Journal of Environmental Management (1990) **31**, 247–259

An Analysis of Anthropogenic Deforestation Using Logistic Regression and GIS

Aaron Kim Ludeke

Texas Natural Resources Information System, P.O. Box 13231, Austin, Texas 78711, U.S.A.

Robert C. Maggio

Department of Forest Science, Texas A&M University, College Station, Texas 77843, U.S.A.

and Leslie M. Reid

Department of Recreation, Park and Tourism Sciences, Texas A&M University, College Station, Texas 77843, U.S.A.

Received 16 *April* 1989

The anthropogenic deforestation occurring throughout the tropical world is one of the great crises of our time. To better understand the dimensions of this problem, a number of studies have sought to quantify the extent and rate of tropical forest loss. However, environmental planners and managers need to know more than the extent and rate of tropical deforestation. There is a need for predictability, i.e. which areas are most susceptible to deforestation. This study demonstrates a methodology for predicting those areas with the greatest propensity for deforestation based on natural and cultural landscape variables. Logistic regression analysis was used to determine variables most closely associated with deforestation. The independent variables most strongly associated with the forested/deforested dependent variable were location of forested areas in relation to the forested/deforested edge in an earlier time period, to the location of access in an earlier time period, and to the location of habitation in an earlier time period. GIS analysis was then used to verify spatially the close statistical relationship between the dependent variable and each of the independent variables selected by the logistic regression modeling.

Keywords: GIS, logistic regression, tropical deforestation, Honduras.

1. Introduction

The continuing loss of the world's tropical forests is leading to massive environmental disruptions, including an historically unprecedented rate of species extinction. Loss of

0301–4797/90/070247 + 13 $03.00/0

protective forest cover reduces the sustainable yield of watersheds which can lead to alternating floods and water shortages. Deforestation also leads to depletion of soil nutrients, to soil erosion, and to changes in local and, some evidence indicates, global climate (U.S. Interagency Task Force on Tropical Forests, 1980).

These environmental disruptions increase the risk to human populations from both natural and man-made disasters. Most directly affected are local populations in developing countries—those populations most closely tied to the natural resource base for their livelihood and economic development. The threats posed by loss of the tropical forests range from the specter of short-term resource depletion to widespread ecological catastrophes. Further, the continued depletion of the natural resource base decreases the options for a long-term sustainable society.

Several recent studies have attempted the difficult task of inventorying the extent of tropical forests and their rate of destruction (Persson, 1974; Sommer, 1976; Myers, 1980; Lanly, 1983). Other studies have combined the results of inventories with other data in an attempt to understand deforestation in developing countries. Allen and Barnes (1985) concluded that, in the short term, deforestation is the result of population growth and agricultural expansion. They identified wood harvesting for fuel and export as long-term factors affecting deforestation. Bowonder (1985–1986) stated that deforestation in developing countries is mainly the result of poverty and underdevelopment.

Environmental planners and managers need to know the rate and extent of deforestation for a given region and what conditions in the society lead to uncontrolled anthropogenic deforestation. Also important is the location of deforestation in relation to natural and cultural landscape variables. Understanding the relationship between natural and cultural landscape variables and deforestation is an important step in understanding the deforestation process. Better insight into tropical deforestation is critical since deforestation has been called "the greatest tragedy that the Third World has to face during this century" (Bowonder, 1985–1986, p. 171).

Two major research efforts have been undertaken to map tropical deforestation in relation to natural and cultural landscape variables. In the 1970s remotely sensed data were used in conjunction with geographic information system (GIS) analysis to develop a model of land use change in northern Thailand. Nualchawee *et al.* (1981) used the dependent variable—land cover (including six forest cover types and three agricultural land cover types), and the independent variables—cultural features, transportation networks, and physiographic features to develop a land cover change model. Forest land cover maps were developed from remotely sensed data for six different dates. Two different approaches were used to develop a multivariate, multi-temporal model of deforestation patterns. The first was a Markov trend model used to project future land use from trends in the recent past. This descriptive model could only indicate a general trend, not provide detailed spatial land cover change information. It could not identify those areas with a higher probability of deforestation. The second methodology used discriminant analysis to correlate the dependent variable "change in land cover" with the independent variables slope, elevation, proximity to housing, proximity to rivers and streams, and proximity to roads and trails. The authors concluded that the "results provide encouragement that the approach can project the spatial occurrence and alteration of shifting cultivation in tropical forests." (Nualchawee *et al.*, 1981, p. 77).

Sader and Joyce (1985) used published maps depicting the dependent variable forest area and the independent variables—life zone, soil class, slope class, and proximity to transportation networks as inputs for a GIS analysis of deforestation in Costa Rica. They proposed that the propensity of a forested area to be cleared of primary forest is a function of both biophysical characteristics of the landscape and human-related factors. By quantifying these relationships, they sought to identify predictors of forest cover change for four different time periods. The resulting tabular summaries revealed both rates and trends of forest clearing over a period of 40 years ending in 1977. In general, they found less deforestation had occurred: (1) in life zones with the lowest PET/P (potential-evapotransportation/precipitation) ratios; (2) as the slope gradient increased; (3) at locations not adjacent to roads and railroads; (4) on soils with moderate to strong limitations for crop and pasture use; and (5) in areas assigned forest protection status such as national parks and biological reserves (Sader and Joyce, 1985).

In an extension of this study, stepwise descriminant analysis was used to determine the relative strength of the assocations between the independent variables and deforestation. Road development was the most significant variable associated with deforestation in three of the time periods. Slope gradient was found to be the second most significant variable associated with deforestation in the last three time periods. Life zone was the most important landscape variable associated with deforestation in the first two time periods, but was of decreased importance in subsequent periods (Sader, 1987).

2. Setting

This study of the association between deforestation and natural/cultural landscape variables is set in the Cordillera Nombre de Dios of Honduras. This region is the site of a new national park/biosphere reserve with Pico Bonito, the highest peak in this range, providing the name and focus for the park. The boundaries of this study are delineated by the La Ceiba to Tela coastal highway to the north (as depicted by the La Ceiba and La Masica Instituto Geografico Nacional (IGN) first edition map sheets) and the Rio Aguan to the south. The eastern boundary is defined by the Rio Cangrejal and the Rio Yaruca. The western boundary follows the Rio Cuero south, then trends southeast to the Rio Aguan. A map depicting the location of the study area is given in Figure 1.

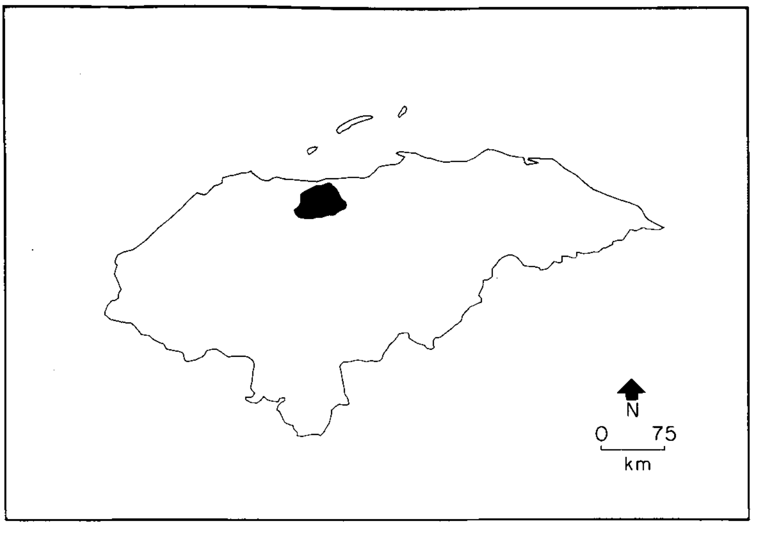

Figure 1. Map of Honduras. Study area is in black.

3. Methodology

3.1. MAP DEVELOPMENT

Base maps of the study area were developed from 1:50 000 quadrangle maps produced by the Instituto Geografico Nacional (IGN) of Honduras. All or part of the following map sheets were included in this study: (1) Arenal; (2) Jimia; (3) Jutiapa; (4) La Ceiba; (5) La Masica; (6) Olanchito; (7) Pico Bonito; and (8) Yaruca as shown in Figure 2.

These maps were compiled using 1:60 000 black and white aerial photographs covering the 1954 to 1965 (T_1) time period. Available aerial photography used to compile these maps was acquired from IGN and photo-interpreted for comparison with the published maps. Areas deforested and areas with intact forest cover were delineated on map overlays of each of the base maps. Additional overlays developed for this study included the following classes:

1. proximity of areas forested in T_1 to the nearest forested/deforested edge;
2. proximity to the nearest T_1 access (road, trail, or railroad right of way);
3. proximity to the nearest T_1 road;
4. proximity to the nearest T_1 house/shelter;
5. proximity to the nearest T_1 urban center;
6. proximity to the nearest T_1 perennial river or stream;
7. soils class based on Miller's (1982) soils map of Honduras;
8. life zone class based on the Holdridge Life Zone System (Holdridge, 1967);
9. department (state);
10. elevation class; and
11. slope class.

In addition, 1:40 000 black and white aerial photographs of the study area for the 1977–1978 time period (T_2) was acquired from the Honduras National Cadaster. Forested/deforested areas were delineated using photointerpretation and the resulting

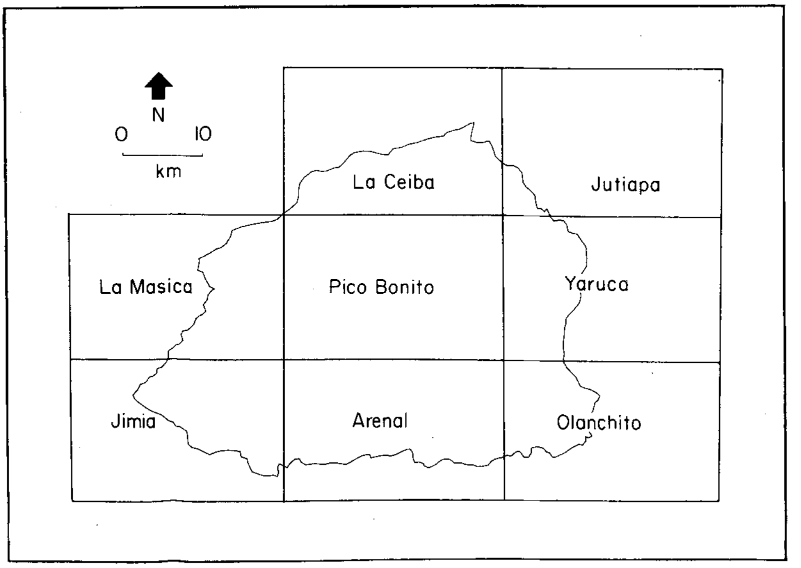

Figure 2. Base maps of the study area—developed from 1:50 000 scale quadrangle maps produced by the Instituto Geographico Nacional (IGN) of Honduras.

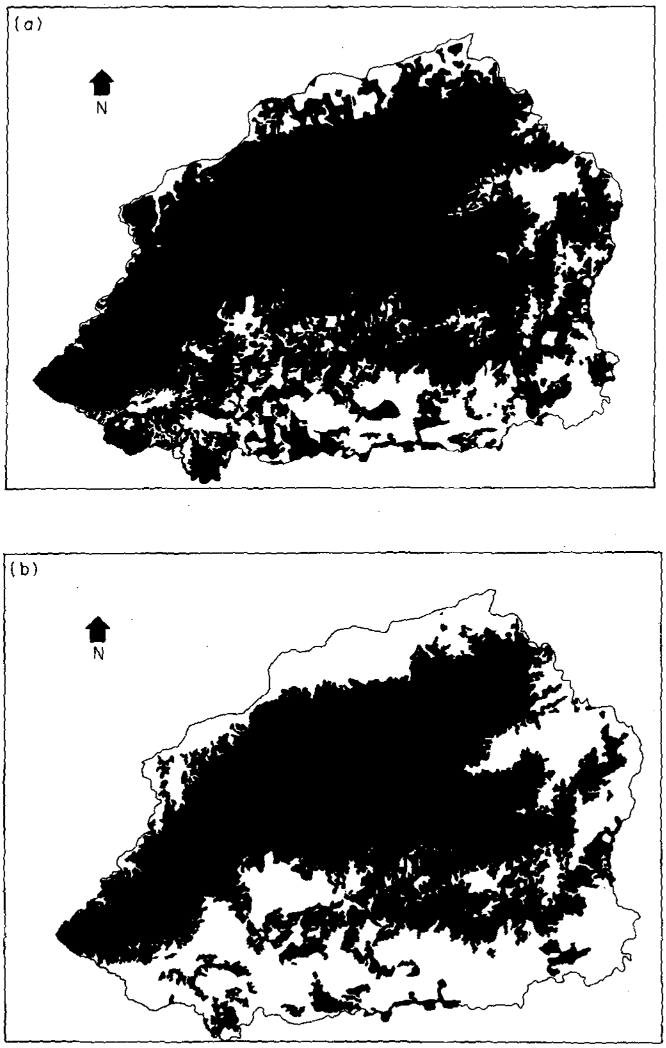

Figure 3. (a) Forested/deforested in T_1. Black indicates areas forested. White indicates areas deforested. (b) Forested/deforested in T_2. Black indicates areas forested. White indicates areas deforested.

maps compared with the T_1 forested/deforested maps. Areas deforested in T_1 but appearing forested in T_2 were reclassified as deforested in T_2 since the focus of this study was on intact primary forest. Figure 3 provides a comparison of the forest cover between these two time periods [Figure 3(a)—T_1 and Figure 3(b)—T_2].

The resulting thematic maps were entered into a GIS data base using micro-GIS, a microcomputer-based software package developed by the Texas A&M Department of Forest Science (Maggio and Wunneburger, 1986). These data files were then reformatted and transferred to the SAGIS software package (a public domain GIS package developed by the National Park Service) on the VAX 8600 super-minicomputer at the Texas A&M University Computing Services Center.

3.2. SYSTEMATIC SAMPLING

Each map was divided into 0·25 km^2 grid cells based on the Universal Transverse

Mercator (UTM) system. A 10% systematic sample of these grid cells was manually entered into a SAS (Statistical Analysis System) data base. Only those grid cells falling completely within the study area were included in the systematic sample. Each sampled grid cell was categorized by soils class, department, and life zone, based on the predominant class found within that grid cell. Only those sampled grid cells which appeared to retain 100% of their natural forest cover were classified as forested. Otherwise they were classified as deforested. The proximity variables were recorded as the distance from the centroid of a sampled grid cell to the nearest forested/deforested edge, the nearest road, trail or railroad (access), the nearest road, the nearest house or shelter, the nearest urban center, and the nearest perennial river or stream. The resulting values were recorded as both categorical and continuous data. Slope, elevation, and aspect data were recorded for the sampled grid cells using topographical information on the original IGN maps. Aspect (not included on a thematic map but still recorded for sampled grid cells) was recorded as categorical data whereas elevation and slope were recorded as both continuous and categorical data.

4. Statistical Analysis

4.1. LOGISTIC REGRESSION

Logistic regression was used because the object of this research was to test the relationship between a dichotomous dependent variable (forested/deforested) and independent variables which cannot be assumed to satisfy the required assumptions of discriminant analysis (Press and Wilson, 1978). Two different SAS procedures were used to compute models for comparison; (1) LOGIST (Harrell, 1983) was used when the independent variables were continuous; and (2) CATMOD (SAS Institute Inc., 1985) was used when the independent variables were categorical or both categorical and continuous (SAS Institute Inc., 1985).

Model comparison focused on several criteria. First, the overall ability of the independent variable(s) to predict the dependent variable was compared. Second, the model R^2 for each model was calculated. The model R^2 is not computed the same way as the R^2 used in regular regression. It does give an indication of the fit of the model, but does not give as much information as the regression R^2 on the scatter of the data about the fitted line. Third, the parsimony or conciseness of each model was compared with the other models. Finally, the model χ^2 (Chi squared) was used to test the hypothesis that the model is non-predictive.

Based on the above statistical criteria, a model comparison table (Table 1) was constructed. From Table 1 it is clear that, if the location of access in T_1 is known, the dependent variable, forested/deforested, can be predicted for T_2 with 87% accuracy. Information on the location of the nearest T_1 house/shelter gives a T_2 forested/deforested prediction of 86%. A third model which gives a 91% overall predicted—the best of any model tested, with a model R^2 of 0·64—should also be noted. It was not included in the model comparison table (Table 1) since, by definition, it cannot be used to test the T_1 dependent variable. This model demonstrates that information on the location of the forested/deforested edge in T_1 can be used to predict forested/deforested patterns in T_2. The model R^2s are large relative to the degrees of freedom, and the probability of obtaining a greater value of χ^2 at random is less than 0·005 for these models, each developed using a single independent variable.

TABLE 1. Model comparison table

Cover T_2											Cover T_1	
Overall correct (%)	Model R^2	Proximity to access T_1	Proximity to access class T_1	Proximity to house T_1	Proximity to house class T_1	Proximity to road T_1	Slope	Slope class	Soil class	Life zone	Model R_2	Overall correct (%)
87	0·56	†									0·40	79
86	0·54			†							0·37	78
87	0·56			†		†					0·36	80
88	0·58			†			†				0·39	80
86	0·97		†		†						0·97	80
80	0·95								†	†	0·87	81
79	0·86						†			†	0·84	80
89	0·58			†		†			†		0·41	82
87	0·90		†					†	†		0·87	82
84	0·45					†	†		†		0·34	79

† Independent variable(s) used to construct each model of T_1/T_2 cover.

TABLE 2. Model Contingency Table for proximity to edge in T_1. This table compares the actual vs. the predicted for both forested and deforested in T_2

		Predicted: Forested	Predicted: Deforested
Actual	Forested	258 92·1%	37 10·3%
	Deforested	22 7·8%	322 89·7%

The use of model contingency tables to break down the dependent variable into a forested and deforested component help to examine how well these models predict deforestation. Table 2 gives the breakdown for the "proximity of the T_2 forested to the T_1 forested/deforested edge" model. This model gives a T_2 deforestation prediction of almost 90% indicating the strong relationship between deforestation and the forested/deforested edge.

TABLE 3. Model Contingency Table for proximity to access in T_1. This table compares actual vs. predicted for both forested and deforested in T_2

		Predicted: Forested	Predicted: Deforested
Actual	Forested	238 90·2%	57 15·2%
	Deforested	26 9·8%	318 84·8%

Table 3 gives an almost 85% correct deforestation prediction for T_2 given information of "proximity to access in T_1". This indicates a strong relationship between the location of access and deforestation.

TABLE 4. Model Contingency Table for proximity to house/shelter in T_1. This table compares actual vs. predicted for both forested and deforested in T_2

		Predicted: Forested	Predicted: Deforested
Actual	Forested	234 89·7%	61 16·1%
	Deforested	27 10·3%	317 83·9%

Table 4 gives a correct deforestation prediction of almost 84% given information on the "proximity of house/shelter in T_1". There appears to be a strong relationship between the location of houses and shelters and deforestation.

4.2. DESCRIPTIVE STATISTICS

To better understand the results of the logistic regression modeling, it is useful to examine descriptive statistics for the best models. These descriptive statistics are calculated using categorical or class data. The change in the percent deforested from T_1 to T_2 for areas forested in T_1 located within different proximity classes from the forested/deforested edge in T_1 was computed. The results are as follows:

1. >0–0·5 km—78·2%
2. >0·5–1 km—41·4%
3. >1 km—5·3%.

Since lands in these proximity classes were forested in T_1 these figures also give the total per cent of each deforested in T_2. It is clear that areas adjacent to areas previously deforested are most susceptible to deforestation pressure. This pressure decreased with distance from the edge. Only slightly more that 5% of the land more than 1 km from the T_1 forested/deforested edge had been deforested by T_2.

Table 5 shows the relationship between deforestation and the proximity of the sampled grid cells to access routes. From this table, it is clear that areas closer to access routes are more susceptible to deforestation pressures in both time periods. However, the highest deforestation rate from T_1 to T_2 does not occur in the nearest proximity class, but in the next one out. There was a steep drop in the per cent area deforested and actual area deforested beyond 2 km from access routes.

Table 6 shows the relationship between deforestation and the nearest house or shelter. As with the proximity to access class, there was a clear relationship between the per cent deforested and the "proximity to house/shelter" class. A grid cell located nearer to a house or shelter was more likely to be deforested in both T_1 and T_2. Again, the highest deforestation rate between T_1 and T_2 occurred not in the nearest class, but the next class out. There was a rapid drop in the per cent area deforested and the actual area deforested beyond 2 km.

TABLE 5. This table shows five different proximity to access classes, the total percentage of the study area in each class, the number of grid cells sampled in this study, the per cent of each class deforested in T_1, in T_2, and in the period from T_1 to T_2

Proximity to Access Class-T_1 (km)	Area in Class (GIS) (%)	No. of Sample Grid Cells	Deforested T_1 (%)	Deforested T_2 (%)	Deforested T_1 to T_2 (%)
1 (0–0·5)	34·7	210	75·3	92·9	17·6
2 (>0·5–1)	16·0	105	60·0	82·9	22·9
3 (>1–2)	16·2	112	27·7	49·1	21·4
4 (>2–5)	22·8	142	2·8	4·9	2·1
5 (>5)	10·4	70	0·0	0·0	0·0
Total	99·9†	639	40·1	53·8	13·8†

† Rounding error.

TABLE 6. This table shows five different proximity to house/shelter classes, the total percentage of the study area in each class, the number of grid cells sampled in this study, the per cent of each class deforested sampled in this study, the per cent in each class deforested in T_1 and T_2, and in the period from T_1 to T_2

Proximity to House/shelter Class-T_1 (km)	Area in Class (GIS) (%)	No. of Sample Grid Cells	Deforested T_1 (%)	Deforested T_2 (%)	Deforested T_1 to T_2 (%)
1 (0–0·5)	22·1	122	81·2	98·4	17·2
2 (>0·5–1)	19·6	130	63·1	86·9	23·8
3 (>1–2)	22·8	151	42·4	63·6	21·2
4 (>2–4)	17·6	117	9·4	12·8	3·4
5 (>4)	18·0	119	0·0	0·0	0·0
Total	100·1†	639	40·1	53·8	13·8†

† Rounding error.

5. GIS Analysis

The SAGIS software was used to compute the percentage of the study area which had been deforested for each of the time periods. These results were compared with the SAS output (see Tables 5 and 6). The GIS analysis indicated the area deforested by T_1 was 23·8% of the total study area. This contrasted with the SAS figure of 40·1%. The GIS T_2 deforestation figure was 41·1% in contrast to the SAS figure of 53·8%. The deforestation rate from T_1 to T_2 is 13·8% according to the SAS statistics but 17·4% according to the GIS output. The larger deforested area given by the SAS output is an artifact of the method used to discriminate between grid cells which had been deforested and those with no deforestation. Since a partially deforested grid cell was counted as deforested, a high bias for deforestation using the sampling procedure is expected. The lower rate of deforestation from T_1 to T_2 using SAS is consistent with this methodological bias. Sampled grid cells with a small percentage of their total areas deforested in T_1 were counted as deforested. Additional deforestation occurring within that grid cell from T_1 to T_2 did not affect the per cent deforested since the cell was already counted as deforested.

Finally, the SAGIS software was used to produce overlay maps displaying the intersection of the dependent variable in T_2 with each of the most significant independent variables. Figure 4 gives the location of areas deforested in T_2 in relation to areas within 1 km of the T_1 forested/deforested edge. The SAS statistics showed that the highest percentage of deforestation from T_1 to T_2 occurred within 1 km of the T_1 forested/deforested edge. Figure 4 provides a spatial verification of this strong relationship.

Figure 5 gives the location of areas deforested in T_2 in relation to areas within 2 km of the nearest T_1 access. As Table 5 demonstrates, most of the deforestation from T_1 to T_2 was concentrated in areas within 2 km from access in T_1. Figure 5 gives spatial support for the strong statistical relationship.

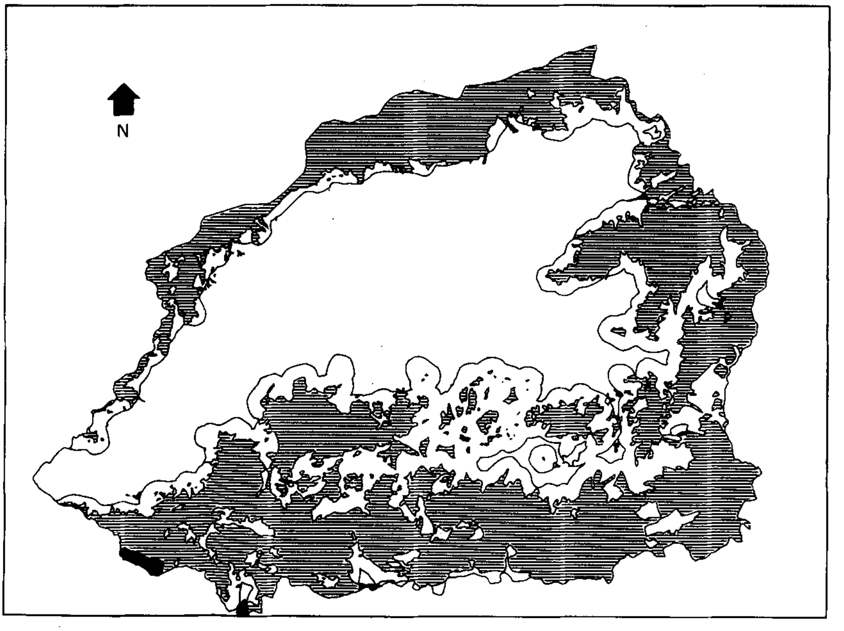

Figure 4. Proximity of T_2 deforested to T_1 forested/deforested edge. Shaded areas were deforested in T_2 and within 1 km to T_1 edge. Black areas were deforested in T_2 and more than 1 km from T_1 forested/deforested edge.

Figure 6 gives the location of areas deforested in T_2 in relation to areas within 2 km from the nearest T_1 house/shelter. Table 6 shows that most of the deforestation from T_1 to T_2 was concentrated in areas within 2 km from the nearest T_1 house or shelter. Figure 6 provides corroboration for the strong statistical relationship.

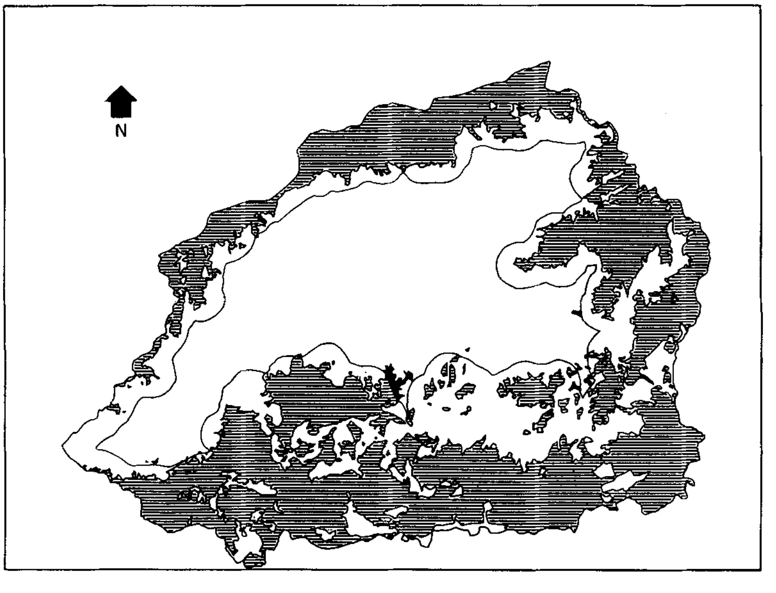

Figure 5. Proximity of T_2 deforested to T_1 access. Shaded areas were within 2 km of T_1 access and deforested. Black areas were more than 2 km from T_1 access and deforested.

Figure 6. Proximity of T_2 deforested to T_1 house/shelter. Shaded areas were deforested and within 2 km of T_1 house/shelter. Black areas were deforested and greater than 2 km from T_1 house/shelter.

6. Conclusion

This study provides both statistical and spatial confirmation of the strong spread effect of deforestation indicated by other studies (Conklin, 1961; Myers, 1980). Information of the location of the forested/deforested edge in one time period provided a 90% correct prediction of areas deforested in a later time period.

This study also provides both statistical and spatial confirmation of the importance of access in the location of deforestation. This observation has been made by other studies: (Ruthenberg, 1971; Golley *et al.*, 1971; Sader and Joyce, 1985; Allen and Barnes, 1985). However, most of these studies have focused on road access, whereas this study examined both roads and a combination of roads, trails, and railroad access. This study showed that the location of combined access provided a higher prediction of deforestation than the independent variable location of roads. Information on the location of access in T_1 provided an 85% correct prediction of deforestation in T_2.

Finally, location of houses and shelters was shown to be spatially and statistically related to deforestation. Duncan (1978), in his case study of highland Honduras, found that most shifting cultivation sites were within 4 km from the cultivator's home. The descriptive statistics produced by this study confirmed Duncan's finding (see Table 6). Further, this study demonstrated that information on the location of houses and shelters in T_1 allowed a T_2 deforestation prediction of 84%.

Although these single independent variable models provide strong predictions of deforestation, Table 1 shows eight multivariate models which also provide strong predictions of deforestation. However, the single variable models on which this study focused provide strong predictions of deforestation without the need for the additional variable(s) found in the other models. The results of this study indicate that a planner or manager with the limited information found in the best single variable models can predict those areas most susceptible to deforestation pressure.

References

Allen, J. C. and Barnes, D. F. (1985). The causes of deforestation in developing countries. *Annals of the Association of American Geographers,* **75**, 163–184.

Bowonder, B. (1985–86). Deforestation in developing countries. *Journal of Environmental Systems* **15**, 171–192.

Conklin, H. C. (1961) The study of shifting cultivation. *Current Anthropology* **2**, 27–61.

Duncan, T. W. (1978). Adaptive strategies of swidden cultivators in Honduras. PhD dissertation, Wayne State University, Detroit.

Golley, F. B., Olien, M. D. and Hoy, D. R. (1971). Cognized environments of San Carlos valley settlers. *Revista Geografica* **74**, 33–50.

Harrell, F. E., Jr. (1983). The logist procedure. In *SAS SUGI Supplemental Library Users Guide*. pp. 181–202. Cary, North Carolina: SAS Institute Inc.

Holdridge, L. R. (1967). *Life Zone Ecology*. San Jose, Costa Rica: Tropical Science Center.

Lanley, J. P. (1983). Assessment of the forest resources of the tropics. *Forestry Abstracts* **44**, 287–318.

Maggio, R. C. and Wunneburger, D. F. (1986). A micro-computer based geographic information system for natural resource managers. In *Conference Proceedings, Geographic Information Systems Workshop*. pp. 269–300. April 1–4, Atlanta, Georgia.

Miller, L. L. (1982). *Unidades de Mapeo Para el Mapa de 'Suelos de Honduras la Escala 1:1,000,000*. Tegucigalpa, Honduras: Direccion Ejecutiva de Recursos Naturales.

Myers, N. (1980). *Conservation of Tropical Moist Forests*. Washington D.C. National Academy of Sciences.

Nualchawee, K., Miller, L. D., Tom, C., Christenson, J. and Williams, D. (1981). Spatial inventory and modeling of shifting cultivation and forest land cover of northern Thailand with inputs from maps, airphotos, and Landsat. *Remote Sensing Center Technical Report No. 4177*. Texas: Texas A&M University, College Station.

Persson, R. (1974). World forest resources: review of the world's forest resources in the early 1970s. *Research Notes No. 17*. Stockholm: Department of Forestry Survey, Royal College of Forestry.

Press, S. J. and Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association* **73**, 699–705.

Ruthenberg, H. (1971). *Farming Systems in the Tropics*. Oxford: Clarendon Press.

Sader, S. A. (1987). Monitoring historical change in Costa Rican forests from 1940 to present. In *Conference Proceedings, Land and Resource Evaluation for National Planning in the Tropics, International Conference and Workshop*. Chetumal, Mexico.

Sader, S. A. and Joyce, A. T. (1985). Relationship between forest clearing and biophysical, factors in a tropical environment. *Technical Commission I, Subtheme 6.1.B, IX World Forestry Congress*. Mexico City.

SAS Institute Inc. (1985). *SAS User's Guide: Statistics Version 5 Edition*. Cary, North Carolina: SAS Institute Inc.

Sommer, A. (1976). Attempt at an assessment of the world's tropical moist forests. *Unasylva* **28**, 5–24.

U.S. Interagency Task Force on Tropical Forests. (1980). The world's tropical forests: A policy, strategy, and program for the United States. *Department of State Publication 9117*. U.S. Government Printing Office, 0-311-568(627).