [10]

### 4.1. Satellite and aggregate data model

The satellite data model presented comprises a 1600 $km^2$ area of 16 *ejidos* situated on the western edge of the study region (Fig. 1). [11] This window is employed because the larger project involves various stages of data collection and processing that require adjustments based upon results of modeling trials. The window in question was one of the first in which sufficient, but not yet complete, data to run a trial were available. This trial satellite data model, therefore, should be viewed as a step towards the more complete model(s) to follow in the future.

As described above, there are many target classes, but for this model the satellite data are aggregated into three land-uses/covers: forest (includes all extant forest and regrowth); agriculture (includes cropland and pasture); and other (all remaining land-use/cover categories), to develop a dataset for a simple binary choice model of "remain forested" (all forest and regrowth) and "deforested" (land use change into agriculture).

The statistical model developed for this is a simple binomial logit model of these two choices, where the unit of observation is the TM pixel. Assume that there is an underlying response variable $y_i{}^*$ that is given by the following regression:

$$y_i{}^* = \beta' x_i + u_i$$

where the $y_i{}^*$ is the economic returns from a deforested pixel, the $x_i$'s are the exogenous (explanatory) variables, the $\beta$ are coefficients to be estimated, and the $u_i$'s are random errors. Clearly, the economic returns from a pixel is not observable from space, so $y_i{}^*$ is actually unobserved, therefore a dummy variable, $y$, is created:

$$y = \begin{cases} 1 & \text{if } y_i{}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

That is, a deforested pixel is observed if the economic returns to deforestation are greater than zero. If it is assumed that the cumulative distribution function of the $u_i$'s are distributed as a logistic distribution, then this is a binomial logit of the form (see Maddala, 1983):

$$\Pr(Y = 1\ \text{(i.e., deforested)}) = \frac{\exp(\beta' x)}{1 + \exp(\beta' x)}$$

One of the great strengths of modeling with satellite data and GIS is not just simply as a source of the endogenous (dependent) variable, but also the ability to create other explanatory variables from the data and data overlays. As is often done, measures such as distance from pixel to road, market, and so forth were created, but in addition spatial indices of the land uses surrounding a pixel to capture the effect of diversity or fragmentation in land-use in an area were calculated. [12] This is an attempt to go beyond simple one-dimensional measures of space, and try to capture more of the spatial complexity of the landscape in the model and to control for how they might affect choice.

[10] For the area included in the satellite model in this paper, the land under the rice and cattle projects remained cleared of forest and converted to pasture for the period of assessment. For the survey data model, however, the one *ejido* sampled that had experienced such project-driven deforestation permitted significant portions of that cleared land to revert to secondary growth. A portion of this secondary growth area has just recently been converted to crop uses.

[11] This smaller test site used a cloud-free TM scene path 20/row 47 from 4/27/1988, 5/8/1992, and 3/22/1995.

[12] See Geoghegan et al. (1997) for the first use of these measures in an econometric model.

The variables included in the model for each individual pixel are: elevation; slope; soil type;[13] initial type of forest cover;[14] distances to road, village, market; distance to nearest agricultural land use; diversity and richness of neighboring land uses;[15] population, population density, cattle density, and truck density of the *ejido* that the pixel is in, as well as the changes in the levels of these variables.[16] These variables are included in the model for two reasons: (i) various theories of agricultural and land-use change indicate that they will affect deforestation, such as environmental variables affecting the choice of particular areas within a parcel for agricultural use (i.e., Ricardo), distance measures to affect the amount of integration into the market system, which in turn can affect the amount of land under agricultural production (i.e., von Thünen), population and population density to control for local demand for agricultural production (i.e., Chayanov, Boserup (1965), induced intensification), and cattle and trucks to be proxies for wealth (i.e., structural themes) which affects farmers' choices;[17] and (ii) at this stage in the SYPR project these variables were the most accessible (improved variables are under construction).

Separate logit models for each of the two time periods (1988–1992, 1992–1995) were performed for the northern half of the sample in each time period. A separate model for each time period is estimated to compare the estimated coefficients from the two time periods, to test if the effects of the exogenous variables are consistent over time. As the first time period is longer than the second, one would expect the coefficients to change in size; here, the test is simply if any estimated coefficient changes sign over the two time periods. The estimated coefficients from each of these models can be found in Tables 1 and 2, respectively. As this is a discrete choice model, the signs on the coefficients are interpreted (given statistical significance) that a positive coefficient means an increase in the probability of deforestation and a negative sign, a decrease in the probability.

The overall fits of the models are decent for this type of analysis. The pseudo-$R^2$ for the 1988–1992 model[18] is 0.25, and for the 1992–1995 model is 0.37. Most of the signs on the coefficients remain consistent over time and meet expectations: the higher the elevation, the smaller the probability of deforestation;[19] the further a pixel is from the road, the less

[13] The Mexican (INEGI) soil maps for the region are complex, offering many geographical "clusters" of soils based on order, type plus other characteristics, and multiple types headed by the primary. The uplands are dominated everywhere in the region by rendzinas or rendolls (mollisol order) of various qualities; this base type forms over limestone, is high in organic matter and soft when dry, and maintains a base saturation of greater than 50%. Exceptionally thin and rocky soils, often near the top of hills, are classified as litosols, and are largely avoided for cultivation. The bajo or seasonal wetland soils are almost everywhere of the order vertisols; such soils are heavy, cracking clays in excess of 35% clay, and display various degrees of shrinking and swelling based on their seasonal moisture. For this exercise, the many types and subtypes have been aggregated into three that reflect the basic conditions facing farmers: rendzina and other mollisols, and vertisols.

[14] Five dummy variables: seasonally inundated, semi-evergreen lowland forest; semi-evergreen upland forest; herbaceous secondary vegetation; shrub dominated secondary vegetation; tree-dominated secondary vegetation. These are included in the model in an attempt to control for the dynamic fallow cycle decisions.

[15] For these two variables, using all original 11 classes, a $3 \times 3$ window of each pixel was used to calculate each variable: Diversity $= -\sum_{\text{classes}} p \ln p$, where $p$ is the proportion of each class in the window;

$$\text{Richness} = \frac{\text{number of different classes present}}{\text{maximum number of classes}} \times 100.$$

[16] The data are derived from the 1980, 1990 and 1995 censuses, using a linear extrapolation to estimate the values of the variables for 1988 and 1992. Each pixel in each of the 16 *ejidos* were assigned the average value for these census variables. We recognize that change in population is potentially endogenous to change in forest cover, which would produce biases in the estimated coefficients. As Pfaff (1999) suggests, however, it is plausible that population does not result from the clearing process itself but rather is a function of other explanatory factors such as roads or environmental conditions. Since such a relationship implies multicollinearity, the potential problem is really one of sufficient sample size.

[17] Most of these theories and themes are self-explanatory. For induced intensification see Turner and Brush (1987), Turner and Ali (1996).

[18] Similar to the tradition $R^2$ from an OLS regression, the pseudo-$R^2$ is a measure of goodness-of-fit. There are many measures of goodness-of-fit for discrete choice models. Here, the measure used is: $1-(\log L_\Omega)/(\log L_\omega)$, where $L_\Omega$ is the maximum of the unconstrained likelihood function and $L_\omega$ is the maximum of the constrained likelihood function. The two models are not nested models, so one cannot directly compare the psuedo-$R^2$ from each model.

[19] Higher elevated lands tend to be related to more rugged the terrain, shallow soils and caprock, and less soil moisture, all factors inhibiting yields.

Table 1
Binomial logit model of deforestation, time period 1988–1992. Unit of observation: the pixel ($n = 805,898$)

| Variable name | Estimated coefficient | $z$-Statistic |
|---|---|---|
| Constant | 2.112 | 42.026 |
| Elevation | −0.021 | −87.615 |
| Slope | 0.037 | 14.799 |
| Soil 2 | −0.619 | −25.046 |
| Soil 3 | −0.586 | −64.146 |
| Bajo | −1.043 | −39.519 |
| Mediana | −0.216 | −8.580 |
| Secondary 2 | 0.327 | 13.333 |
| Secondary 3 | −0.727 | −14.146 |
| Distance to road | −0.021 | −102.072 |
| Distance to market | 0.001 | 4.218 |
| Distance to village | 0.305 | 16.553 |
| Distance to closest agriculture | −0.001 | −79.685 |
| Diversity | 0.111 | 8.740 |
| Richness | −0.001 | −0.339 |
| Population | −0.002 | −32.011 |
| Population change | 3.415 | 18.249 |
| Population density | 23.449 | 44.716 |
| Cattle density | −2.321 | −24.728 |
| Cattle density change | 9.393 | 15.985 |
| Truck density | 172.765 | 19.791 |
| Truck density change | −5.904 | −10.451 |
| Psuedo-$R^2$ | 0.25 | |

Table 2
Binomial logit model of deforestation, time period 1992–1995. Unit of observation: the pixel ($n = 744,530$)

| Variable name | Estimated coefficient | $z$-Statistic |
|---|---|---|
| Constant | −1.453 | −13.572 |
| Elevation | −0.013 | −26.044 |
| Slope | 0.014 | 2.828 |
| Soil 1 | 0.766 | 18.306 |
| Soil 3 | −0.150 | −3.734 |
| Bajo | −1.005 | −23.472 |
| Mediana | −0.673 | −11.929 |
| Secondary 1 | 0.664 | 23.836 |
| Secondary 2 | 1.127 | 35.564 |
| Distance to road | −0.007 | −15.072 |
| Distance to market | 0.002 | 7.452 |
| Distance to village | −0.215 | −5.998 |
| Distance to closest agriculture | −0.005 | −60.516 |
| Diversity | 0.274 | 12.617 |
| Richness | −0.075 | −33.309 |
| Population | −0.005 | −27.462 |
| Population change | 27.347 | 16.750 |
| Population density | 14.639 | 18.625 |
| Cattle density | −5.280 | −37.591 |
| Cattle density change | −51.684 | −24.302 |
| Truck density | 252.249 | 17.052 |
| Truck density change | 78.919 | 39.872 |
| Pseudo-$R^2$ | 0.37 | |

likelihood of deforestation; the closer a pixel is to a market or a village, the greater the probability of deforestation. The coefficient on distance to nearest cropland is negative and statistically significant in both models; this result is consistent with that expected from farmers who, in most cases, are assigned a segment of an *ejido* to crop and pursue a crop-fallow cycle on the lands assigned. Other results show that the deforestation pixels are more likely to occur on early secondary regrowth, capturing an important dynamic of the fallow cycle in the region. The diversity and richness of surrounding land uses are statistically significant, but further ecological field research is necessary in order to interpret these more completely.

While the results above met a priori expectations, the results on the slope coefficient are counter-intuitive: the greater the slope, the greater the probability of deforestation. [20] The results on the cattle variable are also counter-intuitive: the higher the cattle density, the decreased probability of deforestation. This result differs by the period considered, however. In the 1988–1992 model, the change in cattle density is consistent with expectations: the greater the increase in the change in this variable, the greater the probability of deforestation. This result does not hold for the 1992–1995 model. The reasons for this last result are under investigation but may involve an intensification of cattle on the rice project-land converted to pasture before the assessment period investigated here and the inability of locals to expand pasture owing to the establishment of the Calakmul Biosphere Reserve. Also, the coefficient on population appears counter-intuitive, but the coefficient on population density meets the a priori expectation. That is, the higher the population density, the higher the probability of deforestation, so that the population–land ratio is important for the amount of forest change.

Prediction in a logit model is different than in a regression model, in that the prediction for any observation is a *probability* of an event or choice occurring.

[20] This results is being explored in several ways, as this may reflect the distinction between lowland and upland forests that is not currently controlled for in the model.

In the model here, the probability of a particular pixel becoming deforested was estimated. In the "actual" event, either a pixel was deforested or not, so there is no clear way of "creating" residuals to compare predicted value versus actual value. That is, the prediction is a probability of change, while the "actual" is either "yes" the pixel did change, or "no" it did not.[21] In addition, as the prediction results in a probability that ranges from 0 to 100% (in theory), what is the critical value of the percent so that the pixel "counts" as a deforested pixel?

Often researchers take 50% as the critical threshold, but a different approach is taken here, that can conceptually be divided into two steps. Instead of imposing an ad hoc percentage to consider a pixel a "deforested" pixel, first the total number of actual pixels, $n$, that were deforested were identified. Secondly, the spatial distribution of highest probability pixels were sequentially identify until $n$ have been selected. The results of this are presented in the maps below for the in-sample prediction (Fig. 2 for 1988–1992, corresponding to the model results in Table 1 and Fig. 3 for 1992–1995, corresponding to the model results in Table 2). Hence, this model does not attempt to predict the actual amount of deforestation, but takes that amount as given, and predicts the spatial distribution of that deforestation over the landscape. Using this technique, then, pixels having less than a 50% chance of being deforested could conceivably be included in that distribution. For time period 1, this critical value of probability was 31%, for the second it was 42%. For the out-of-sample predictions (the one half observations in the southern half of the scene not used in the estimation) a similar approach was taken in predicting deforestation, using the estimated coefficients from the models and predicting to probability of deforestation for each pixel in the southern scene with that pixel's explanatory variables. A second map for each time period is created using the critical value from the in-sample predictions as the critical value for the out-of-sample predictions. The out-of-sample prediction maps are Figs. 4 and 5 for 1988–1992 and Figs. 6 and 7 for 1992–1995 (see Table 3 for associated data).

By visually inspecting Figs. 2 and 3, some intuition is gained on what could be potential problems and limitations of the current model. These maps contain the roads layer in addition, as the roads have been used to create a number of important spatial variables in the models. The importance of the road network for the creation of agricultural land is clear, which is not surprising, given the statistical strength of that variable in the models. Probably the most striking feature of the difference between the 1988–1992 and 1992–1995 predictions is the dispersed nature of the pixels that were predicted to remain forested that were in actuality deforested (the lightest gray pixels in Figs. 2 and 3). Perhaps this is due to the much larger estimated coefficient on the distance to nearest agricultural plot in the 1992–1995 model, but clearly much more research into the underlying spatial process of the decision-making is needed (e.g., resource institutions, assigned lands within an *ejido*, and so forth).

The same general observations hold for the out-of-sample prediction maps (Figs. 4 and 5 for 1988–1992 and Figs. 6 and 7 for 1992–1995): the earlier time period is much more dispersed in the incorrect predictions. As discussed above, two approaches in deriving the "critical probability value" were used for the out-of-sample prediction maps. First, for each time period, the in-sample critical threshold probability was taken and applied to the out-of-sample predictions. These results can be seen in Fig. 4 for 1988–1992 and Fig. 6 for 1992–1995. The same approach for the in-sample predictions was then used as well for the out-of-sample predictions: the total number of pixels that changed in the southern scene were used to create the threshold probability value. That is, in Fig. 5 for 1988–1992 and Fig. 7 for 1992–1995, the total number of change pixels matches the actual number of change pixels, but the spatial distribution of those changes is predicted from the estimated model. What is interesting here is that for 1988–1992, the threshold probability decreases from 31 to 26%, meaning that using the threshold probability from the estimated model resulted in not enough total change being predicted, while for 1992–1995 the threshold probability increases from 41 to 58% so that in this time period there is an over-prediction of the total

[21] As compared to the regressions performed in the next section on amount of deforestation per farmer, the estimated equation can be used for prediction, as there is an "actual" amount of deforestation from the survey data, so the "actual" amount of over-prediction or under-predictions (or predicted residuals) can be calculated.

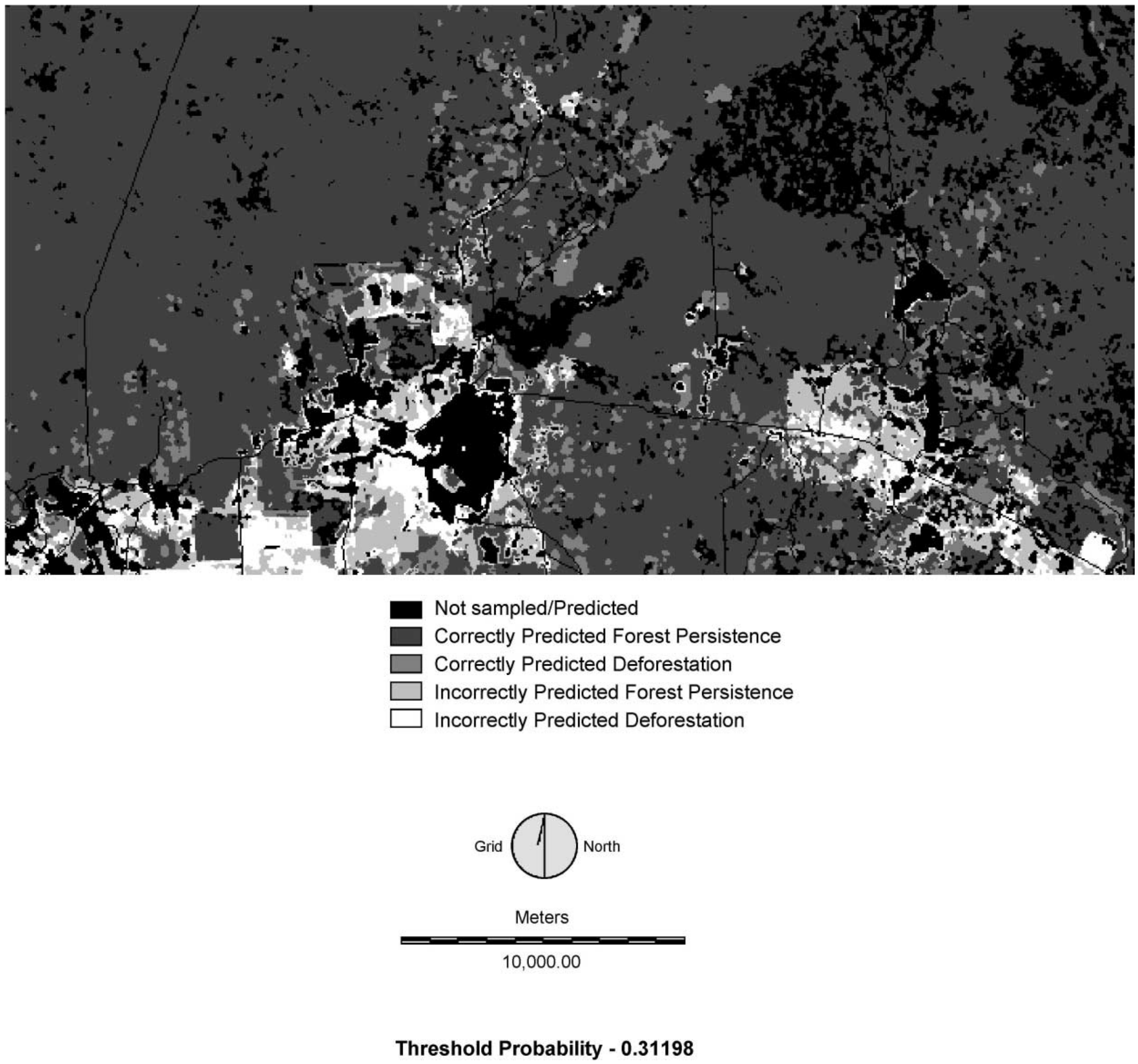



Fig. 2. Actual and predicated deforestation 1988–1992 (in sample).

amount of change using the threshold value from the in-sample predictions. These errors demonstrate a need for a further understanding of the underlying dynamic process of deforestation, to complement the need for further understanding of the underlying spatial process. As a reminder, however, that this modeling exercise is exploratory and the SYPR project has been developing the means to refine the understanding of the processes in question.

### 4.2. *Survey data*

Next, a model was estimated using the survey data that as closely as possible "matched" the satellite data model. This step tests whether the same variables that increase the probability of deforestation in the aggregate model (above) have the same general effect in explaining the amount of deforestation that an individual land manager (*ejidatarios*) undertakes. To

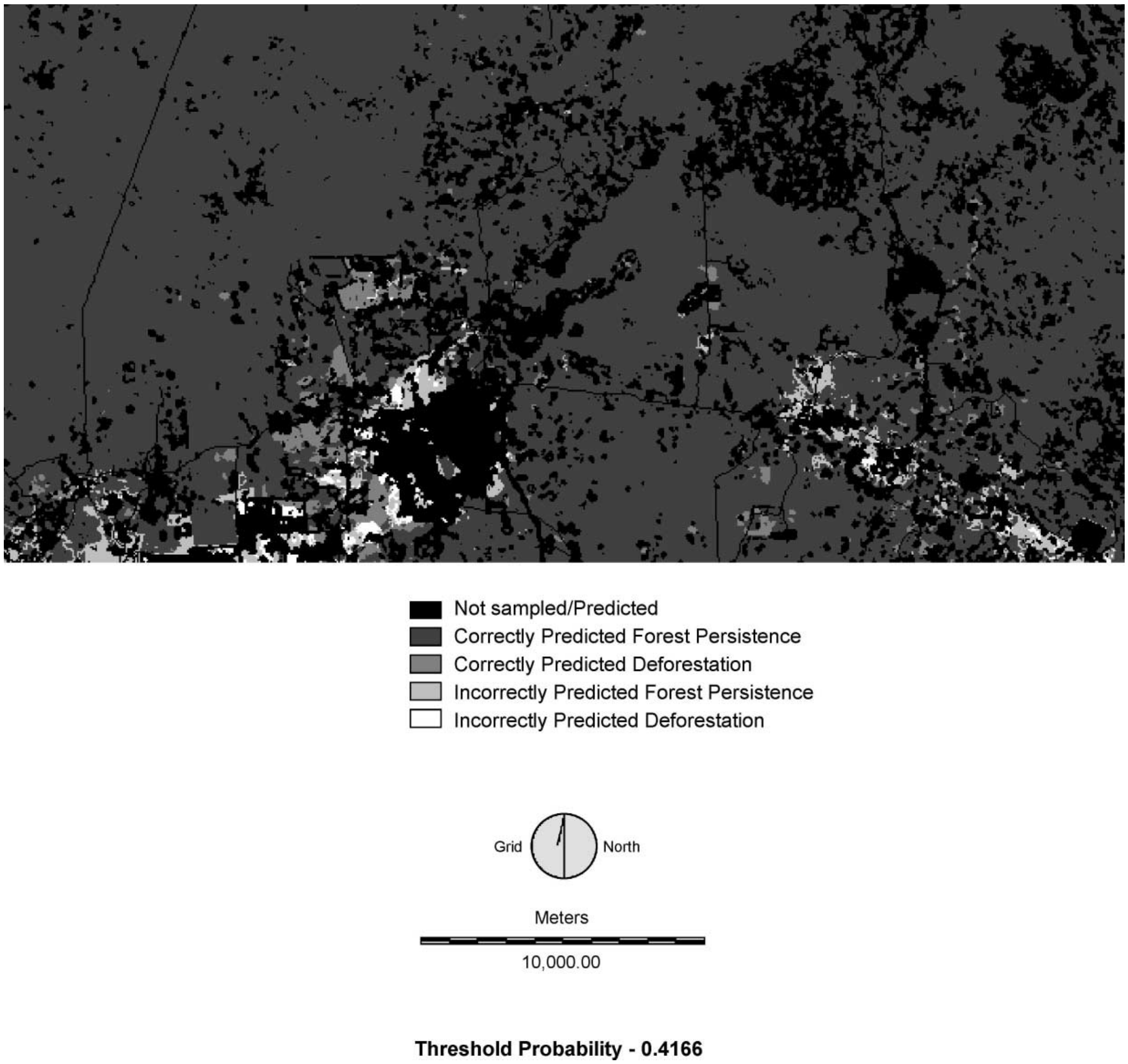



Fig. 3. Actual and predicated deforestation 1992–1995 (in sample).

create the dependent variable for this model, the total amount of deforestation was calculated for each *ejidatarios* during the previous 5 years, calculated from the survey responses to questions concerning the total land cleared during the previous 5 years, resulting in a variable which includes the amount of old growth and secondary forest cut during 1993–1998. As this is a continuous dependent variable, a traditional ordinary least squares regressions was used, modified to take into account the survey design strategy.[22] It should also be remembered that the *ejidatarios* throughout the region display mixed production — more-or-less self-sufficiency in maize, beans, and cucurbits, and

[22] As explained above, a stratified random, not random, sample was employed, and this kind of sample must be taken into account in the estimation procedure.

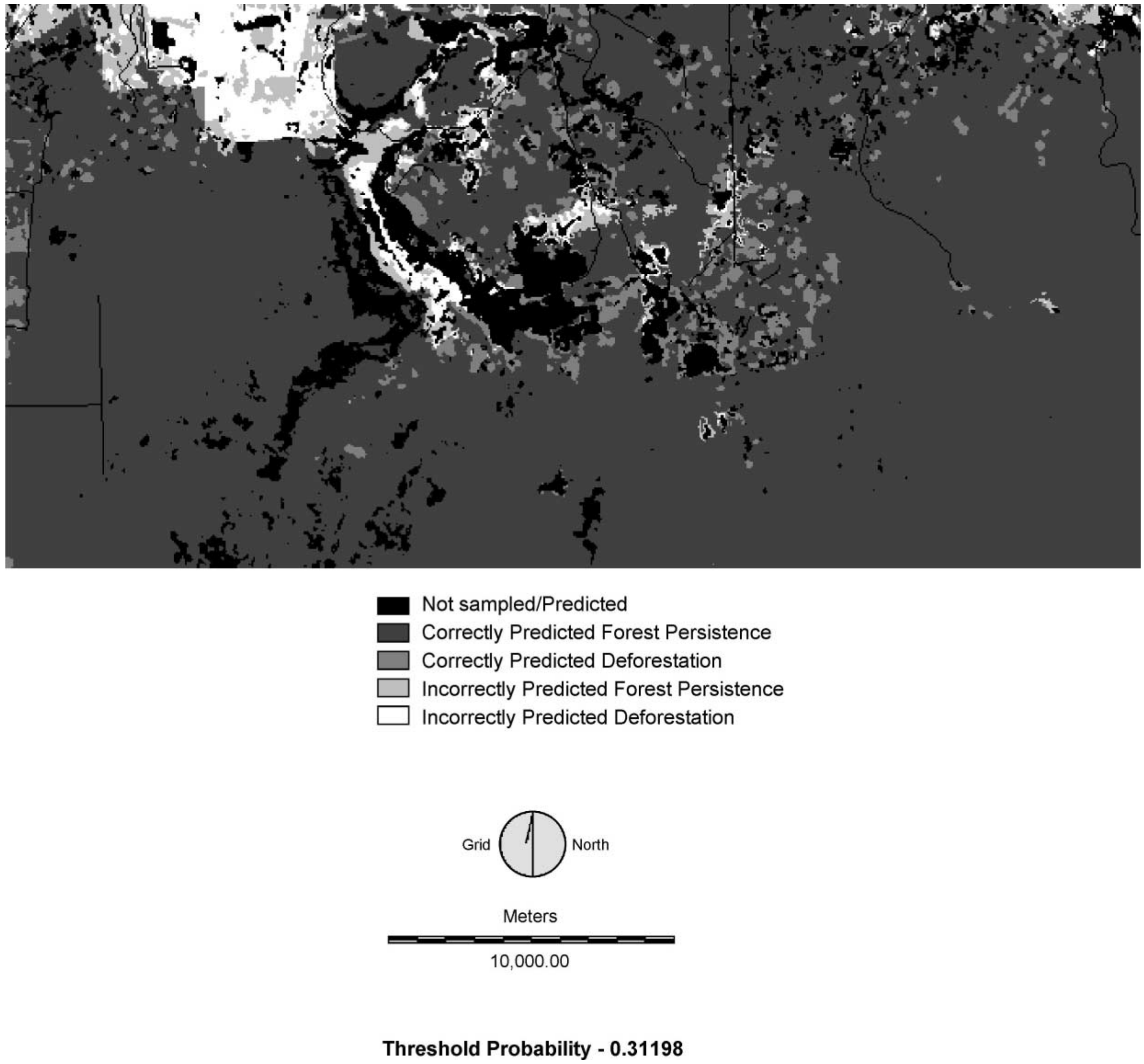



Fig. 4. Actual and predicated deforestation 1988–1992 (out of sample).

experimental market crops — and hence the exploration of variables linked to both production types.

The survey data variables that are "closest" to the satellite data model variables are:[23] elevation; slope; soil type (derived from the same sources used for the satellite data model above);[24] distance from house to agricultural plot; distance from plot to major road;

[23] The survey data from the two southern *ejidos* are not included as currently data is missing on elevation, slope and soil type for those observations.

[24] In the satellite data models, soil types were aggregated into three variables, as noted in Footnote 12. For the survey data model the soil types are reduced to two because areas dominated by litosols were not apparent among the lands cultivated by the households (as would be expected). The two types used, therefore, are rendzina and gleysol–vertisols, both of which were apparent on household controlled lands.

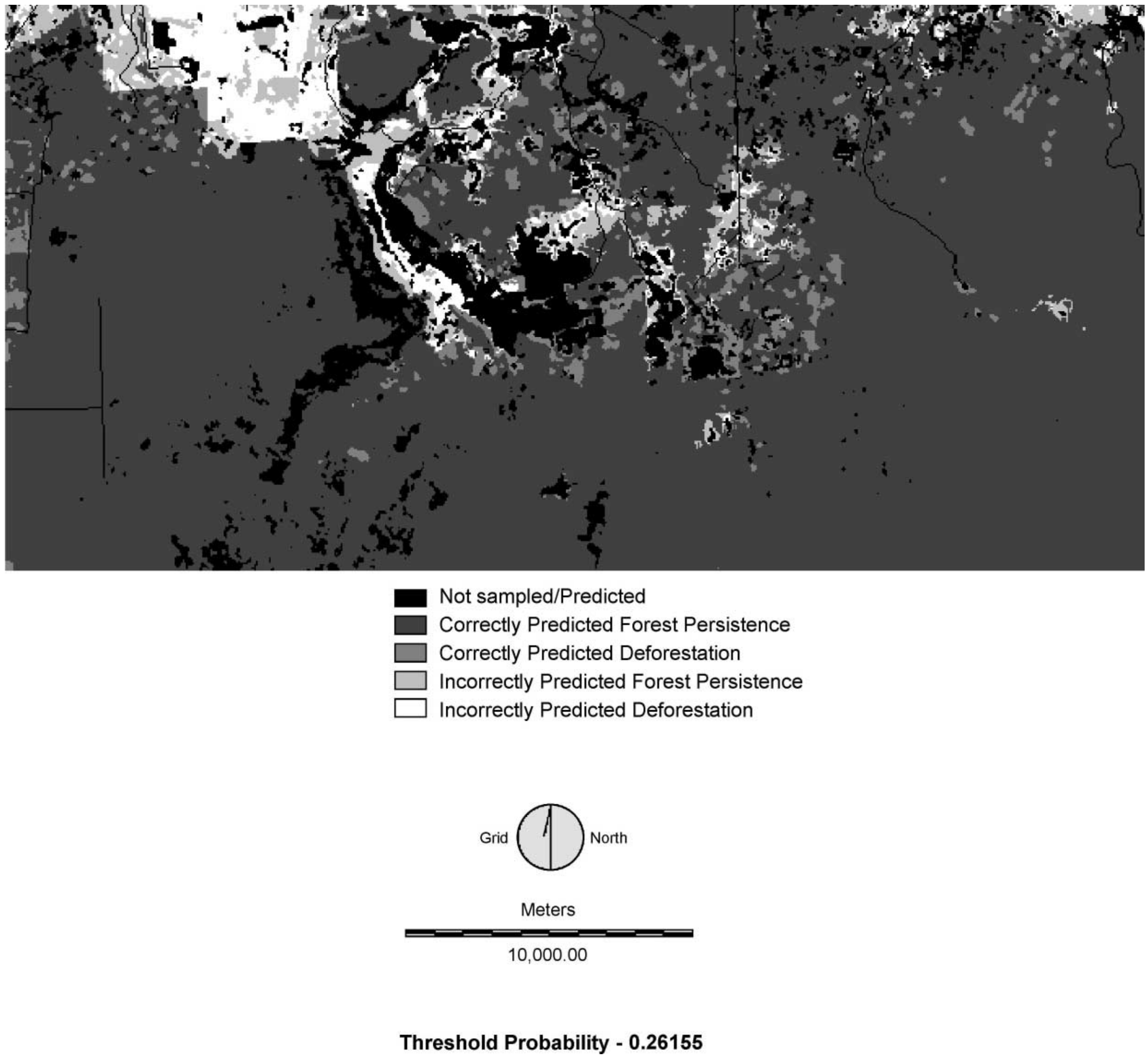



Fig. 5. Actual and predicated deforestation 1988–1992 (out of sample).

road distance to nearest major market; household (population); size of household land holdings; number of cattle owned; and truck access. The results from this model can be found in Table 4. While many of the signs on the estimated coefficients are as expected, the only statistically significant variables at the 95% level are elevation, soil type, and plot size. Similarly to the discrete choice models, the higher the elevation, the smaller the amount of deforestation; for soil type, *bajo* soils are negative and significant — unlike the wet rice projects of late 1970s to early 1980s, *ejidatarios* tend to avoid the seasonally inundated and massive clay *bajo* soils — and the larger the area available to an individual, the greater the amount of deforestation, implying extensification of agricultural use in a swidden-based system. None of the distance variables or socio-demographic variables are statistically significant. Overall, the model fits poorly, with an $R^2$ of 0.13.

The next model developed in this exploration includes other socio-demographic and economic

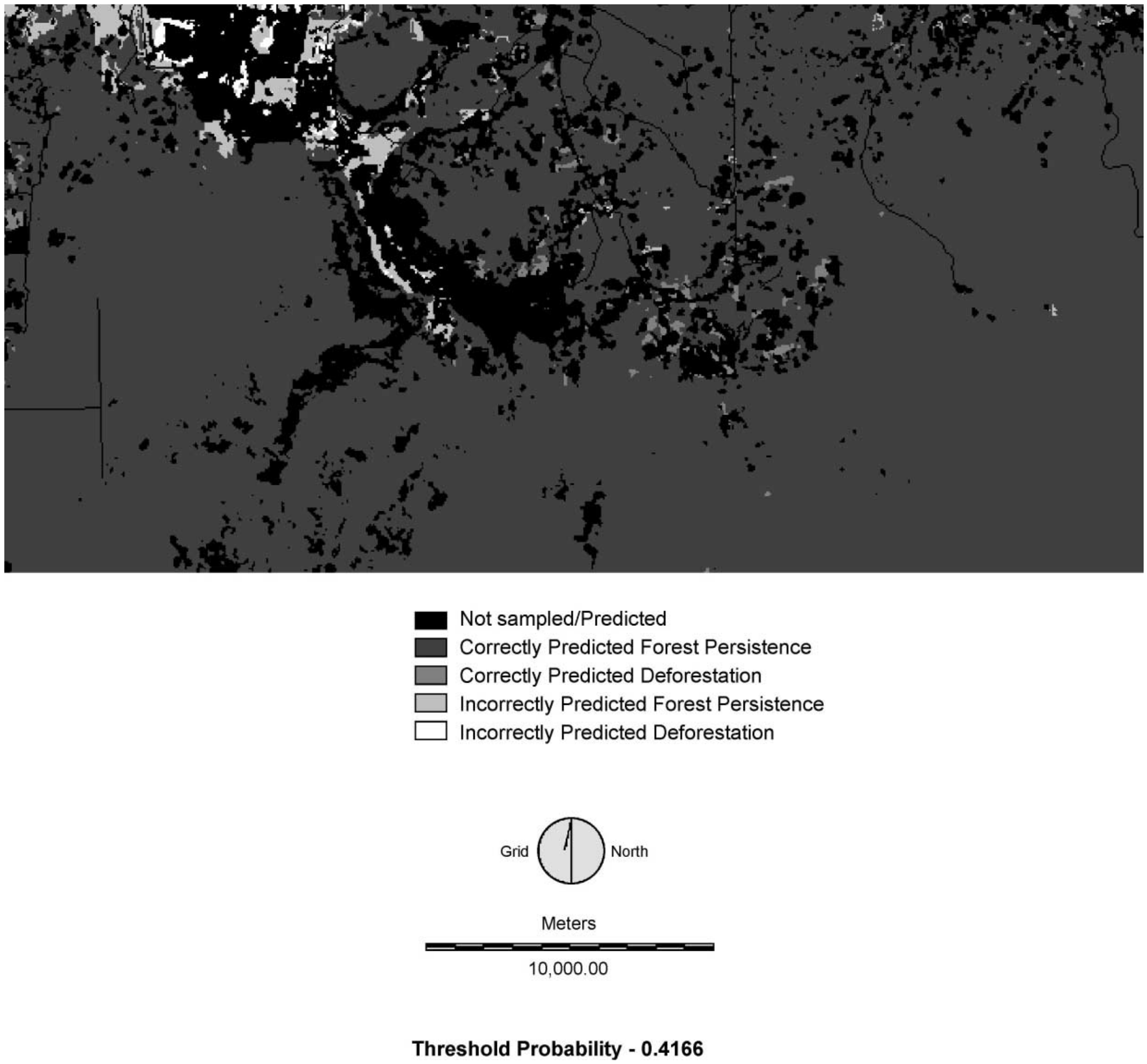



Fig. 6. Actual and predicated deforestation 1992–1995 (out of sample).

variables from the survey data, as it is hypothesized that these other variables better capture the true features of the underlying decision-making process. The distance measures and the geophysical measures are kept, but a number of other variables are replaced. Instead of household size, this model includes a variable that is a measure of individuals who are dependent upon the agricultural output to better capture the demands on subsistence production (induced intensification thesis), as there are extended family ties throughout the region. The number of cattle variable is replaced with the total value of livestock sold during the year, also as a more complete measure of total demand for agricultural resources. Also included are the education level of the head of household and the amount of income generated off-farm for the household, as it is expected that these features will affect managerial skill of the farmers and the non-farm opportunities available to the family. Lastly, a variable on chainsaw access is included, as this technology

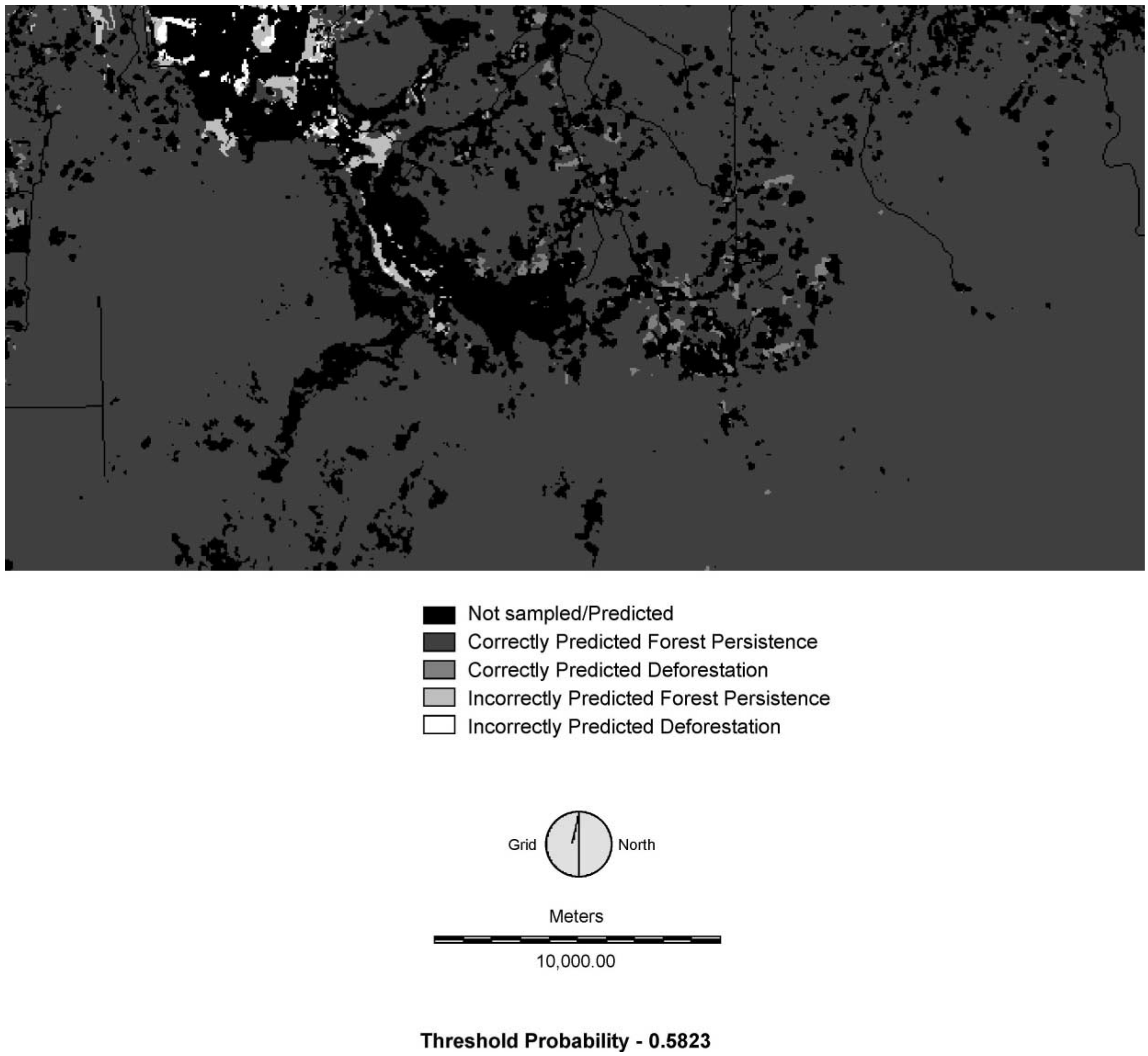



Fig. 7. Actual and predicated deforestation 1992–1995 (out of sample).

decreases the labor costs associated with clearing forests. The results from this model are found in Table 5.

The results from this model demonstrate the value of using the richer survey data. Now, the estimated coefficients on the variables that measure the overall demand on the output of the plot, the number of household members who are dependent on the crop output, and the total value of livestock sold are positive and significant, meeting a priori expectation. The estimated coefficients on truck and chainsaw availability are not statistically significant; the hypothesis was that these variables would increase the amount of land deforested because truck availability should make access to markets easier and chainsaws decrease the labor cost of land clearing. Individual restrictions on land access (*ejido* tenure institutions) and the use of fallow to regenerate soil quality may interfere with these expectations. The two other household variables, education of head of household and off-farm income, have

Table 3
Prediction results for binomial logit models

| Prediction type | In-sample (%) | Out-of-sample (1) (%) | Out-of-sample (2) (%) |
|---|---|---|---|
| *Time period 1 (1988–1992)* | | | |
| Correct forest persistence | 92 | 95 | 94 |
| Incorrect forest persistence | 8 | 5 | 6 |
| Correct deforestation | 42 | 49 | 53 |
| Incorrect deforestation | 58 | 51 | 47 |
| *Time period 2 (1992–1995)* | | | |
| Correct forest persistence | 98 | 99 | 99 |
| Incorrect forest persistence | 2 | 1 | 1 |
| Correct deforestation | 19 | 24 | 16 |
| Incorrect deforestation | 81 | 76 | 84 |

estimated coefficients that meet expectations; they are both negative and significant, demonstrating as other economic activities are available to the household, the less the demand on clearing the forest. The elevation and area of land available variables are both statistically significant and of the same sign as in the previous model, demonstrating the importance of these variables in both specifications. In this model, the soil type variable is not statistically significant at the 95% level as it was in the previous model, but is extremely close to the critical value (it is statistically significant at the 94.6% level). An interesting feature is the lack of statistical significance on any of the distance/accessibility measures. Perhaps this is not surprising, given the institutional framework of the *ejido* sector. *Ejido* land is not purchased in a market, but is assigned through a different process that is not observed in this model. In addition, it is not possible are to predict the specific parcels assigned to any individual *ejido* member.

Table 4
Regression model of deforestation, "similar" explanatory variables unit of observation: the household ($n = 145$)

| Variable name | Estimated coefficient | *t*-Statistic |
|---|---|---|
| Constant | 13.800 | 3.469 |
| Elevation | −0.030 | −2.335 |
| Slope | −0.519 | −0.612 |
| Soil type | −4.094 | −2.059 |
| Distance from plot to road | 0.059 | 0.416 |
| Distance to plot to market | −0.011 | −1.009 |
| Distance to house to plot | 0.088 | 0.250 |
| Household size | 0.324 | 1.099 |
| Number of cattle | −0.079 | −1.141 |
| Size of plot | 0.038 | 2.753 |
| Truck availability | 2.384 | 1.194 |
| $R^2$ | 0.13 | |

Table 5
Regression model of deforestation, "refined" and additional explanatory variables. Unit of observation: the household ($n = 145$)

| Variable name | Estimated coefficient | *t*-Statistic |
|---|---|---|
| Constant | 15.093 | 4.300 |
| Elevation | −0.029 | −2.649 |
| Slope | −0.731 | −0.943 |
| Soil type | −3.338 | −1.942 |
| Distance from plot to road | 0.054 | 0.426 |
| Distance to plot to market | −0.012 | −1.149 |
| Distance to house to plot | 0.146 | 0.464 |
| Number of plot dependents | 0.504 | 2.602 |
| Value of livestock sold | 0.0001 | 5.020 |
| Size of plot | 0.023 | 2.123 |
| Truck availability | −0.587 | −0.280 |
| Chainsaw availability | 1.679 | 1.108 |
| Education of head of household | −0.518 | −2.326 |
| Off-farm income | −0.0002 | −3.664 |
| $R^2$ | 0.33 | |

Given the small sample set from the survey data, the dataset was not split into a estimation set and a prediction set, as was done with the satellite data model. Therefore, only in-sample predictions were calculated. Fig. 8 (*ejido* boundaries demarcated) shows the predicted residuals from the second regression.[25] Visually, inspecting the predicted residuals

[25] The predicted residuals were calculated by subtracting the predicted dependent variable from the estimated model from the observed value.

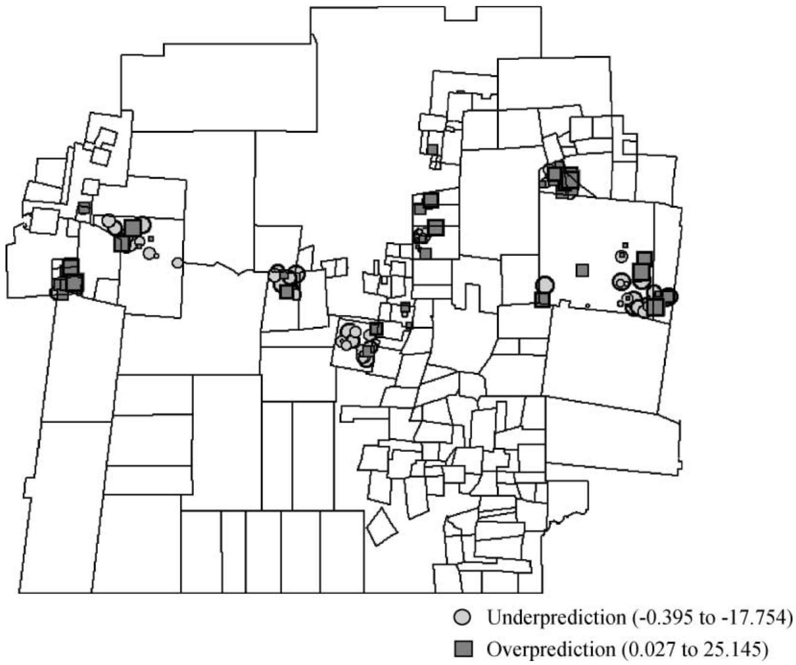



Fig. 8. Predicated residuals from enhanced survey regression model of deforestation.

gives insights into potential problems caused by spatial autocorrelation in the residuals. There are some *ejido*s where the mix of over prediction and under prediction are fairly evenly mixed, while in others there is a distinct amount of one over the other. Therefore, there may indeed be spatial autocorrelation in the error terms, possibly due to omitted variables and/or not modeling the spatial process correctly.

### 4.3. Comparing the models

To reiterate, these two sets of models ask different questions. The unit of observation for the discrete choice models is the pixel; the unit of observation for the continuous model is the land manager. The discrete choice model estimates the probability of deforestation of a pixel, given its location and biophysical characteristics as well as the aggregate socio-demographic characteristics of the *ejido* in which the pixel resides. In contrast, the continuous model estimates the amount of land a farmer will clear of forest and successional growth, given the location and biophysical characteristics of the plot, as well as information on family size and structure.

While the models ask different questions using different data, several broad comparisons seem useful. While most variables are statistically significant in the discrete choice model, none of the location variables are statistically significant in the continuous model. Therefore, while location affects the overall probability of deforestation, it does not explain the amount of deforestation on a given location by an individual. This result is perhaps explained by the institutional framework and semi-subsistence behavior of the farmers. As described above, *ejido*s are essentially geographically defined, federal land grants to groups, and members of an *ejido* are usually assigned usufruct rights to specified segments of the *ejido*. Further research on how and why a particular farmer is at a particular location would be needed to supplement the model results that explain given the parcel of land and the agricultural choices on it.

Comparing the two regression models using the survey data, none of the socio-demographic variables are statistically significant in the first model which included the explanatory variables that most closely "matched" the discrete choice model. However, in the second regression model, the "improved" variables of household dependents and amount of livestock sold, as well as the additional variables of education and off-farm income, are statistically significant, demonstrating both the importance of measuring the direct variables that are theoretically hypothesized to affect choice, instead of proxy measures, as well as controlling for other individual factors that can affect the choices available to an individual.

## 5. Conclusions and future research

The purpose of this paper was to demonstrate the methodology that is being developed in the SYPR project. Elsewhere, we have referred to this kind of effort as "socializing the pixel" and "pixelizing the social" (Geoghegan et al., 1998), a centerpiece of proposed work by the larger land-use/cover change community of researchers. The aim of this body of work is to improve understanding and explanation of the variation in the regional dynamics of land change, including improvements in spatially explicit models that move beyond the magnitude of changes within a region to the locations of that change.

There remains much data and modeling work to be done to achieve these aims. At this time the SYPR project is completing the classification of the full spatial extent for different years of the TM Landsat images covering the region as well as enhancing the project's GIS-layered data sets commensurate with the problems encountered in the region. For example, the project can enrich its analysis by use of *ejido* and Calakmul Biosphere Reserve boundaries, accounting for within-*ejido* settlement locations and *ejido* forest reserves, and variations in *ejido* usufruct rules, improved rainfall variation data, and awaited data on soil–vegetation nutrient cycling under different land uses. These and other data enhancements are underway or are being planned.

For the satellite data model, the unit of observation is the pixel, but clearly no land manager makes decisions on a pixel basis, so further research into developing a better unit of observation, perhaps, e.g., by simply determining the number of pixels in an average agricultural plot in the region, and using that land area as the unit of observation. Additional considerations include modeling other land changes, such as forest regrowth, to attempt to better capture the dynamic of crop–fallow cycles. In the current models "initial forest type" is included as an explanatory variable in an attempt to capture some of this dynamic. But the survey work demonstrates that crop–fallow cycles vary by environmental and demographic conditions, and by crop: *milpa* is often cut from secondary growth, but approximately one half of the observations on chili plots are planted on recently cut old growth forest. Further field research on this topic is currently underway. [26] As more characteristics of the forest are included in the model, [27] more detailed analysis of how these characteristics affect land use/cover change can be analyzed. Moreover, while each time period in this trial was modeled separately, the data are in a panel format, so future models will use statistical time-series techniques that link the pixels through time for more robust estimation.

Further models will also be developed using the survey data to better understand the land dynamics associated with individual land managers. As in the case of the satellite data model, survey data will be used to model jointly the decision to cut down primary and secondary forest and how this decision relates to fallow cycles. Further data development includes linking the individual land managers to the satellite data through the use of GPS when creating the sketch maps of their lands. By doing so, a spatially explicit land-use/cover history by farmer can be linked to the time series of satellite images. This way, it can be seen how the spatial pattern of land uses has changed over time, and the results used to reformulate these models.

Very little policy analysis has been included to date, although the region has been profoundly affected by changes in federal policies, e.g., promotion of forestry, *ejidos*, mechanized rice production, cattle, and, more recently tourism. Macroeconomic variables (that may or may not differ across the landscape), such as agricultural loans, changes in prices, and so forth, will be explored to capture how decisions made

[26] Eric Keys, a doctoral candidate in the Graduate School of Geography and project member is currently detailing the history and characteristics of chili cultivation in the SYPR.

[27] Harvard Forest leads the forest ecology work under the direction of David Foster. Diego Pérez-Salicrup is examining the structure of different forest types, and Deborah Lawrence (now of University of Virginia) is investigating nutrient cycling. In addition, Audrey Barker Plotkin is undertaking a photogrammetric reconstruction of the region.

at the macro-level affect individual land-use choices. Current field research into the market structure and individual choice of chili production will lead to a better understanding of that phenomenon and how to integrate this into a land-use change model.

Finally, the project intends to explore the issue of relative or proximate location. To date, success in spatially explicit land models has been determined by getting the precise location correct, in this case a $28\,m \times 28\,m$ pixel. Less attention has been paid to the value of determining the proximate location (e.g., within X number of pixels from the actual pixel of change). Some of this is clearly related to the unit of observation issue raised above, i.e., what scale of analysis, the pixel or some group of pixels is "best" and which variables remain exogenous or perhaps become endogenous at different scales of analysis. These facets of land-use/cover change models will also be explored in the future.

## Acknowledgements

This work was undertaken through the auspices of the Southern Yucatán Peninsular Region project with core sponsorship from NASA's LCLUC (Land-Cover and Land-Use Change) program (NAG 56406) and the Center for Integrated Studies on Global Change, Carnegie Mellon University (CIS-CMU; NSF-SBR 95-21914). Additional funding from NASA's New Investigator Program (NAG5-8559) also supported the specific research in this paper. SYPR is a collaborative project of El Colegio de la Frontera Sur (ECOSUR), Harvard Forest, Harvard University, the George Perkins Marsh Institute, Clark University and CIS-CMU. The views expressed in this paper are those of the authors and do not necessarily represent those of the US Environmental Protection Agency. We thank Laura Schneider for her assistance with this paper.

## References

Achard, F., Eva, H., Glinni, A., Mayaux, P., Richards, T., Stibig, H.J. (Eds.), 1998. Identification of Deforestation Hot Spot Areas in the Humid Tropics. Trees Publ. Series B. Research Report No. 4. Space Application Institute, Global Vegetation Monitoring Unit. Joint Research Centre, European Commission, Brussels.

Boserup, E., 1965. The Conditions of Agricultural Growth. Aldine, Chicago, IL.

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

Deaton, A., 1997. The Analysis of Household Surveys: A Microeconometric Approach to Development Policy. Johns Hopkins University Press, Baltimore, MD.

Ericson, J., Freudenberger, M.S., Boege, E., 1999. Population dynamics, migration, and the future of the calakmul biosphere reserve. Occasional Paper No. 1. Program on Population and Sustainable Development. American Association for the Advancement of Science, Washington, DC.

Ewell, P., Merrill-Sands, D., 1987. Milpa in Yucatán: a long-fallow maize system and its alternatives in the Maya peasant economy. In: Turner II, B.L., Brush, S.B. (Eds.), Comparative Farming Systems. Guilford Press, New York, pp. 95–129.

Foley, M.W., 1995. Privatizing the countryside: the Mexican peasant movement and neoliberal reform. Lat. Am. Pers. 22, 59–76.

Geoghegan, J., Wainger, L., Bockstael, N., 1997. Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. Ecol. Econ. 23, 251–264.

Geoghegan, J., Pritchard Jr., L., Ogneva-Himmelberger, Y., Roy Chowdhury, R., Sanderson, S., Turner II, B.L., 1998. Socializing the pixel and pixelizing the social in land-use and land-cover change. In: Liverman, D., Moran, E., Rindfuss, R., Stern, P. (Eds.), 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, pp. 51–69.

Goldring, L., 1995. State sponsored land certification and titling programs and changing state-producer relations: *Ejido* reform in Mexico. Common Property Resour. Digest 33, 2–5.

IGBP Report 35-IHDP Report 7, 1995. Land-use and Land-cover Change. Science/Research Plan, Stockholm and Geneva.

IGBP Report 48-IHDP Report 10, 1999. Land-use and Land-cover Change. Implementation Plan, Stockholm and Bonn.

Kaimowitz, D., Angelsen, A., 1998. Economic Models of Tropical Deforestation: A Review. CIFOR, Bogor, Indonesia.

Klepeis, P., 2000. Deforesting the once deforested: land transformation in southeastern México. Doctoral Dissertation. Graduate School of Geography, Clark University.

Klepeis, P., Vance, C., 2000. The impact of land reform and PROCAMPO on land transformation in southeastern México. Paper Presented at the Conference of Latin American Geographers, Austin, TX.

Kremen, C., Niles, J.O., Dalton, M.G., Daily, G.C., Ehrlich, P.R., Fay, J.P., Grewal, D., Guillery, R.P., 2000. Economic incentives for rain forest conservation across scales. Science 288, 1828–1832.

Liverman, D., Moran, E.F., Rinfuss, R.R., Stern, P.C. (Eds.), 1998. People and Pixels: Linking Remote Sensing and Social Science. National Academy Press, Washington, DC.

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

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

Pfaff, A., 1999. What drives deforestation in the Brazilian Amazon? J. Environ. Econ. Mgmt. 37, 26–43.

Primack, R.B., Bray, D., Galletti, H.A., Paciano, I. (Eds.), 1998. Timber, Tourists, and Temples: Conservation and Development in the Maya Forest of Belize, Guatemala, and Mexico. Island Press, Washington, DC.

Stephen, L., 1998. The cultural and political dynamics of agrarian reform in Oaxaca and Chiapas. In: Snyder, R., Torres, G. (Eds.), The Future Role of the Ejido in Rural Mexico. Center for US–Mexican Studies, La Jolla, CA, pp. 7–30.

Turner II, B.L., Ali, A.M.S., 1996. Induced intensification: agricultural change in Bangladesh with implications for Malthus and Boserup. Proc. Natl. Acad. Sci. USA 93, 14984–14991.

Turner II B.L., Brush, S.B., 1987. Comparative Farming Systems. Guilford Press, New York.

Vance, C., 2000. Land-use change, deforestation, and peasant farm systems: a case study of México's southern Yucatán peninsular region. Doctoral Dissertation. Department of Economics, Clark University.

Vance, C., Geoghegan, J., 2000. Testing the separability hypothesis: an analysis of semi-subsistent and commercial land uses in an agricultural frontier of southern Mexico. Paper Presented in an Association of Environmental and Resource Economists Session at the Allied Social Science Association Meeting, Boston, MA.

Warwick, D.P., Lininger, C.A., 1975. The Sample Survey: Theory and Practice. McGraw-Hill, New York.

Whitmore, T., Turner II, B.L., Johnson, D., Kates, R.W., Gottschang, T., 1990. Long-term population change. In: Turner II, B.L., Clark, W.C., Kates, R.W., Richards, J.F., Mathews, J.T., Meyer, W.B. (Eds.), The Earth as Transformed by Human Action. Cambridge University Press, Cambridge, pp. 25–39.