



---

---

The Review  
of Regional Studies

---

---

*Oklahoma State University*

## **Examining the Reliability of Survey Data with Remote Sensing and Geographic Information Systems to Improve Deforestation Modeling**

**Jill L. Caviglia-Harris**

*Department of Economics and Finance, Salisbury University, Salisbury, Maryland  
21801-6860, e-mail: jlcaviglia-harris@salisbury.edu*

**Daniel W. Harris**

*Department of Geography and Geosciences, Salisbury University, Salisbury, Maryland  
21801-6860, e-mail: dwharris@salisbury.edu*

---

### **Abstract**

Tropical deforestation has environmental consequences at local, regional and global scales. The Brazilian Amazon's deforestation resulted largely from conversion to farmland by landholders. These conversions caused deforestation that researchers have examined locally by interviews with landowners and regionally by satellite remote sensing. This paper merges data from these methods to validate survey-based deforestation levels with remote-sensing information. We determine household characteristics associated with misreporting of land use. After identifying errors, we modify the data to better estimate influences on local deforestation. Although individuals are not found to intentionally misrepresent land use, incorporating differences between the two data sources improves the estimations of deforestation.

*Keywords: Land use change; Deforestation; Amazon; Satellite remote sensing; GIS*

*JEL classification: Q12; Q23; R14*

---

We thank Dar Roberts, University of California, Santa Barbara, for the classified 1996 land cover images. National Security Education Program, Organization of American States, Institute for the Study of World Politics, and McClure Fund Foundation grants helped support the administration of the household surveys. The data can be found at the archive of social science data –Inter-university Consortium for Political and Social Research, University of Michigan. All location identifiers have been removed.

## 1. INTRODUCTION

Tropical deforestation remains a critical concern due to its role in global climate change and biodiversity loss. Because deforestation results from complex social and physical processes, researchers from a variety of disciplines collect and analyze a substantial array of socioeconomic, demographic, and satellite-derived land use data. Researchers rely on accurate measurements to determine the causes and extent of forest loss and to understand the consequences of land use change. Most deforestation measurements are collected at two very different scales: 1) satellite remote sensing measurements of entire regions, and 2) on-site interviews of households (Brondizio et al. 1996; Lambin 1997; Lambin, Rounsevell, and Geist 2002; Kaimowitz and Angelsen 1998; Walker, Moran, and Anselin 2000; Evans et al. 2001; Caviglia-Harris 2004). Although analyzing land use change with data collected from different scales has advantages, a more robust understanding of deforestation dynamics and interactions can be gained by merging remotely sensed data, processed with geographic information systems (GIS), with survey observations. The union of these data enables socioeconomic processes associated with individual landholders and regional markets to be examined in relation to their impact on the broader framework of large-scale deforestation.

Emerging research in the social sciences examines the utility of linking of satellite and survey data (Mendelsohn 1994; McCracken et al. 1999, 2002; Vance and Geoghegan 2002; Muller and Zeller 2002; Munroe, Southworth, and Tucker 2002; Staal et al. 2002). In the context of tropical deforestation, many recent studies have combined survey and satellite data (Chomitz and Gray 1996; Pfaff 1999; Geoghegan et al. 2001; Rudel, Bates, and Machinguiashi 2002; Perz and Skole 2003; Chomitz and Thomas 2003). Inconsistencies in the units of analysis between satellite and survey data, however, can lead to potential interpretation errors (Rindfuss et al. 2003a; Brondizio et al. 1996, 2002). The ability to evaluate and verify deforestation and land use data collected by survey is an important issue that has been overlooked (Rindfuss et al. 2003b). Error exists in any data sample but can be minimized by careful survey design and the implementation of statistical techniques to test for and correct bias. In addition, the comparison of an independent source of data can provide a powerful method for evaluating survey responses. In this paper, we implement an approach to statistically validate survey-derived estimates of deforestation with remotely sensed land use data incorporated into geographic information systems (GIS).

While household-level observations are also used to assess data collected through remote sensing, our goal is to evaluate possible falsified or erroneous responses in household reported values of deforestation. Relatively large landholdings in Ouro Preto do Oeste, Rondônia, provide the opportunity to merge satellite and survey data using the household as the common unit of analysis. Landholdings in this study area average 71 hectares (a minimum of 10 and a maximum of 325), and therefore are large enough to analyze with Thematic Mapper (TM)-derived land cover data with a spatial resolution of 30 meters.

We begin our study with a review of current research examining data measurement error. Next, we examine the geography and settlement history of the survey area. The methodology section identifies data sources, discusses data processing, and then describes the application of the reliability index to our data. Based on our results, we then estimate response error, correct for bias, and develop a methodology to improve survey responses.

## 2. MEASUREMENT AND RESPONSE ERROR

Data errors in the form of inaccurate responses by subjects, erroneous data entry, or miscoding are common in the various stages of data collection. These issues generally fall under “classical measurement error” and are often ignored in statistical analyses if the inaccuracies are unrelated to the error term or uncorrelated with the true value of the variable (Aiger 1973; Hyslop and Imbens 2001; Bollinger 2001). Alternatively, if autocorrelation or collinearity results from misreporting, errors cannot be treated as random in statistical analysis because the results can over- or under-estimate the magnitude of the relationship between the true variables and the regressors (Messonnier et al. 2000; Bollinger and David 1997). Such survey response bias can lead to erroneous empirical results and misguided policy suggestions.

Studies on the estimation of survey response bias have focused on the verification of interview data, in particular on responses to questions on sensitive topics and income (Bound and Kruger 1991; Bollinger and David 1997; Bollinger 1998, 2001). For example, it is often assumed that respondents have an incentive to underreport illegal drug use, criminal history, and embarrassing medical conditions. Validity studies estimate this bias using external criterion information such as official records or drug test results (Marquis et al. 1986). By applying the reliability index developed from these studies, we test the validity of our survey responses on deforestation with measurements obtained from satellite images. Satellite data are used as the criterion data (or comparison) in the calculation of the reliability index since they are collected from an objective source and because we believe that households may not accurately report deforestation since it is illegal to clear more than 30 percent of the landholding.

The model of measurement error used in this analysis is based on Marquis, Marquis, and Polich (1986) and Bound and Kruger (1991). Reported values of deforestation for household  $i$ ,  $D_i$  are assumed to equal the true value,  $T_i$  plus an error,  $e_i$ :

$$(1) \quad D_i = T_i + e_i.$$

In the case of classical measurement error, it is assumed that  $e_i$  is the random response error, uncorrelated to the true and criterion values of the variables, and has an expected value of zero and variance of  $\sigma_e^2(i)$ .

Our objective is to determine whether any measurement error is of the classical form or correlated with the error term, and whether the estimates of the dependent variable are

biased. We do this by estimating the measurement error with regression analysis and identify any observable correlates. Unlike previous studies on measurement error and response bias in which non-response and missing data often result in errors, data measurement error exists for all respondents in our sample due to rounding, differences in reported units of measurement, and the minimum spatial resolution of the satellite sensor. Land use derived from TM reports area in square meters, while respondents answered in hectares, a much larger and therefore less precise measurement. Additionally, deforestation derived from satellite imagery is limited by the platform's (TM) spatial resolution, which is 30 meters. Thus, pasture, secondary forest, and primary forest fragments less than 900 square meters are not differentiated and are represented by a single land use value. For example, a household that reported 40 hectares (400,000 m<sup>2</sup>) of deforestation might have 399,600 m<sup>2</sup> of cleared land according to the classified satellite image. Therefore, due to the use of different units of measurement and the constraint of the sensor's spatial resolution, none of the observations were found to contain exact matches.

The criterion values of land cover for household  $i$ ,  $C_i$  are assumed to equal the true value,  $T_i$  plus an error,  $v_i$ . Although our satellite-derived values are used as the criterion data, they may also contain error caused by misclassification, data entry errors, coverage misalignment, and map inaccuracies caused by survey changes.<sup>1</sup> Again, these errors are expected to have a value of zero and variance of  $\sigma_v^2(i)$ :

$$(2) \quad C_i = T_i + v_i.$$

The most frequently implemented test of data quality is the reliability index,  $R$ . In this context, the index is the ratio of the variance of the true values of deforestation and the variance of the values reported in the survey. When no error exists, this value is equal to one. Since the true values are unknown,  $\hat{R}$  is estimated as the ratio of the covariance of the criterion and survey data and the variance of the values reported in the survey.  $\hat{R}$  is derived from equations 1 and 2 assuming that  $v_i$  and  $e_i$  are uncorrelated with each other and  $T_i$ . Under this assumption  $\sigma_{DC} = \sigma_T^2$  (the variance of the true values) and

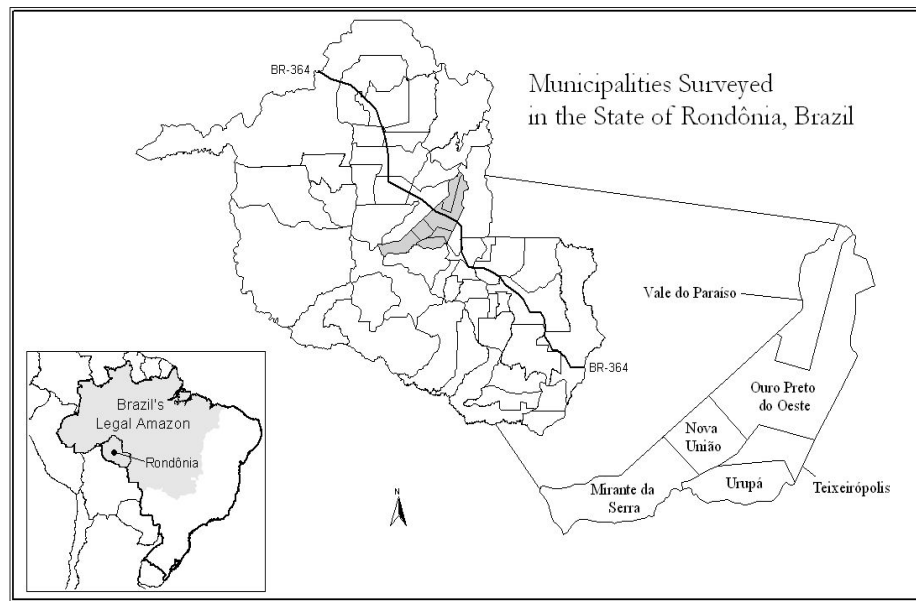
$$(3) \quad \hat{R} = \frac{\sigma_{DC}}{\sigma_D^2}$$

### 3. STUDY AREA AND SURVEY ADMINISTRATION

Rondônia is located in the southwestern Amazon basin of Brazil (Figure 1). The state includes the northwestern portion of the Serra dos Parecis and is characterized by elevations between 50 and 1,000 meters, although most elevations are below 300 m.

---

<sup>1</sup> This is discussed in greater detail in the methods section of the paper.



**FIGURE 1: Map of the Survey Site in Ouro Preto do Oeste, Rondônia, Brazil**

Climate is classified as humid equatorial with a dry winter season. Average annual temperature is above 25°C, and annual precipitation totals over 2,200 mm. (Peterson 2003).

The Ouro Preto do Oeste region of Rondônia is located along the Cuiabá-Porto Velho highway (BR-364) that traverses north-south through the state (Figure 1). Six municipalities comprise the study region: Ouro Preto do Oeste, Vale do Paraíso, Urupá, Mirante da Serra, Nova União, and Teixeiraópolis. These municipalities have a combined population of greater than 92,000, of which 58 percent is rural. Most urban population (68 percent) is concentrated in the municipality of Ouro Preto do Oeste. Major roads in the municipalities are paved, although secondary roads leading to most households remain unpaved and difficult to navigate during the rainy season. Year-round bus service is available along BR-364 and many major side roads.

Large-scale migration to Ouro Preto do Oeste began in the early 1970s as part of the government colonization program Operation Amazonia. Considered a model colonization project, the original land subdivisions included lots for 500 families. However, by 1974, 4,000 families had settled in the city and surrounding region. The lots, averaging 100 hectares, were planned along a square grid (creating secondary roads along boundaries) with little consideration for topography, hydrology, soil type, or additional environmental constraints. During the 1980s, the World Bank-sponsored project POLONO-ROESTE paved BR-364, providing greater access to the region attracting further immigration. During settlement, colonists were able to establish informal property rights by

burning or “improving” up to one-half of their landholdings. Laws now limit “improvements” to a maximum of 30 percent of lots, although many households continue to clear more than the allotted percentages for agriculture and pasture (Caviglia 1999; Alves et al. 1999; Fearnside 2001; Caviglia-Harris 2004).

#### **4. METHODOLOGY: DATA AND GIS PROCESSING**

The data used in the analyses include a stratified random sample of households surveyed in 1996,<sup>2</sup> classified satellite land use data for these households in the same year, and a geographic information system polygon coverage of the individual landholdings in the study area. The survey data include a sample of households interviewed over a five-month period (September 1996 through January 1997), stratified by the rural population of each of the six municipalities and uniformly distributed over space in the municipalities (Figure 2). The distance between individual lots was designed to represent the municipalities’ rural population equally and to obtain spatial variation in topography, soil type, and distance to market. Survey questions were designed to examine household decisions related to land use and collect information on household and lot characteristics, agricultural production, household income, and the use of agroforestry techniques.<sup>3</sup>

Land use data were obtained from classified Thematic Mapper (TM) satellite imagery and verified based on in situ ground truth and digital airborne videography.<sup>4</sup> Land use was categorized into eight classes: primary forest, second growth forest, pasture, green pasture, clouds, urban/soil, burn (pasture), and rock/savanna (Figure 3) (see Roberts et al. 2002). Land cover accuracy exceeded 85 percent in all classes, and therefore we were confident in imagery’s ability to distinguish between primary forest cover (primary forest) and the non-forest classes (second growth forest, pasture, green pasture, clouds, urban/soil, burn (pasture), and rock/savanna), the fundamental difference examined in our research question.

The GIS polygon coverage of landholdings in the region was digitized from a paper map produced during the settlement period by the Brazilian National Institute of Colonization and Agrarian Reform (Instituto Nacional de Colonização e Reforma Agrária – INCRA). GIS processing began with the identification of surveyed lots within the polygon coverage and the assignment of a unique identifier to the interviewed lots in the attribute table (Figure 2). Once the surveyed lots were identified, area calculations derived from the coverage were compared to responses collected as part of the interview

---

<sup>2</sup> The 1996 data collection was supported by grants from the National Security Education Program, the Organization of American States, the Institute for the Study of World Politics, and the McClure Fund Foundation.

<sup>3</sup> See Caviglia (1999), Caviglia and Kahn (2001), and Caviglia-Harris (2004) for further details.

<sup>4</sup> The 1996 Rondônia classified mosaic of land cover was obtained from Dar Roberts, Department of Geography at the University of California, Santa Barbara, and principle investigator of “Hyperspectral Analysis of Landcover in Rondônia,” part of the Large-Scale Biosphere-Atmosphere Experiment in Amazonia (LBA).

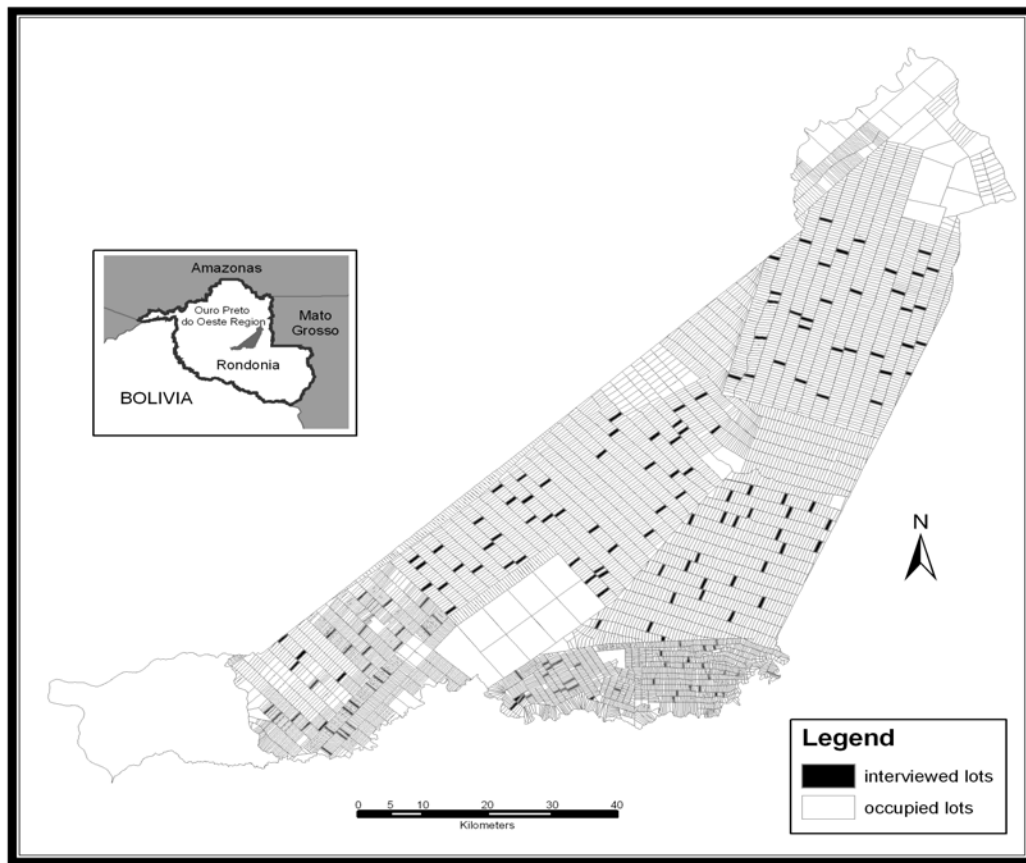


FIGURE 2: Landholdings in Ouro Preto do Oeste Region:  
Interviewed Households Highlighted

process. Of the 171 surveyed lots, 39 varied by more than 25 percent (or 10 hectares, whichever was smaller) and were eliminated from the sample, resulting in a sub-sample of 132 observations.

Several issues identified in the comparative process explain lot size differences and errors. One type of error is attributed to the subdivision and/or aggregation of individual farm lots that occurs when land is sold or bequeathed to family members. The initial lot size of 100, 50, or 25 hectares changes to a percentage of the initial allotment and does not correspond to the INCRA colonization map. Error also stems from the reality that the “usable” lot area does not always conform to the surveyed boundaries. Rocky outcrops and wetland areas cannot be cultivated and are often omitted when summing reported land area during the interview process. In addition, landholdings of 100 hectares (reduced to 50 and 25 hectares in the 1980s) were first distributed to households by

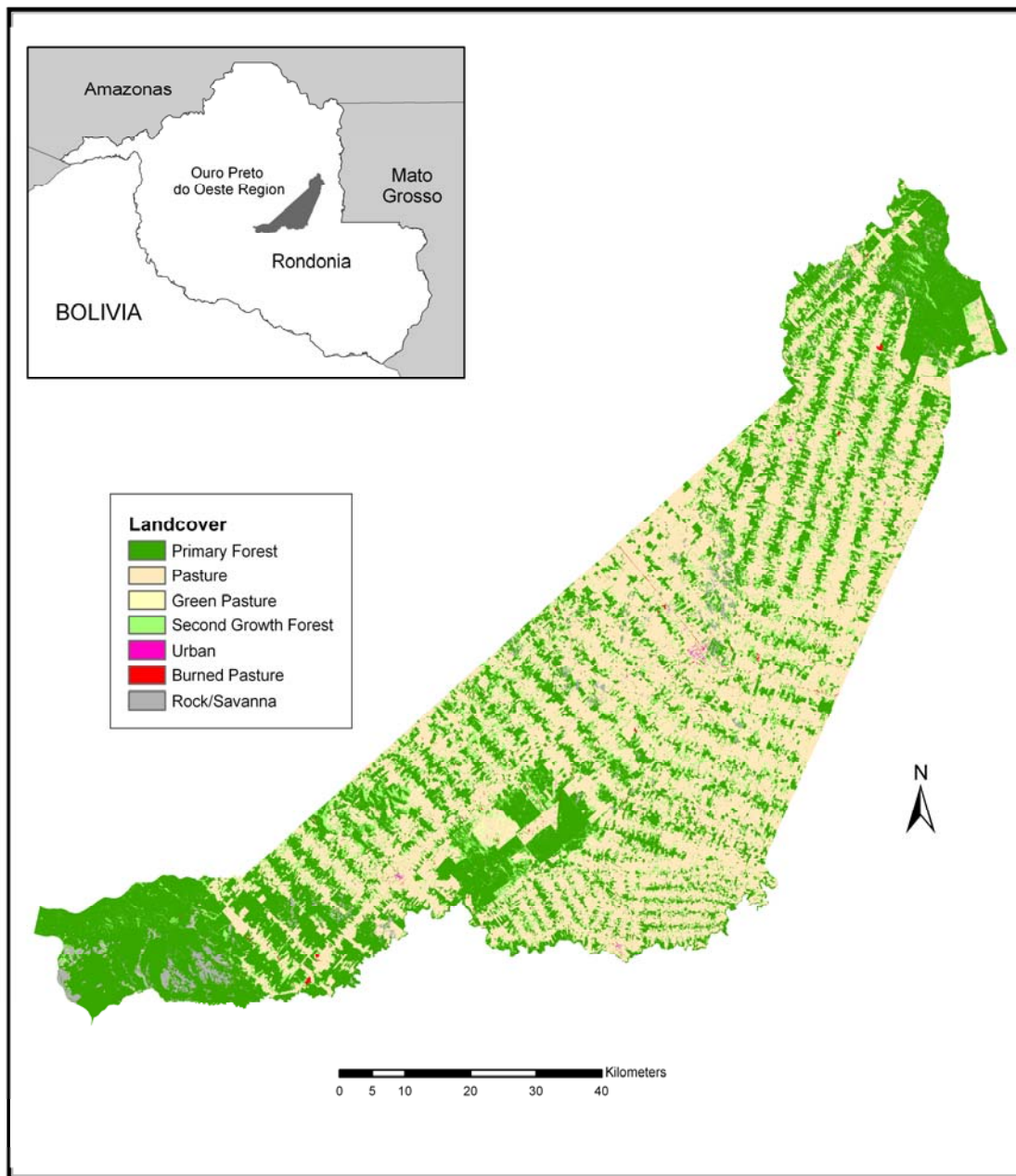


FIGURE 3: Landcover in the Ouro Preto do Oeste Region



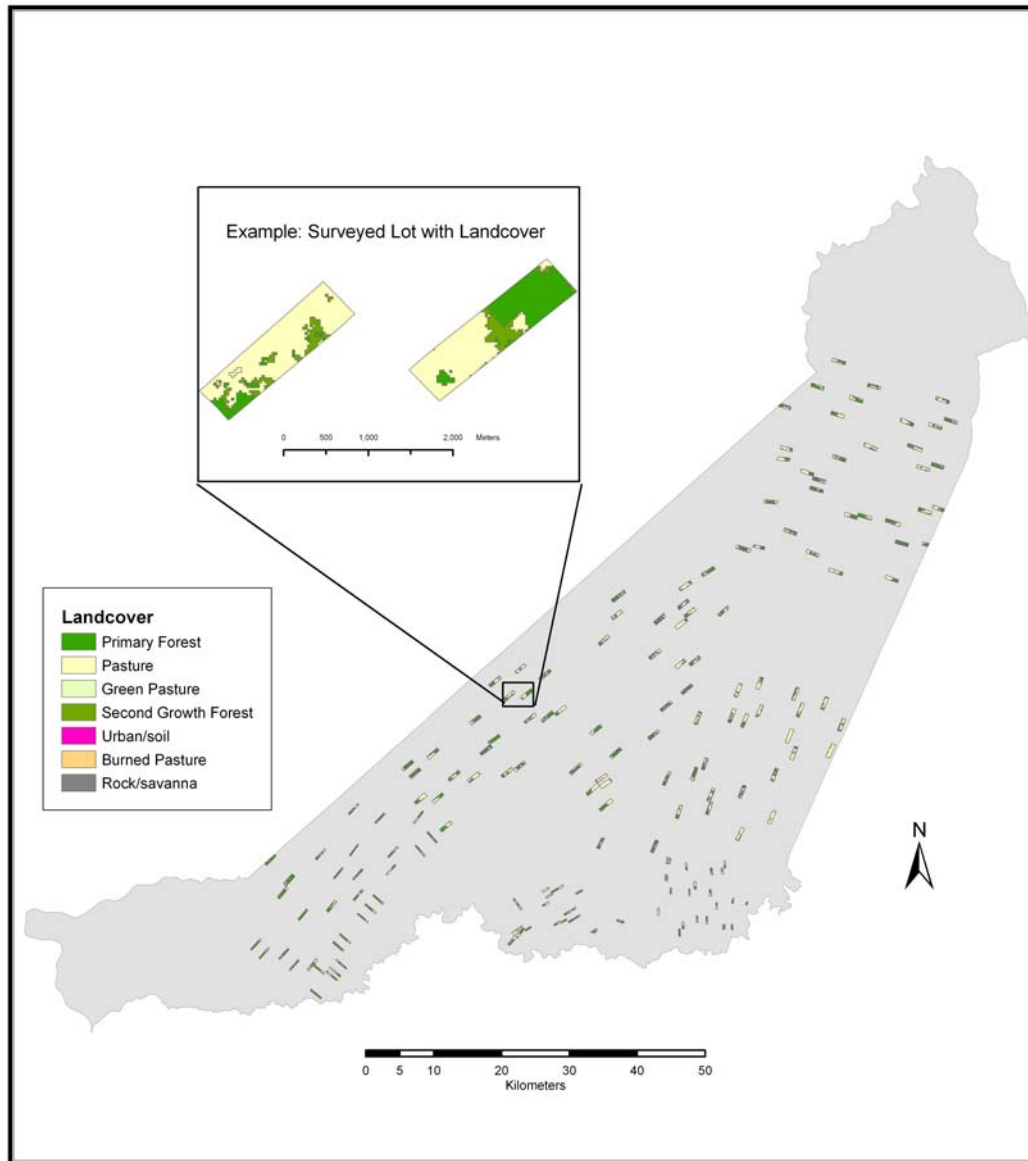


FIGURE 4: Landcover in Interviewed Lots: Ouro Preto do Oeste Region

INCRA according to the grid-like colonization map that ignored the true nonlinear positions of roads, hydrographic features, and topographical irregularities related to rock outcrops. As a result, actual lot sizes as reported by the surveyed inhabitants may differ from sizes derived from the INCRA map. Spatial data constraints also introduce errors into lot size calculations. The GIS coverage was digitized from a paper map with a very small scale (1:100,000). The inherent instability of the paper medium (e.g., expansion and contraction due to humidity changes and use), coupled with small lot sizes in the municipality Urupá, introduces significant distortions in certain lot size measurements.

Once the sub-sample of 132 lots was extracted, survey reported forest and non-forest areas were appended to the associated lots in the GIS coverage's database. Next, the classified land use mosaic was converted from a raster image to a GIS polygon coverage and clipped by the boundaries of the interviewed lots (Figure 4). The resulting land cover classes were aggregated into two categories: forest and non-forest.

Primary forest is defined in the survey data as any area of forest that has not been cleared since occupation (and does not include secondary growth forest) but may include intercropping that requires a forest canopy. Primary forest in the satellite-derived data corresponds to dense tropical forest (Roberts et al. 2002). In our research context, deforestation is defined as primary forest that has been cleared for pasture and agriculture. In the survey data, this includes pasture, crops, land in fallow, and homes, while satellite-derived land cover classes include pasture, green pasture, burned pasture, second growth forest, and urban land.

Table 1 provides a brief overview of the demographic and physical characteristics of the sub-sample data used in the verification process. In the sample, the average age of the head of household is 48 years, and this individual has about 2.5 years of education. Approximately nine people live on each lot, and the residents have lived on the lots for 11 years. Each lot is almost 70 hectares in size, of which nearly 50 hectares is deforested, with the remaining area in forest cover. These households earn R\$7,000 (US\$3,000) per year, possess over 60 cattle, and on average own less than one vehicle.

## **5. EMPIRICAL ANALYSIS**

The primary question of survey response accuracy is analyzed using the sub-sample of 132 landholdings. We investigate the quality of our survey responses using the reliability index to examine the differences between reported deforestation and forest cover. These classes were examined individually because these land uses are not mutually exclusive in the satellite data. The satellite data lot size includes a more detailed classification, including eight categories that were aggregated into four classes, deforestation, forest cover, rock, and savanna. Alternatively, the sum of the deforestation and forest cover in the survey data did sum to total lot size. Due to this difference and differences in the lot size between the samples, the reliability of these two land classifications were calculated individually.

TABLE 1

## Variable Definitions and Descriptive Statistics for the Sub Sample

Variable	Definition	Mean ( <i>n</i> = 132)	Standard Deviation
Education	education of the household head, in years	2.48	2.91
Age	age of the household head, in years	47.64	12.96
Family	number of family members residing on the lot	8.51	6.33
Years on Lot	number of years the household has resided on the lot	11.30	6.64
Distance	distance from the city center (or highway BR-364), kilometers	48.44	23.67
Total Income	total yearly household income, 2000 reais <sup>5</sup>	6957.01	6665.23
Vehicles Owned	number of motorcycles, cars, and trucks owned by the household	0.40	0.63
Cattle	number of cattle owned by the household	63.46	59.57
Reported Lot Size	lot size reported by the household, in hectares	69.31	32.45
GIS Lot Size	lot size, calculated in GIS from satellite images, in hectares	68.45	31.01
Reported Deforestation	deforestation reported by the household, in hectares	52.42	28.86
GIS Deforestation	deforestation, calculated in GIS from satellite images, in hectares	49.64	27.74
Reported Forest Cover	forest cover reported by the household, in hectares	16.89	15.30
GIS Forest Cover	forest cover, calculated in GIS from satellite images, in hectares	18.81	14.29

The reliability of the deforestation responses is estimated to be 83.3 percent, well above the 70 percent cutoff suggested to identify problem variables (Marquis, Marquis, and Polich 1986; Bound and Krueger 1991). Alternatively, the reliability of reported forest cover is only 52.5 percent, suggesting that sufficient error exists to warrant suspicion about the quality of these data.<sup>6</sup>

<sup>5</sup> Inflation rates are based on the annual average consumer price index (*World Economic Outlook May 1999* and *May 2000*, IMFund, Washington D.C.). In 2000 US\$1 was approximately equal to R\$1.83 according to the National Trade Data Bank, U.S. Department of Commerce.

<sup>6</sup> The reliability index relies on the assumption that the errors between the two estimations of deforestation are uncorrelated. Since these deforestation measurements are obtained independently with expected errors attributed to different causes, we do not believe this assumption to be violated.

Since the reliability indices suggest that data improvements are possible, we begin by estimating deforestation and primary forest, based on variables defined in Table 1.<sup>7</sup> Differences between the two estimations are the dependent variables, measured by the survey response and satellite, and the independent variable size of the lot, again measured by the two means. We find the estimates using satellite data to be slightly more efficient and to contain fewer significant variables (Table 2).<sup>8</sup> We also use a Wald test and find the coefficients of the regressions of deforestation and forest cover to be significantly different for the two data sets.

An important implication of these models is that the policy recommendations based on results would differ significantly depending on the dependent variable chosen. For example, in the estimation of deforestation using the household reported level of deforestation, we find that education is a significant determinate of deforestation. Those households with higher levels of education have deforested significantly less of their lots. This implies that programs that target improvements to the education system may also result in reducing deforestation levels. However, this is not the case if one examines the estimations based on the satellite data. Additional independent variables are found to significantly impact forest cover and deforestation. For example, policies to reduce the incentive to own cattle and produce beef and dairy products are likely to aid in reducing deforestation according to both models. In addition, income and wealth (as measured by the ownership of vehicles) are both positively related to deforestation according to the satellite data, implying that improvements in household welfare may also result in increased levels of deforestation.

Since the estimations of deforestation using the survey and satellite data are significantly different, we next analyze the determinants of the error. We estimate the net measurement error, defined as the absolute value of the difference between the survey and satellite values for deforestation, forest, and lot size using the same variables as in the deforestation models (Table 2). Specifically, we test for systemic differences in household characteristics that lead to differences in reported data. The mean error is 11 hectares for deforestation levels, 10 for forest, and 5 for the total lot size. The strongest indicator of error in each regression is lot size. In other words, as an individual's lot area increases, the accuracy of their reported data decreases. We do not find survey error to be significantly influenced by any variables indicating purposeful manipulation (e.g., education) but rather find variables demonstrating the owner's lack of knowledge about

---

<sup>7</sup> The variables included in the deforestation model are determined by the estimation of the derived demand for land for these households that serve as both consumers and producers of goods. See Caviglia-Harris and Sills (2005) for the conceptual framework.

<sup>8</sup> Lot size is an exogenous variable in this context since settlers were originally assigned lots during the occupation period. Some lots have changed hands through generations, and several were sold. However, a majority of the occupants in the 1996 survey were original settlers. Since 1996 a more developed land market has been noted, although still in the initial stages (Caviglia-Harris 2004).

TABLE 2

## Estimations of Deforestation and Measurement Error

	Deforestation (in hectares) Survey Data	Deforestation (in hectares) Satellite Data	Measurement Error Deforestation (in hectares)	Measurement Error Primary Forest (in hectares)	Measurement Error Lot Size (in hectares)
Constant	18.33*** (7.19)	20.58** (6.93)	-8.58 (5.79)	-5.52 (5.36)	3.20 (2.35)
Education	-0.93** (0.41)	-0.61 (0.38)	0.24 (0.33)	0.38 (0.31)	-0.01 (0.13)
Age	-0.04 (0.09)	-0.03 (0.08)	0.12* (0.07)	0.08 (0.07)	-0.02 (0.03)
Family	0.17 (0.19)	-0.22 (0.18)	-0.02 (0.15)	-0.11 (0.14)	0.01 (0.06)
Years on Lot	-0.67*** (0.19)	-0.12 (0.18)	-0.16 (0.16)	-0.14 (0.14)	-0.05 (0.06)
Distance	-0.15** (0.06)	-0.23*** (0.06)	0.07 (0.05)	0.04 (0.05)	-0.02 (0.02)
Total Income	-0.00 (0.00)	-0.00* (0.00)	0.00 (0.00)	0.00* (0.00)	0.00 (0.00)
Vehicles Owned	2.45 (1.94)	3.44* (1.82)	-0.12 (1.57)	-0.01 (1.45)	0.06 (0.64)
Cattle	0.09*** (0.03)	0.07*** (0.02)	-0.03 (0.02)	-0.05** (0.02)	-0.02* (0.00)
Reported Lot Size	0.65*** (0.05)		0.17*** (0.04)	0.18** (0.03)	0.07*** (0.02)
GIS Lot Size		0.63*** (0.05)			
R-Squared	.84	.85	.21	.24	.22
Adj R-Squared	.83	.84	.15	.19	.16
Observations	132	132	132	132	132

\*, \*\*, \*\*\* indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively; standard errors in parenthesis.

the lot to be key determinants. We expected households to overreport primary forest due to recent legislation making it illegal to clear and/or burn more than 30 percent of the lot. However, we found most households underreported forest levels and overreported deforestation, contradictory to what Walker, Moran, and Anselin (2000) presume.

Since we identify at least one variable that is strongly and consistently related to error, we use the results of the estimations of the errors to derive weights to improve the survey responses about deforestation and forest cover. Survey weights are commonly used to correct for unbalanced data or to undo disproportional endogenous sampling (Fitzgerald, Gottschalk, and Moffit 1998; Nevo 2003).<sup>9</sup> They are generally calculated to ensure that some characteristic of the sub-population is similar, on average, to the full population. For example, if the sub-population includes a cohort that consists of 30 percent of the population but only 10 percent of these individuals are represented in the subsample, a weight can be calculated to increase the representation of these individuals proportionally. We calculate a weight in a similar manner, by accounting for the increased error attributed to lot size. We use the regression coefficients found in Table 2 to estimate errors for the reporting of deforestation and forest cover by each household and apply these values in the weighted least squares estimation of deforestation and forest cover for our entire sample of 171 observations (Table 3).<sup>10</sup>

Using defined weights, we re-estimate non-forest and forest using our full sample and find the explanatory power of the regressions to increase (although efficiency of the estimators slightly decreases) (Table 3). The estimation results suggest the error in reported land use leads to an overestimation of the importance of lot size in non-forest land and an underestimation of the importance of household and lot characteristics such as age and income, distance to the city center, and wealth (in the form of vehicles and cattle). We also find that the influence of lot size, age, education, and distance to city center on forested land are underestimated with the uncorrected data. We recalculate the reliability index using the corrected data and find that the index increases to 91.7 for the non-forest data and to 77.2 for the forest data.<sup>11</sup> Note, the corrected survey data indicates that household characteristics, particularly income, wealth, and education, are more important than if they were predicted with survey data or satellite data alone. These results are particularly important to policy makers because they suggest different avenues to reduce pressure on tropical forests. Since income and the welfare of households are suggested to be closely linked to deforestation, policy recommendations to reduce

---

<sup>9</sup> There are two methods for correcting similar data problems noted in the literature. The first is the use of instrumental variables (Marquis, Marquis, and Polich 1986; Bollinger 2001; Hyslop and Imbens 2001). This method is appropriate when correlation exists between the measurement error and another variable such that independent information can be used to correct the variance-covariance matrix. Otherwise, weights can be used to correct for any resulting bias.

<sup>10</sup> The weight for the estimation of deforestation is therefore calculated as:  $\text{weight} = -8.58 + \text{Education} * 0.24 + \text{Age} * 0.12 + \text{Family} * -0.02 + \text{Years on lot} * -0.16 + \text{Distance} * 0.07 + \text{Total Income} * 0.00 + \text{Vehicles} * -0.13 + \text{Cattle} * -0.03 + \text{Reported Lot Size} * 0.17$ .

<sup>11</sup> Again, this reliability index is based on the assumption that the errors are uncorrelated.

TABLE 3  
Re-Estimation of Land Cleared and Primary Forest – Corrected Models

	<u>Estimations of Deforestation (in hectares)</u>				<u>Estimations of Primary Forest (in hectares)</u>			
	<u>Survey Data</u>		<u>Corrected Survey Data</u>		<u>Survey Data</u>		<u>Corrected Survey Data</u>	
	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.	Coeff.	Std.Err.
Constant	19.79***	6.04	27.92***	7.47	-19.79***	6.04	-29.05***	7.28
Education	-1.14***	0.40	-2.02***	0.47	1.14***	0.40	2.05***	0.46
Age	-0.10	0.09	-0.16	0.10	0.10	0.09	0.17	0.10
Family	0.12	0.18	0.17	0.20	-0.12	0.18	-0.20	0.21
Years on Lot	-0.25	0.18	-0.26	0.21	0.25	0.18	0.25	0.21
Distance	-0.14***	0.05	-0.12**	0.06	0.14***	0.05	0.13**	0.06
Total Income	-0.00**	0.00	-0.00***	0.00	0.00*	0.00	0.00***	0.00
Vehicles Owned	3.64**	1.83	5.60***	2.05	-3.64**	1.82	-5.70***	2.04
Cattle	0.11***	0.02	0.16***	0.02	-0.11***	0.02	-0.17***	0.02
Reported Lot Size	0.61***	0.03	0.55***	0.03	0.39***	0.03	0.47***	0.03
R-Squared	.89		.91		.56		.71	
Adj R-Squared	.88		.90		.53		.70	
Observations	171		171		171		171	

\*, \*\*, \*\*\* indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

pressure on forests that contain programs to maintain or improve current welfare levels are likely to be more successful.

## 6. CONCLUSIONS

Research on the underlying causes of deforestation has been undertaken at the national, state, municipal, and household scales (see Barbier and Burgess 2001; Kaimowitz and Angelsen 1998; Rudel, Bates, and Machinguashi 2002 for reviews). The use of deforestation data collected at various scales, from different sources, is becoming more common, thus providing greater opportunities for researchers to evaluate and validate data sources. Previous studies of deforestation that combine satellite and survey data at a more macro scale have found roads, agroclimate conditions, and government policy to be major determinants of deforestation in the Amazon (Chomitz and Gray 1996; Pfaff 1999; Chomitz and Thomas 2003; Geist and Lambin 2001). We identify several determinants of deforestation at the household level, including distance to markets and indicators of household welfare. More importantly, we find that additional determinants are identified when our data are corrected, leaving us to recommend that similar processes be established in future research to better inform policy.

In testing the reliability of a stratified random sample of household responses on deforestation and forested land, we find the reported survey data to be reliable when verified against satellite imagery of the same landholdings. This result is encouraging for research conducted in developing regions, which is often dependent upon self-reported survey data. Additionally, it appears that survey error is not significantly influenced by any variables indicating purposeful manipulation. Most error is found to be attributable to variables indicating a lack of knowledge about very large landholdings. Our findings therefore support the notion that households answered the land use survey questions truthfully using their best available knowledge.

The analysis also suggests that the utilization of an independent source of data such as satellite imagery can improve data reliability and estimations. The major influences of survey error were identified using the GIS estimates and corrected using the error estimates. We find that the uncorrected estimations of deforestation underestimate the importance of some household characteristics while overestimating the impact of the landholding size. These results are significant for policy makers because the corrected data may suggest a shift in policy focus. Rather than directing policy as reducing the impact and accumulation of cattle in the region, the results suggest that investment in human capital may also help reduce pressure on these forests.

## REFERENCES

- Alves, D.S., J.L.G. Pereira, C.L. De Sousa, J.V. Soares, and F. Yamaguchi, 1999. "Characterizing Landscape Changes in Central Rondônia Using Landsat TM Imagery," *International Journal of Remote Sensing* 20, 2877-2882.



- Aiger, D.J., 1973. "Regression with a Binary Independent Variable Subject to Errors of Observation," *Journal of Econometrics* 1, 49-59.
- Barbier, E.B. and J.C. Burgess, eds., 2001. "Special Issue on Tropical Deforestation," *Land Economics* 77(2), 155-314.
- Bollinger, C., 1998. "Measurement Error in the Current Population Survey," *Journal of Labor Economics* 16, 576-594.
- Bollinger, C., 2001. "Response Error in the Union Wage Differential," *Southern Economic Journal* 68, 60-76.
- \_\_\_\_ and M. David, 1997. "Modeling Discrete Choice with Response Error: Food Stamp Participation," *Journal of the American Statistical Association* 92, 827-835.
- Bound, J. and A. Kruger, 1991. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics* 9, 1-24.
- Brondizio, E.S., S.D. McCracken, E. Moran, A.D. Siqueira, D.R. Nelson, and C. Rodriguez-Pedraza, 2002. "The Colonist Footprint: Towards a Conceptual Framework of Land Use and Deforestation Trajectories Among Small Farmers in Frontier Amazonia," in *Deforestation and Land Use in the Amazon*, C.H. Wood and R. Porro (eds.). University Press of Florida: Gainesville, FL.
- Brondizio, E.S., E. Moran, P. Mausel, and Y. Wu, 1996. "Land Cover in the Amazon Estuary: Linking of the Thematic Mapper with Botanical and Historical Data," *Photogrammetric Engineering and Remote Sensing* 62(8), 921-929.
- Caviglia, J.L., 1999. *Sustainable Agriculture in Brazil: Economic Development and Deforestation*, New Horizons in Environmental Economics series. Edward Elgar: United Kingdom.
- \_\_\_\_ and J.R. Kahn, 2001. "Diffusion of Sustainable Agriculture in the Brazilian Tropical Rain Forest: A Discrete Choice Analysis," *Economic Development and Cultural Change* 49(2), 311-333.
- Caviglia-Harris, J.L., 2004. "Household Production and Forest Clearing: The Role of Farming in the Development of the Amazon," *Environment and Development Economics* 9(2), 181-202.
- \_\_\_\_ and E. Sills, 2005. "Land Use and Income Diversification: Comparing Traditional and Colonist Populations in the Brazilian Amazon," *Agricultural Economics* 32, 221-237.
- Chomitz, K.M. and D.A. Gray, 1996. "Roads, Land Use and Deforestation: A Spatial Model Applied to Belize," *The World Bank Economic Review* 10(3), 487-512.
- Chomitz, K.M. and T.S. Thomas, 2003. "Determinants of Land Use in Amazônia: A Fine-Scale Spatial Analysis," *American Journal of Agricultural Economics* 85(4), 1016-1028.
- Evans, T., A. Manire, F. de Castro, E. Brondizio, and S. McCracken, 2001. "A Dynamic Model of Household Decision-Making and Parcel Level Landcover Change in the Eastern Amazon," *Ecological Modeling* 143, 95-113.
- Fearnside, P., 2001. "Land-Tenure Issues as Factors in Environmental Destruction in the Brazilian Amazonia: The Case of Southern Para," *World Development* 29(8), 1361-1372.

- Fitzgerald, J., P. Gottschalk, and R. Moffit, 1998. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics," *The Journal of Human Resources* 33, 251-299.
- Geist, H.J. and E.F. Lambin, 2001. *What Drives Tropical Deforestation? A Meta Analysis of Proximate and Underlying Causes of Deforestation Based on Subnational Case Study Evidence*, LUCC Report Series No. 4. University of Louvain.
- Geoghegan, J., S.C. Villar, P. Klepeis, P.M. Mendoza, Y. Ogneva-Himmelberger, R.R. Chowdhury, B.L. Turner II, and C. Vance, 2001. "Modeling Tropical Deforestation in the Southern Yucatan Peninsular Region: Comparing Survey and Satellite Data," *Agriculture, Ecosystems, and Environment* 84, 25-46.
- Hyslop D.R. and G.W. Imbens, 2001. "Bias Form Classical and Other Forms of Measurement Error," *Journal of Business and Economic Statistics* 19, 475-481.
- Instituto Brasileiro de Geografia e Estatística, "Censo Demográfico 2000 - Sinopse Preliminary," Tabela - População residente, em valores absolutos e relativos, total, em situação urbana e em situação urbana na sede municipal, área total e densidade demográfica, segundo as Unidades da Federação e Municípios, Web site: [www.ibge.gov.br](http://www.ibge.gov.br), June 2001.
- International Monetary Fund, 1999. *World Economic Outlook May 1999*, Washington, D.C.
- International Monetary Fund, 2000. *World Economic Outlook May 2000*, Washington, D.C.
- Kaimowitz, D. and A. Angelsen, 1998. *Economic Models of Tropical Deforestation: A Review*. Center for International Forestry Research: Indonesia.
- Lambin, E.F., 1997. "Modelling and Monitoring Land-Cover Change Processes in Tropical Regions," *Progress in Physical Geography* 21(3), 375-393.
- \_\_\_\_\_, M.D.A. Rounsevell, H. Geist, 2002. "Are Agricultural Land-Use Models Able to Predict Change in Land-Use Intensity," *Agricultural Ecosystems and the Environment* 82, 321-331.
- Marquis, K.H., M.S. Marquis, and J.M. Polich, 1986. "Response Bias and Reliability in Sensitive Topic Surveys," *Journal of the American Statistical Association* 81, 381-389.
- McCracken, S., A.D. Siqueira, E.F. Moran, and E.S. Brondizio, 2002. "Land-Use Patterns on an Agricultural Frontier in Brazil: Insights and Examples from a Demographic Perspective," in *Deforestation and Land Use in the Amazon*, C.H. Wood and R. Porro (eds.). University Press of Florida: Gainesville, FL.
- Mendelsohn, R., 1994. "Property Rights and Tropical Deforestation," *Oxford Economic Papers* 46, 750-756.
- Messonnier, M.L., J.C. Bergstrom, C.M. Cornwell, R.J. Teasley, and H.K. Cordell, 2000. "Survey Response-Related Biases in Contingent Valuation: Concepts, Remedies, and Empirical Application to Valuing Aquatic Plant Management," *American Journal of Agricultural Economics* 83, 438-450.
- Muller, D. and M. Zeller, 2002. "Land Use Dynamics in the Central Highlands of Vietnam: A Spatial Model Combining Village Survey Data with Satellite Image Interpretation," *Agricultural Economics* 27, 333-354.

- Munroe, D.K., J. Southworth, and C.M. Tucker, 2002. "The Dynamics of Land-Cover Change in Western Honduras: Exploring Spatial and Temporal Complexity," *Agricultural Economics* 27, 355-369.
- Nevo, A., 2003. "Using Weights to Adjust for Sample Selection When Auxiliary Information is Available," *Journal of Business and Economic Statistics* 21, 43-52.
- Perz, S.G. and D.L. Skole, 2003. "Secondary Forest Expansion in the Brazilian Amazon and the Refinement of Transition Theory," *Society and Natural Resources* 16, 277-294.
- Peterson, W., 2003. "Climatological Overview of the Rondônia Region," available at [www.radarmet.atmos.colostate.edu/~walt/brzclim.html](http://www.radarmet.atmos.colostate.edu/~walt/brzclim.html), July 24, 2003.
- Pfaff, A., 1999. "What Drives Deforestation in the Brazilian Amazon?" *Journal of Environmental Economics and Management* 37(1), 26-43.
- Rindfuss, R.R., P. Prasartkul, S.J. Walsh, B. Entwisle, Y. Sawangdee, and J.B. Vogler, 2003a. "Household Parcel Linkages in Nang Rong, Thailand: Challenges of Large Samples," in *People and the Environment: Approaches for Linking Household and Community Surveys to Remote Sensing and GIS*, J. Fox, R.R. Rindfuss, S.J. Walsh, and V. Mishra (eds.). Kluwer Academic Publishing Group: Dordrecht.
- Rindfuss, R.R., S.J. Walsh, V. Mishra, J. Fox, and G.P. Dolcemascolo, 2003b. "Linking Household and Remotely Sensed Data: Methodological and Practical Problems, in *People and the Environment: Approaches for Linking Household and Community Surveys to Remote Sensing and GIS*, J. Fox, R.R. Rindfuss, S.J. Walsh, and V. Mishra (eds.). Kluwer Academic Press: Dordrecht, Netherlands.
- Roberts, D.A., I. Numata, K. Holmes, G. Batista, T. Krug, A. Monteiro, B. Powell, and O.A. Chadwick, 2002. "Large Area Mapping of Land-Cover Change in Rondônia Using Multitemporal Spectral Mixture Analysis and Decision Tree Classifiers," *Journal of Geophysical Research* 107, 401-418.
- Rudel, T.K., D. Bates, and R. Machinguashi, 2002. "A Tropical Forest Transition? Agricultural Change, Out-migration and Secondary Forests in the Ecuadorian Amazon," *Annals of the Association of American Geographers* 92(1), 87-102.
- Staal, S.J., I. Baltenweck, M.M. Waithaka, T. deWolff, and L. Njoroge, 2002. "Location and Uptake: Integrated Household and GIS Analysis of Technology Adoption and Land Use, with Application to Smallholder Dairy Farms in Kenya," *Agricultural Economics* 27, 295-315.
- U.S. Department of Commerce, 2001. National Trade Data Bank, available at [www.stat-usa.gov](http://www.stat-usa.gov), July 2001.
- Vance, C. and J. Geoghegan, 2002. "Temporal and Spatial Modelling of Tropical Deforestation: A Survival Analysis Linking Satellite and Household Survey Data," *Agricultural Economics* 27, 317-332.
- Walker, R., E. Moran, and L. Anselin, 2000. "Deforestation and Cattle Ranching in the Brazilian Amazon: External Capital and Household Processes," *World Development* 28(4), 683-699.