Investigating Tropical Deforestation Using Two-Stage Spatially Misaligned Regression Models

Deepak K. AGARWAL, Alan E. GELFAND, and John A. SILANDER, JR.

Deforestation in the tropics has been a major concern in conservation science for more than two decades. A standard explanation is population pressure, argued through descriptive statistical summaries, but the connection between local population and forest exploitation has not been clearly addressed from a formal modeling perspective. We implement such modeling here using a two-stage specification. At the first stage, we provide a spatial model for population counts. At the second stage, we provide a conditional spatial model for land use given population. A critical problem is misalignment. The population counts are recorded at various administrative unit levels. In particular, we work with town-level counts. The land-use classifications are from remotely sensed satellite images and are provided at a 1-km $\times$ 1-km pixel level. We propose a methodology to implement regressions in this situation. The motivating data are obtained for the tropical wet forest on the eastern coast of Madagascar. This is a designated hotspot rainforest featuring high species diversity and high endemism. A fairly detailed analysis connecting land use with population data for this region is presented.

Key Words: Binomial regression; Conditionally autoregressive prior; Markov chain Monte Carlo; Poisson regression; Rasterization; Spatial resolution.

1. INTRODUCTION

Deforestation in the tropics has been a major concern in conservation science for more than 20 years. Estimates of tropical deforestation over the past few decades have shown an alarming acceleration in forest lost. Concern has focused on the massive loss of biodiversity in the rainforest biome, which is thought to contain more species than all other biomes together (Whitmore 1992). The vast majority of rainforest species remain undescribed. The other major concern is the loss of ecosystem services provided by the world's rainforests. For example, some believe that rainforests play a critical role as a global carbon sink, thus

Deepak K. Agarwal is a member of AT&T Shannon Research Labs, Florham Park, NJ 07932-0971. Alan E. Gelfand is a Professor in the Department of Statistics and John A. Silander, Jr., is a Professor in the Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 7, Number 3, Pages 420–439
DOI: 10.1198/108571102348

providing a buffering capacity for atmospheric CO_2 and hence global climate change.

The standard explanation for deforestation in the tropics has been rapid population growth, associated poverty, and consequential environmental destruction (Leach and Mearns 1988; Richards and Tucker 1988; Mercier 1991; Brown and Pearce 1994; Sponsel, Headland, and Bailey 1996). In selected areas, commercial exploitation and clear cutting are also important causes of deforestation (Torsten 1992). However, conventional explanations for forest loss and environmental degradation have recently been questioned as too simplistic, misleading, or general. In the absence of a clear understanding of the connection between local populations and forest exploitation, it is not surprising that conservation and reforestation schemes have had only modest success (e.g., Olson 1984).

The goal of this article is to develop, fit, and interpret suitable stochastic models that can help clarify the foregoing connections. This is a rather challenging undertaking on several accounts. First, there are numerous factors that have been linked to land use and deforestation. These factors can be socioeconomic, e.g., population growth or economic growth; physical, e.g., topography or proximity of rivers and roads; government intervention, e.g., agriculture and/or forestry policies; or external, e.g., demand for exports or financing conditions. For any given region, typically only some of this information is available.

Second, there is no well-accepted notion of a response variable. In some cases, deforestation rates or forest areas are used. Deforestation rates are typically calculated over time spans that are limited. Areas are usually obtained from cross-sectional studies of different countries or different regions within countries. The alternative we adopt is to partition the study region into disjoint areal units and then attach a variable that is a land-use classification to each unit. Of course, such a variable is not uniquely defined in terms of number of and definition of classifications. Additionally, in describing forest cover, classifications may be partly but not entirely ordered.

Third, the explanatory variables and the response variables are typically measured in different areal units. For instance, in the dataset we investigate, the response variable is land-use classification, which is ascribed to 1-km $\times$ 1-km pixels. On the other hand, population is recorded at various administrative levels. In our case, we use town-level data, which are considered to be the most reliable. How does one develop a regression for data collected on spatially misaligned areal units?

Finally, land use and deforestation are inherently spatial processes. So, too, are many of the explanatory variables, such as population counts. Satisfying stochastic as well as mechanistic modeling should capture association between measurements on areal units in terms of proximity of these units. Introduction of suitable sets of spatial random effects provides one way to accomplish this. A further advantage is that such effects can serve as spatial surrogates for unmeasured or unavailable covariates.

Little formal modeling of deforestation was attempted until the 1980s (Granger 1998). Descriptive statistical summaries of land use or forest area obtained for certain regions at certain time points were customary. Most of the ensuing statistical work that has appeared has been based on standard multiple linear regression models relating deforestation rates or forest area to a laundry list of potential explanatory variables.

In fact, Granger listed at least 28 different variables that have been linked directly or indirectly to deforestation. In these studies, it appears that the major objective has been to maximize R^2; models with 8 or more explanatory variables have been put forward. None of these models are explicitly spatial in nature. The little work with a spatial flavor that does exist has arisen from an econometric perspective as in, e.g., Palloni (1992) or Chomitz and Gray (1996). These models assume that land use will be devoted to the activity yielding the highest rent. The modeling connects output prices to input prices through production functions, with the spatial aspect introduced solely by relating these prices to market distance.

Our approach to addressing the connection between local population and forest exploitation is to formulate a model for the joint distribution of these two variables, given other explanatory variables, that is explicitly spatial. Recalling the incompatibility of the data layers (again, land use is obtained for 1-km $\times$ 1-km pixels while population counts are obtained at the town level), we overlay the town-level map on the pixel-level map and modify town boundaries so that each pixel is contained in one and only one town. Upon such rasterization, we implement the joint modeling at the pixel level. In particular, we provide this joint distribution by modeling the unobserved pixel-level population counts and then the conditional distribution of land use given the associated count. The model for the latent population counts induces a model for the observed counts. With the inclusion of two sets of spatial effects, one associated with the town population counts model, the other with the pixel-level land-use model, a multilevel hierarchical model results.

We indicate how such models can be straightforwardly fitted using Gibbs sampling (Gelfand and Smith 1990), thus enabling full inference regarding all of the modeling levels. The proposed misalignment approach extends recent work of Mugglin and Carlin (1998) and Mugglin, Carlin, and Gelfand (2000). In the present context, the other explanatory variables we employ are also measured at the pixel level. However, in general, rasterization of misaligned data layers to a common set of pixels and then modeling the joint distribution of the variables at this level provides a basis for introduction of further incompatible data layers.

The dataset we work with was collected for a rainforest in eastern Madagascar. Madagascar has been designated as an area of high priority for conservation efforts (Mittermeier 1988; Davies, Heywood, and Hamilton 1994). The eastern tropical wet forest is a global rainforest hotspot (Meyers 1991). In Section 2, we present a detailed description of this dataset as well as some preliminary analysis. In Section 3, we elaborate on the modeling described above. Section 4 summarizes and interprets our analysis of the dataset. Finally, Section 5 describes several anticipated refinements to the dataset and how the modeling of Section 3 will have to be modified.

2. THE DATASET

Madagascar is an area of the world designated as particularly high priority for conservation efforts. It is recognized as one of seven megadiversity countries in the world

(Mittermeier 1988), and the eastern tropical wet forest in Madagascar is one of 12 global rain forest hotspots. This designation reflects a high level of species diversity. In fact, Madagascar may have up to 30% of the plant species of all of Africa but constitutes less than 2% of the area (Lowry, Schatz, and Phillipson 1997) and a high level of endemism (about 80% of the plants species are found nowhere else). At the same time, the forests in Madagascar have been under tremendous threat from deforestation. As a consequence, many of the most well-known plant and animal species are listed by the International Union for the Conservation of Nature (IUCN) as globally threatened.

There is considerable difference of opinion as to how much of the original forest is left, but estimates range down to as little as 10%, and perhaps only 25% of what is left is considered primary or undisturbed forest. Of these systems, the coastal forests are the most threatened. Estimated deforestation rates for Madagascar are quite variable, i.e., 1–5% per year or more. Some have estimated that, within 20 years, little or no forest will remain outside protected areas (about 1.85% of the total land surface). See Green and Sussman (1990) for further discussion.

The focal area for this study is the tropical rainforest biome within Toamasina (or Tamatave) Province of Madagascar. This province is located along the east coast of Madagascar and includes the greatest extent of tropical rainforest on the island nation. The aerial extent of Toamasina Province is roughly 75,000 sq. kms. We constructed four georeferenced GIS coverages (Geographic Information System data layers) for modeling forest cover for the province: town boundaries with associated 1993 population census data, elevation, slope, and land cover. These datasets from which we derived our coverages were all downloaded off the Web. Details and sites are provided in the Appendix.

Working with multiple GIS data layers usually requires geocorrection. Such manipulation necessarily introduces measurement error into the corrected data layers. In fact, the accuracy of any data layer including the template layer (the one viewed as best, against which the others are geocorrected) is not typically quantifiable. The geocorrection process, especially rubbersheeting, involves subjectivity and also introduces unquantifiable error. Explicit assessment of error associated with any GIS geocorrection requires ground truthing, repeated sampling, and so on. General methodology is a field unto itself and would be a topic of a different article. Quantification in a particular context would obviously depend on the region and choice of data layers. Here we only note that considerable effort has been expended to provide the best visual alignment of layers and that our proposed analysis only attempts to relate the resulting variables.

Ultimately, the total number of towns in our dataset was 159 and the total number of pixels was 74,607. For analysis at a lower resolution, we aggregated each of the above final 1-km raster layers into 4-km pixels using ERDAS Imagine Decompose module. For the elevation and slope layers, we obtained average values from the composite 16 pixels for each of the new, larger pixels. For the town boundary layer, we used majority rule to assign each pixel to a town. In so doing, two towns were lost as being less than 16 km^2 in area. For the 4-km $\times$ 4-km forest raster layer, we used the number of forested subpixels $(0, 1, 2, \ldots, 16)$.

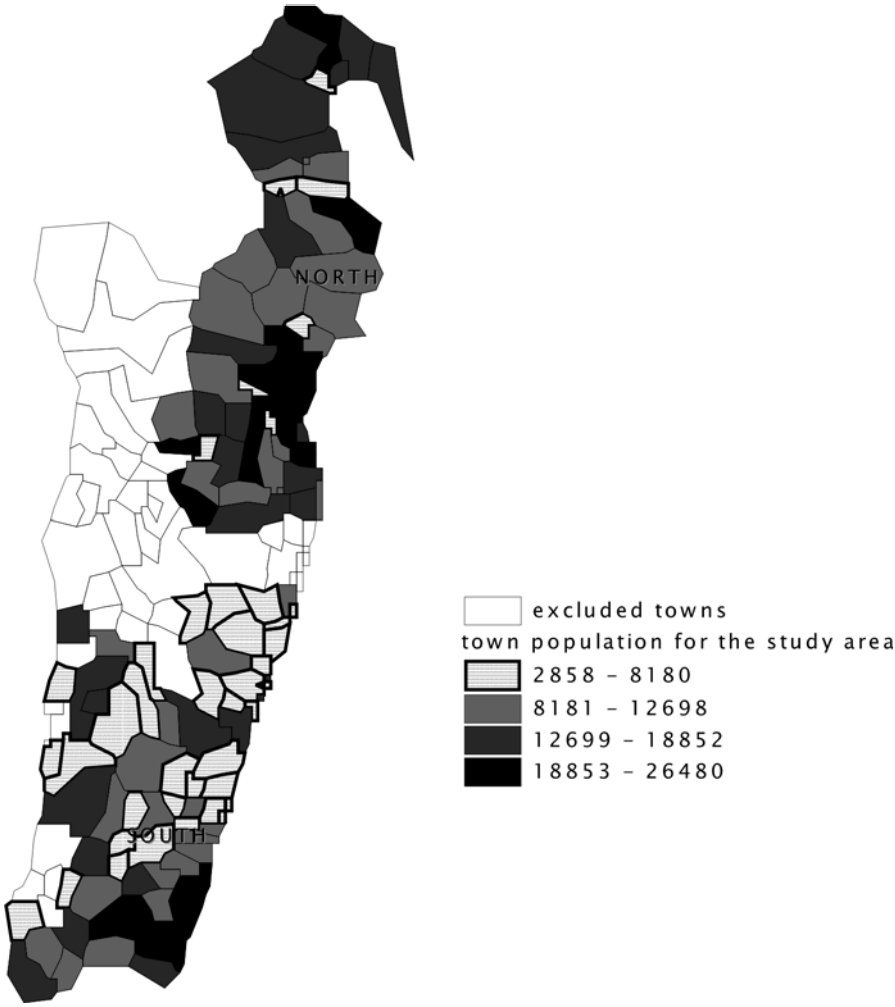

Figure 1. Northern and southern regions within the study region; gray-scale population overlaid.

Figure 1 shows the town-level map for the 159 towns in the study region. In fact, there is an escarpment in the western portion where the climate differs from the rest of the region. It is a seasonally dry grassland/savanna mosaic. Also, the northern part is expected to differ from the southern part. The north has fewer population areas with large forest patches, while the south has more villages with many smaller forest patches and more extensive road development, including commercial routes to the national capital west of the study region. (See also Figure 12 in this regard.) Hence, we excluded the western towns and introduced between north and south a transition region (to provide separation of north and south spatial effects). Finally, we arrived at the illustrative north and south regions, which are identified in Figure 1 with the excluded white areas being a combination of grassland/savanna and transition zone. In the north, there are 50 towns with total area

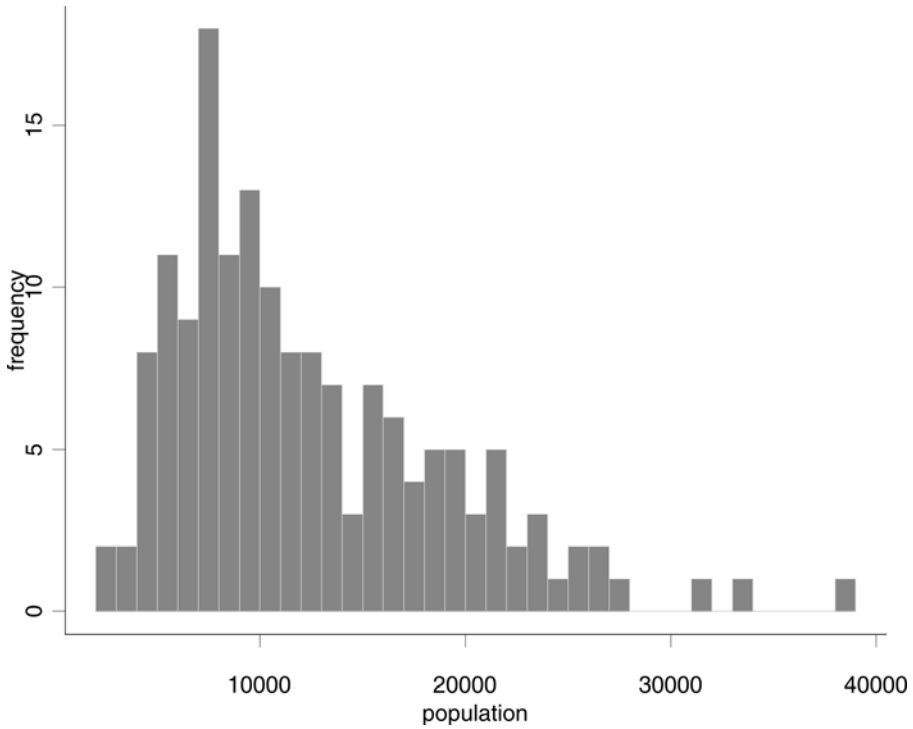

Figure 2. Histogram for population for the 159 towns in the study area.

26,432 km^2 and total population 707,786; in the south, there are 66 towns with total area 25,168 km^2 and total population 664,066. In Figure 1, gray-scale town population data is overlaid on these regions. In addition, Figure 2 provides a histogram of this population data for all 159 towns. In Section 4, we fit the models mentioned in the introduction to each region separately and then make comparisons.

Figure 3 provides the rasterized (at 1-km $\times$ 1-km resolution) north and south regions. The land-cover classification is overlaid. The proportion of forest in the north is .7055; in the south, it is .6448. At the 4-km $\times$ 4-km resolution, the distribution of the land-use variable is shown in Figure 4. Similarly, Figure 5 provides gray-scale maps for elevation for the north and south, while Figures 6 provides gray-scale maps for slope for the north and south.

While the binary map in Figure 3 shows spatial pattern in land use, we develop an additional display to provide quantification. For data on a regular grid or lattice, we calculate binary analogues of the sample autocovariances using the 1-km $\times$ 1-km resolution with four illustrative directions: east (E), northeast (NE), north (N), and northwest (NW). Relative to a given pixel, we can identify all pixels in the region in a specified direction from that pixel and associate with each a distance (Euclidean distance centroid to centroid) from the given pixel. Pairing the response at the given pixel (X) with the response at a directional

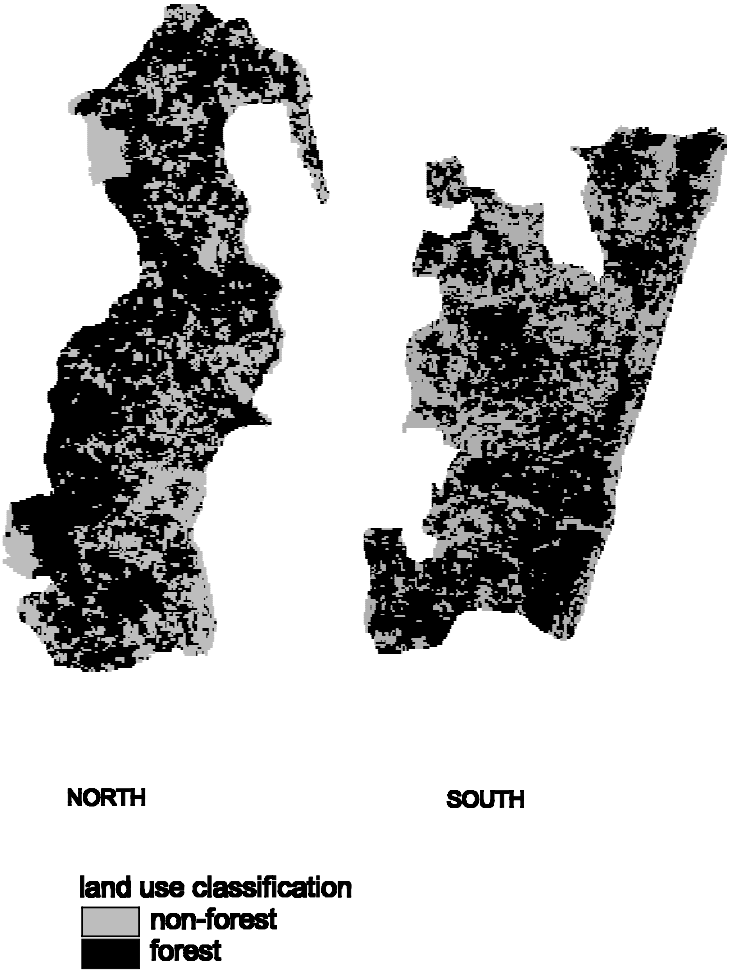

Figure 3. Rasterized north and south regions (1 km × 1 km) with binary land-use classification overlaid.

neighbor (Y), we obtain a correlated binary pair. Collecting all such (X, Y) pairs at a given direction/distance combination yields a 2×2 table of counts. The resultant log-odds ratio measures the association between pairs in that direction at that distance (note that, if we followed the same procedure but reversed direction, e.g., changed from E to W, the corresponding log-odds ratio would be unchanged).

In Figure 7, we plot log-odds ratio against direction for each of the four directions. Note that the spatial association is quite strong, requiring a distance of at least 40 km before it drops to essentially zero. This suggests that we do not lose much spatial information if we work with the lower (4-km × 4-km) resolution. In exchange, we obtain a richer response variable (17 ordered levels) and a substantial reduction in number of pixels (1,652 in the north region, 1,534 in the south region) to facilitate model fitting.

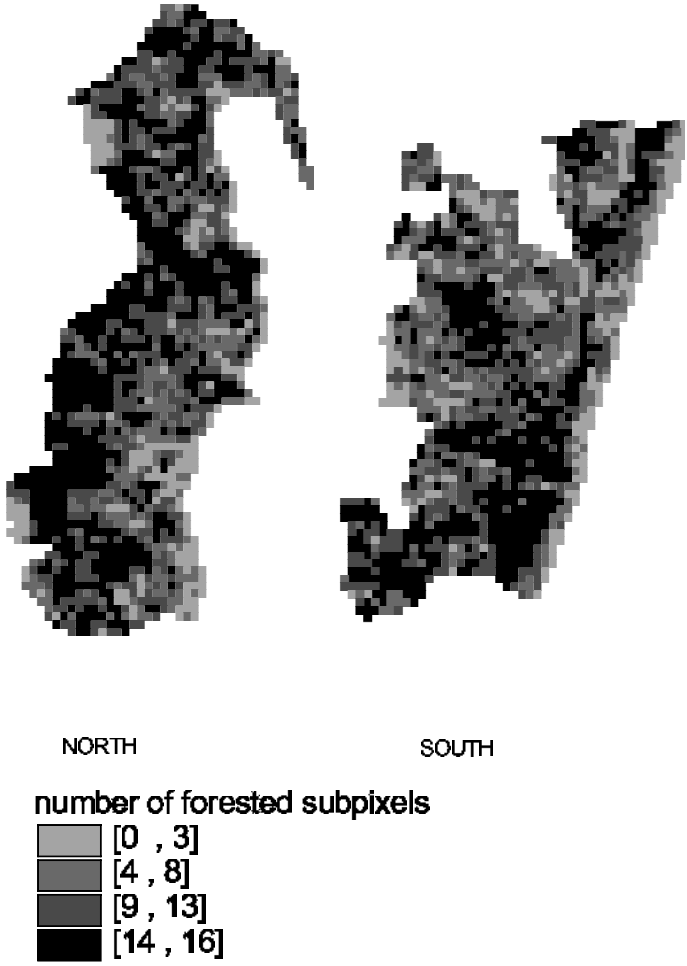

Figure 4. Distribution of number of forested pixels (out of 16) at the 4-km × 4-km resolution for the north and south regions.

3. MODELING DETAILS

We model the joint distribution of land use (L) and population count (P) at the pixel level. Let L_{ij} denote the land-use value for the jth pixel in the ith town and let P_{ij} denote the population count for the jth pixel in the ith town. Again, the L_{ij} are observed, but only $P_{i.} = \Sigma_j P_{ij}$ are observed at the town level. We collect the L_{ij} and P_{ij} into town-level vectors $\boldsymbol{L}_i$ and $\boldsymbol{P}_i$ and overall vectors $\boldsymbol{L}$ and $\boldsymbol{P}$.

As described in Section 2, we also observe at each pixel an elevation, E_{ij}, and a slope, S_{ij}. To capture spatial association between the L_{ij}, we introduce pixel-level spatial effects φ_{ij}; to capture spatial association between the $P_{i.}$, we introduce town-level spatial effects δ_i, i.e., the spatial process governing land use may differ from that for population.

We seek to specify the joint distribution, $\boldsymbol{f}(\boldsymbol{L}, \boldsymbol{P} \mid \{E_{ij}\}, \{S_{ij}\}, \{\varphi_{ij}\}, \{\delta_i\})$. We factor this joint distribution as

$$\boldsymbol{f}(\boldsymbol{P} \mid \{E_{ij}\}, \{S_{ij}\}, \{\delta_i\})\boldsymbol{f}(\boldsymbol{L} \mid \boldsymbol{P}, \{E_{ij}\}, \{S_{ij}\}, \{\varphi_{ij}\}). \tag{3.1}$$

We condition in this fashion because one of our objectives is to explain the effect of population on land use. Of course, we do not assert causality and recognize that, in a different context, the conditioning could be reversed. (Also, implicit in (3.1) is a marginal specification for $\boldsymbol{L}$ and a conditional specification for $\boldsymbol{P} \mid \boldsymbol{L}$.)

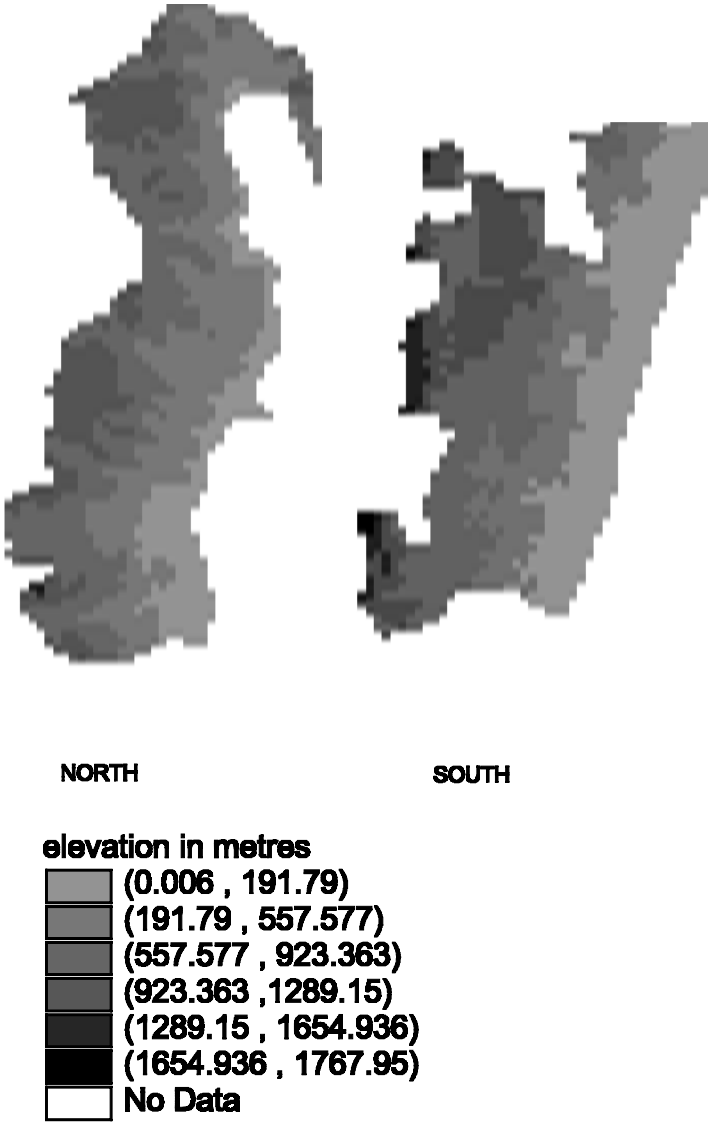

Figure 5. Gray-scale elevation maps for the north and south regions.

Turning to the first term in (3.1), we assume that the P_{ij}'s are conditionally independent given the E's, S's, and δ's. In fact, we assume $P_{ij} \sim \text{Poisson}(\lambda_{ij})$, where

$$\log \lambda_{ij} = \beta_0 + \beta_1 E_{ij} + \beta_2 S_{ij} + \delta_i. \tag{3.2}$$

As a result, $P_{i.} \sim \text{Poisson}(\lambda_{i.})$, where $\log \lambda_{i.} = \log \Sigma_j \lambda_{ij} = \log \Sigma_j \exp(\beta_0 + \beta_1 E_{ij} + \beta_2 S_{ij} + \delta_i)$. In other words, the P_{ij} inherit the spatial effect associated with $P_{i.}$. Also, $\{P_{ij}\} \mid P_{i.} \sim \text{multinomial}(P_{i.}; \{\gamma_{ij}\})$, where $\gamma_{ij} = \lambda_{ij}/\lambda_{i.}$.

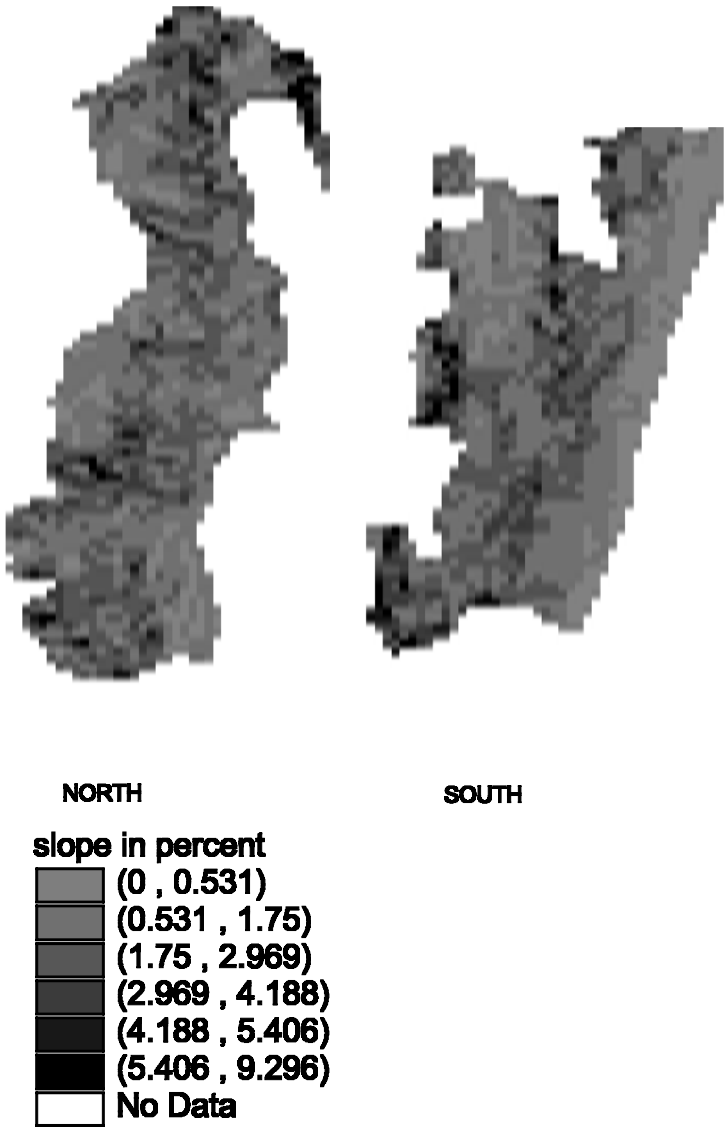

Figure 6. Gray-scale slope maps for the north and south regions.

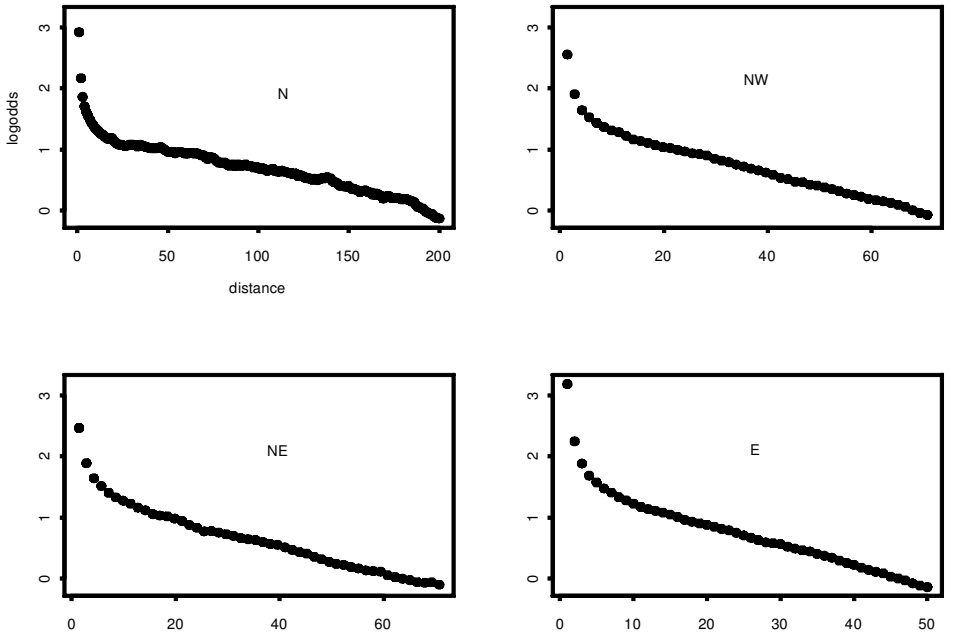

Figure 7. Land-use log-odds ratio versus distance in four directions. See text for details.

In the second term in (3.1), we assume conditional independence of the L_{ij} given the P's , E's, S's, and φ's. For the 4-km $\times$ 4-km resolution, where L_{ij} lies between 0 and 16, we assume $L_{ij} \sim$ binomial$(16, q_{ij})$, i.e., that the sixteen 1-km $\times$ 1-km pixels that comprise a given 4-km $\times$ 4-km pixel are iid Bernoulli with q_{ij} such that

$$\log\left(\frac{q_{ij}}{1-q_{ij}}\right) = \alpha_0 + \alpha_1 E_{ij} + \alpha_2 S_{ij} + \alpha_3 P_{ij} + \varphi_{ij}. \tag{3.3}$$

For the town-level spatial effects, we assume a conditionally autoregressive (CAR) prior using only the adjacent towns for the mean structure, with variance τ_δ^2, and similarly for the pixel effects using only adjacent pixels, with variance τ_φ^2 [see Besag (1974) or Cressie (1993) for further details on CAR models].

To complete the specification of the hierarchical model, we require priors for α, β, τ_δ^2, and τ_φ^2 (when the φ_{ij} are included). Under a binomial, with proper priors for τ_δ^2 and τ_φ^2, a flat prior for α and β will yield a proper posterior. For τ_δ^2 and τ_φ^2, we adopt inverse gamma priors. In particular, $\tau_\delta^2 \sim$ IG$(2, .23)$ and $\tau_\varphi^2 \sim$ IG$(2, 5.86)$. These specifications have infinite variance with mean roughly the sample variability in the log $\hat{\lambda}_i$ (where $\hat{\lambda}_i = \mathrm{P}_{i.}$) and the $\log(\hat{q}_{ij}/(1-\hat{q}_{ij}))$ (where $\hat{q}_{ij} = L_{ij}/16$), respectively. As is customary, to ensure identifiability, i.e., a well-behaved posterior distribution, we impose the constraints $\Sigma_i\, \delta_i = 0$ and $\Sigma_{ij}\, \varphi_{ij} = 0$. The model is fitted using Markov chain Monte Carlo and the constraints are implemented after each iteration (see Besag, Green, Higdon, and Mengersen 1995).

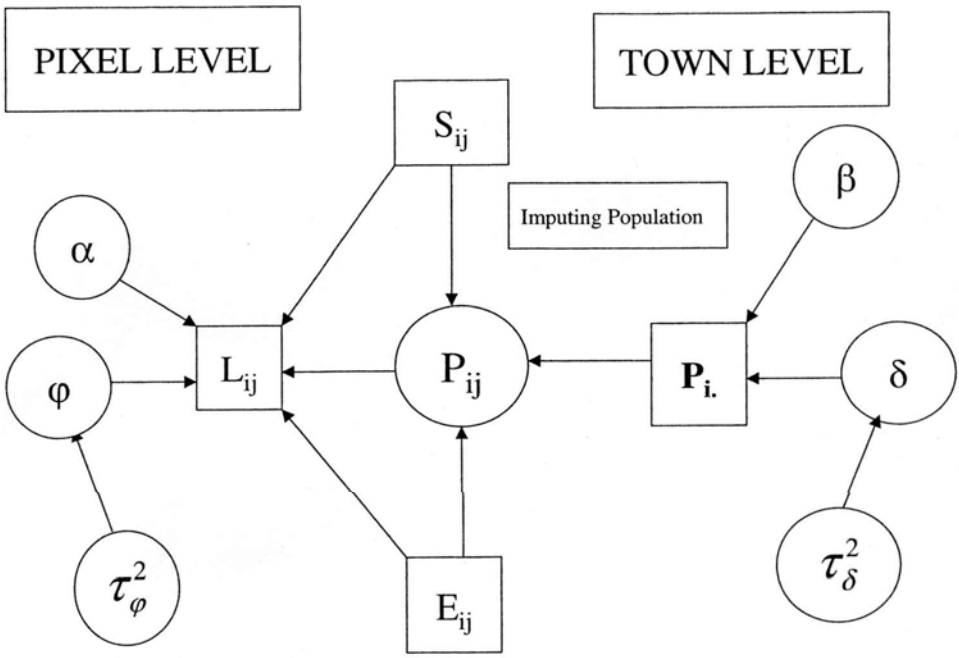

Figure 8. Graphical representation of model in (3.1)–(3.3).

4. ANALYSIS OF THE DATA

At the 4-km × 4-km pixel scale, we fitted two versions of the model in Section 3. In particular, we consider (3.3) with the φ_{ij}'s (model 2) and without them (model 1). We fitted models 1 and 2 separately for the northern and southern regions. The results are summarized in Table 1 with point (posterior median) and interval (95% equal tail) estimate. The population-count model results are little affected by the inclusion of the φ_{ij}. For the land-use model, this is not the case. Interval estimates for the fixed effects coefficients are much wider when the φ_{ij} are included. This is not surprising from the form in (3.3). Though the P_{ij} are modeled and are constrained by summation over j and though the ϕ_{ij} are modeled dependently through the CAR specification, since neither is observed, strong collinearity between the P_{ij} and ϕ_{ij} is expected, inflating the variability of the α's.

Specifically, for the population-count model in (3.2), in all cases, the elevation coefficient is significantly negative; higher elevation yields smaller expected population. Interestingly, the elevation coefficient is more negative in the north. The slope variable is intended to provide a measure of the differential in elevation between a pixel and its neighbors. However, a crude algorithm is used within the Environmental Systems Research Institute (ESRI) software for its calculation, diminishing its value as a covariate. Indeed, higher slope would typically encourage lower expected population. Although this is roughly true for the south under either model, the opposite emerges for the north. The inference for the

Table 1. Parameter Estimation (Point and Interval Estimates) for Models 1 and 2 for the Northern and Southern Regions (See Text for Details)

	Model			
	M_1		*M_2*	
Region	*North*	*South*	*North*	*South*
	Population model parameters			
β_1 (elevation)	−.577 (−.663, −.498)	−.245 (−.419, −.061)	−.592 (−.679, −.500)	−.176 (−.341, .019)
β_2 (slope)	.125 (.027, .209)	−.061 (−.212, .095)	.127 (.014, .220)	−.096 (−.270, .050)
τ_{δ^2}	1.32 (.910, 2.04)	1.67 (1.23, 2.36)	1.33 (.906, 1.94)	1.71 (1.22, 2.41)
	Land use model parameters			
α_1 (elevation)	.406 (.373, .440)	−.081 (−.109, −.053)	.490 (.160, .857)	.130 (−.327, .610)
α_2 (slope)	.015 (−.013, .047)	.157 (.129, .187)	.040 (−.085, .178)	−.011 (−.152, .117)
$\alpha_3 \times 10^{-4}$	−5.10 (−5.76, −4.43)	−3.60 (−4.27, −2.80)	−4.12 (−7.90, −.329)	−8.11 (−14.2, −3.69)
τ_{φ_2}	—	—	6.84 (6.15, 7.65)	5.85 (5.23, 6.54)

town-level spatial variance component τ_δ^2 is consistent across all models. Homogeneity of spatial variance for the population model is acceptable.

Turning to (3.3), in all cases, the coefficient for population is significantly negative. There is a strong relationship between land use and population size; increased population increases the chance of deforestation. The elevation coefficients are mixed with regard to significance. However, for both models 1 and 2, the coefficient is always at least .46 larger in the north. Elevation more strongly encourages forest cover in the north than in the south. This is consistent with the discussion of the preceding paragraph but, apparently, the effect is weaker in the presence of the population effect. Again, the slope covariate provides inconsistent results but is insignificant in the presence of spatial effects. Inference for the pixel-level spatial variance component does not criticize homogeneity across regions. Note that τ_φ^2 is significantly larger than τ_δ^2. Again, this is expected. With a model having four population parameters to explain 3,186 q_{ij}'s as opposed to a model having three population parameters to explain 115 λ_i's, we would expect much more variability in the φ_{ij}'s than in the δ_i's.

To clarify the spatial picture a bit further, Figure 9 presents gray-scale maps of the posterior means of the φ_{ij} for the north and the south. The central interval represents effects within .5 standard deviations from zero. Adjacent intervals are within .5 to 1.5 standard deviations from zero. First, note the presence of association patterns in these figures; φ_{ij}'s of similar magnitude tend to cluster. Second, $\varphi_{ij} < (>) 0$ implies that the explanatory variables E_{ij}, S_{ij}, and P_{ij} provide overestimation (underestimation) of q_{ij}. Comparison with Figure 4 shows that overestimation (underestimation) is more common when L_{ij}, hence q_{ij}, is

small (large). Similarly, Figure 10 presents gray-scale maps of the posterior means of the δ_i for the north and the south. Again, there is presence of an association pattern in the δ_i's, but connection of their magnitudes with associated population sizes is weak.

Last, we examine the implicit imputation of population from town to pixel level. At the 4-km $\times$ 4-km resolution, in Figure 11, we provide a gray-scale rasterized imputed population map on the square-root scale for the north and the south. These maps were developed by obtaining the posterior means of the P_{ij}'s under model 2 and then converting

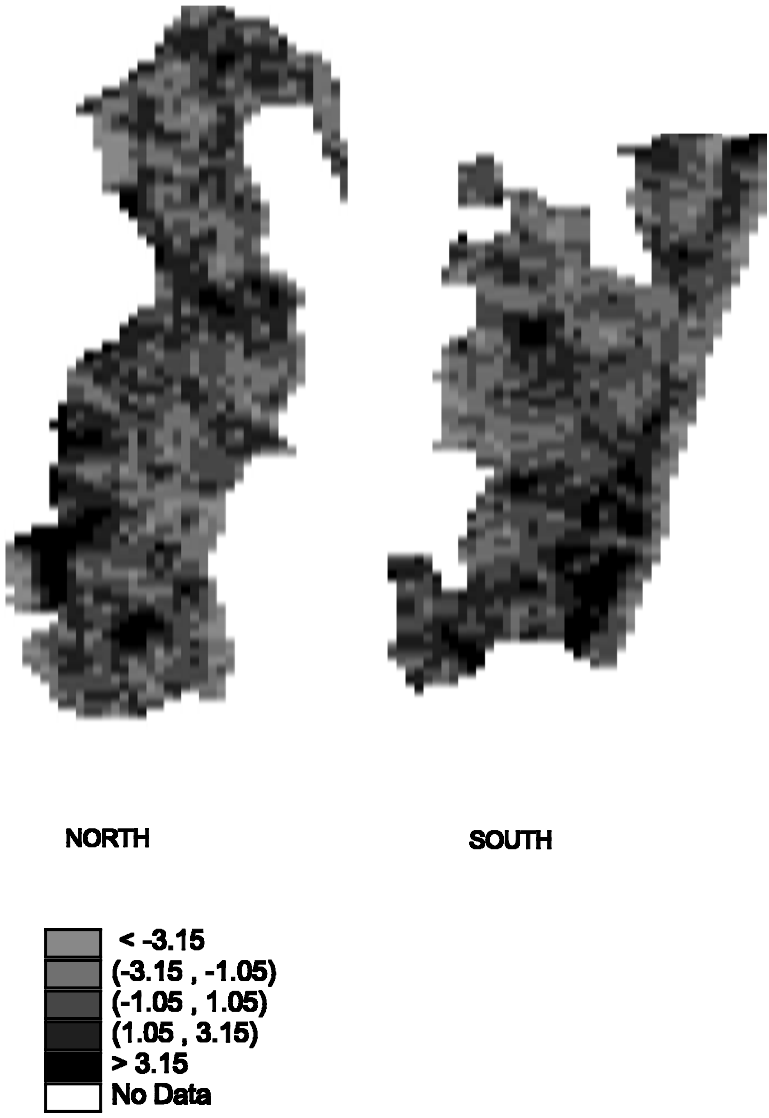

Figure 9. Land-use model spatial random effects.

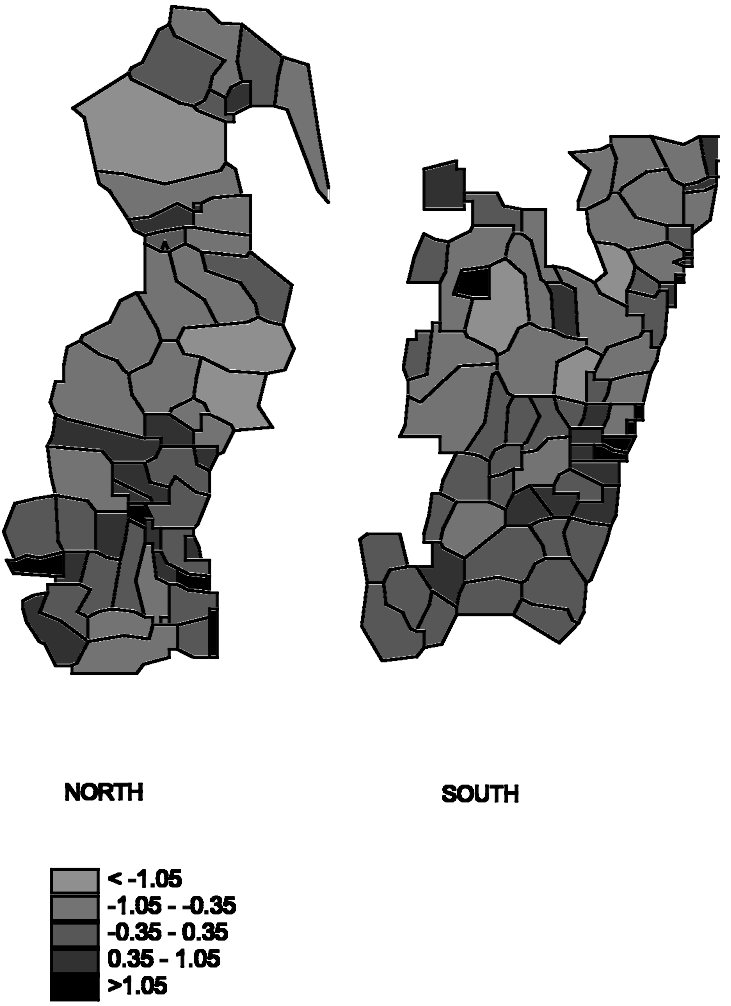

Figure 10. Population model spatial random effects.

to a gray scale. As a form of informal model checking, we obtained Figure 12 to compare with Figure 11. Figure 12 (not developed by us) offers crude spatial information about the distribution or clustering of populations within towns. In particular, villages are denoted by dots, with dot size reflecting village size. The distribution of villages varies greatly across towns, the population ranges associated with dot sizes are not necessarily good choices, and there is, of course, population apart from that in the villages. Nonetheless, our imputed populations do match reasonably well.

5. EXTENSIONS

The foregoing analysis is still rather preliminary, as the dataset is in an ongoing process of refinement. Here we mention some additional work we plan to do and how it can be carried

out within the modeling framework described in Section 3.

First, we are refining the land-use classification. We will eventually obtain five ordinal classifications ranging from severely degraded ($L = 1$) to pristine forest ($L = 5$) . Following the ideas in Albert and Chib (1993), we can introduce a latent variable W_{ij} associated with each L_{ij}. W_{ij} is a conceptual continuous random variable on R^1, interpreted as (transformed) extent of forest cover. Then $L_{ij} = l$ if $W_{ij} \in (\gamma_{l-1}, \gamma_l)$, $l = 1, \ldots, 5$, where $\gamma_0 = -\infty, \gamma_5 = \infty$. For identifiability, we can set the cut point $\gamma_1 = 0$.

Second, we wish to introduce additional explanatory variables. Here there are two possibilities. In one case, the covariate may be definable at the pixel level. For instance,

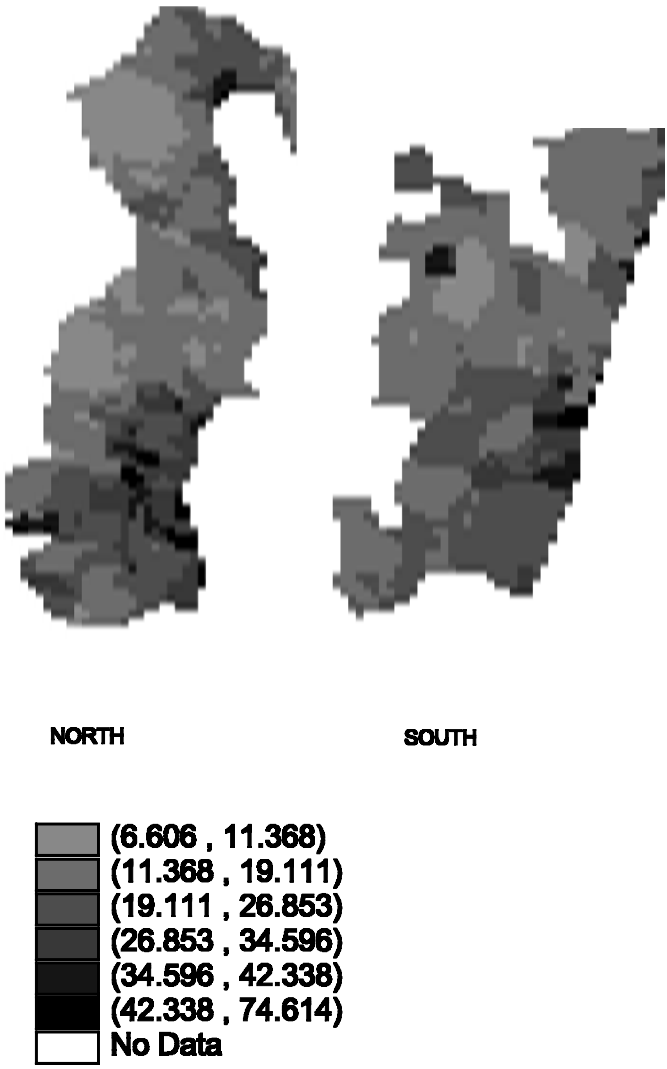

Figure 11. Imputed population (on the square-root scale) for the pixel level for north and south regions.

Figure 12. Population distribution for the north and south regions by village. Dots denote villages; dot size reflects village size.

roads and rivers might be expected to influence land use. A vectorized map of the road network for a region can be converted to a pixel-level road classification in various ways. Our approach will first define an ordinal classification for roads ranging from an ox cart path to a main highway. Then the level assigned to a pixel will be the classification of the most developed road found in the pixel. A similar approach can handle hydrology.

In the other case, the covariate may introduce an additional misaligned data layer. This will be the situation for various socioeconomic and historical variables that may also be of interest with regard to land use. For such variables, again, we propose rasterization of the associated areal units to the pixel level, introducing analogues of (3.2) into the modeling. Careful attention to detail will be critical here since a given pixel will have a different labeling under each rasterization. Look-up tables will need to be created to align the labelings. Edge effects, i.e., pixels in one set of areal units but not in another, must be handled as well [see Mugglin, Carlin, and Gelfand (2000) for related discussion].

APPENDIX: DESCRIPTION OF DATA LAYERS

The town boundary maps and census information were obtained from the United Nations Statistics Division (UNSD) via their software package PopMap for Windows 4.1, which can be downloaded directly from the Web http://www.undp.org/popin/softproj/

software/popmap.htm. The census data were compiled by Direction de la Demographie et Statistiques Sociales of Madagascar and were made available together with town boundary coverages for all of Madagascar by the United Nations Population Fund (UNFPA), the United Nations Department for Economical and Social Affairs (UNDESA), and UNSD. The GIS coverage of towns (firaisana) and census data for Toamasina (160 towns in total) were exported from PopMap as a vector shapefile in geographic projection in decimal degrees, World Geodetic System 1984 (www.wgs84.com).

Elevation data for Madagascar were obtained from the USGS Global 30 Arc Second Elevation Data Set for the world http://edcwww.cr.usgs.gov/landdaac/gtopo30/gtopo30.html from tile E020S10. The sources of the elevation data for Madagascar are both the Digital Chart of the World (DCW) and the Digital Terrain Elevation Data (DTED). Details of the meta-data are provided at this Web site. Unfortunately, these two different data sources, tiled together by USGS for the coverage of Madagascar, have different levels of accuracy. The data were downloaded as a raster grid file in geographic projection as decimal degrees, World Geodetic System 1984, with a pixel resolution of 1 km and elevation expressed to the nearest meter for each pixel. Slope was subsequently derived from this data layer as described below.

Vegetation cover for Madagascar was obtained from the USGS Global Land Cover Characterization for Africa http://edcwww.cr.usgs.gov/landdaac/glcc/af_int.html. This site provides a seasonal land-cover interpretation for 197 cover types for all of Africa with nominal 1-km resolution. The land-cover classification was based on 1-km AVHRR NDVI monthly composite images spanning April 1992 through March 1993. Data were downloaded as a raster image file.

A fourth GIS coverage was used as a template to which the other data layers were geocorrected. This is the Primary Vegetation map of Madagascar, a vector coverage that has been reproduced and is available from several different sources. We downloaded this coverage from the Royal Botanical Gardens, Kew web site http://www.rbgkew.org.uk/herbarium/madagascar/veg_mapping.html. Based on available GIS map layers for Madagascar, we judged this to provide the most accurate outline projection coverage, more accurate that the town-level vector coverage. We downloaded this as a vector shapefile, geographic projection in decimal degrees, World Geodetic System 1984.

The town-level vector coverage was geocorrected to the Primary Vegetation vector coverage in IDRISI (Version 2) using at least 20 control points. The raster elevation and landcover GIS layers were cut to a rectangular area that includes Toamasina Province and the towns it contains. These raster layers were reprojected and geocorrected to the Primary Vegetation vector coverage in ERDAS Imagine (Version 8.1). Slope was obtained from the raster grid elevation coverage directly via the Spatial Analyst module of ESRI ArcView 3.1. The landcover raster layer was reclassified to forest and nonforest. This was achieved by lumping all rainforest classes into a single forest class and all other classes, including degraded rainforest, woodlands, savanna, cropland, etc., as nonforest. The town-level vector coverage was converted to a raster 1-km grid coverage in ArcView. In so doing, one town

was deleted from the data set as a consequence of being less than 1 km^2. The total number of towns was thus 159. All layers were exported to ERDAS Imagine converted to UTM zone 30, spheroid projection, Clark 1866 datum, in meter units. In the final raster layers, we deleted from the data set all coastal islands. These data layers were subsequently exported as a single ASCII file, with pixel coordinates (center) in UTM meters, and associated with each pixel town identity, 1993 population size, elevation (m), slope (%), and landcover (1 = forest, 0 = not forest). The total number of pixels in this file was 74,607.

ACKNOWLEDGMENTS

The work of the second author was supported in part by NSF-DMS 99-71206. The authors thank Robert Dewar for helpful discussions and for assistance with Figure 10 and Eric Ribaira, USAID, Antananarivo, for helping obtain the population and boundary data. The authors thank the referees, associate editor, and editor for their comments and suggestions, which improved the quality of the article.

[Received March 2000. Accepted October 2001.]

REFERENCES

Albert, J., and Chib, S. (1993), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679.

Besag, J. (1974), "Spatial Interaction and the Analysis of Lattice Systems" (with discussion), *Journal of the Royal Statistical Society*, Series B, 36, 192–236.

Besag, J., Green, P. J., Higdon, D., and Mengersen, K. (1995), "Bayesian Computation and Stochastic Systems" (with discussion), *Statistical Science*, 10, 3–66.

Brown, K., and Pearce, D. W. (eds.) (1994), *The Causes of Tropical Deforestation*, Vancouver: University of British Columbia Press.

Chomitz, K. M., and Gray, D. A. (1996), "Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize," *World Bank Economic Review*, 10, 487–512.

Cressie, N. A. C. (1993), *Statistics for Spatial Data* (rev. ed.), New York: Wiley.

Davies, S. D., Heywood, V. H., and Hamilton, A. C. (eds.) (1994), *Centers of Plant Diversity: A Guide and Strategy for Their Conservation* (Vol 1), *Europe, Africa, Southwest Asia and the Middle East*, Cambridge, U.K.: International Union for the Conservation of Nature (IUCN).

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Granger, A. (1998), "Modeling Tropical Land Use Change and Deforestation," in *Tropical Rain Forest: A Wider Perspective*, London: Chapman and Hall.

Green, G. M., and Sussman, R. W. (1990), "Deforestation History of the Eastern Rainforests of Madagascar From Satellite Images," *Science*, 248, 212–215.

Leach, G., and Mearns, R. (1988), *Beyond the Fuelwood Crisis: People, Land and Trees in Africa*, London: Earthscan.

Lowry, P. P., Schatz, G. E., and Phillipson, P. B. (1997), "The Classification of Natural and Anthropogenic Vegetation in Madagascar," in *Natural Change and Human Impact in Madagascar*, eds. S. M. Goodman and B. D. Patterson, Washington, DC: Smithsonian Institute Press, pp. 93–123.

Mercier, J.-R. (1991), *La Deforestation en Afrique: Situation et Perspectives.* Aix-en-Provence, France: Edisud.

Meyers, N. (1991), "The Biodiversity Challenge: Expanding Hotspots Analysis," *Environmentalist*, 10, 243–256.

Mittermeier, R. (1988), "Primate Diversity and the Tropical Forest: Case Studies From Brazil and Madagascar and the Importance of the Megadiversity Country," in *Biodiversity*, ed. E. O. Wilson, Washington, DC: National Academy, pp. 145–154.

Mugglin, A. S., and Carlin, B. P. (1998), "Hierarchical Modeling in Geographic Information Systems: Population Interpolation Over Incompatible Zones," *Journal of Agricultural, Biological, and Environmental Statistics*, 3, 111–130.

Mugglin, A. S., Carlin, B. P., and Gelfand, A. E. (2000), "Fully Model Based Approaches for Spatially Misaligned Data," *Journal of the American Statistical Association*, 95, 877–887.

Olson, S. H. (1984), "The Robes of the Ancestors: Forests in the History of Madagascar," *Journal of Forest History*, 28, 174–186.

Palloni, A. (1992), "The Relation Between Population and Deforestation: Methods for Drawing Causal Inferences From Macro and Micro Studies," CDE Working Paper 92-14, Center for Demography and Ecology: University of Wisconsin, Madison.

Richards, J. F., and Tucker, R. P. (1988), *World Deforestation in the Twentieth Century*, Durham, NC: Duke University Press.

Sponsel, L. E., Headland, T. N., and Bailey, R. C. (1996), *Tropical Deforestation: The Human Dimension*, New York: Columbia University Press.

Torsten, A. (1992), *Deforestation of Tropical Rainforests: Economic Causes and Impact on Development*, Tubingen: J.C.B. Mohr.

Whitmore, T. C. (1992), *Tropical Deforestation and Species Extinction*, New York: Chapman and Hall.