# Do roads cause deforestation? Using satellite images in econometric analysis of land use

Nelson, Gerald and Hellerstein, Daniel

Economic Research Service, USDA

February 1997

# Do Roads Cause Deforestation?

## Using Satellite Images in Econometric Analysis of Land Use

Gerald C. Nelson and Daniel Hellerstein

June 1995

(Revised May 1996)

---

Gerald C. Nelson, Associate Professor, University of Illinonis, Department of Agricultural and Consumer Economics.

Daniel Hellerstein, Natural Resource Economist, Economic Research Service, USDA.

Support for this paper was provided by Cooperative Agreement 43-3AEM-3-80137 between USDA and the University of Illinois. It does not reflect the views of either institution. This draft has benefitted from discussions with Ken Chomitz, Peter Feather, Charles Hallahan, Ralph Heimlich, Laurian Unnevehr, and participants in seminars at IFPRI, the World Bank, Resources for the Future, the University of Illinois, Universidad Nacional Autonoma de Mexico, Centro de Investigación y Docencia Económicas, and Secretaría de Medio Ambiente, Recursos Naturales y Pesca, Mexico. A more detailed version of this paper, with color illustrations, is available on the web at http://w3.aces.uiuc.edu/ACE/faculty/GNelson/papers/lu_paper/description.htm.

# Do Roads Cause Deforestation?

## Using Satellite Images in Econometric Analysis of Land Use

**Gerald C. Nelson and Daniel Hellerstein**

**Abstract**

This paper demonstrates how satellite images and other geographic data can be used to predict land use. A cross-section model of land use is estimated with data for a region in central Mexico. Parameters from the model are used to examine the effects of reduced human activity. If variables that proxy human influence are changed to reflect reduced impact, "forest" area increases and "irrigated crop" area is reduced.

Key words: geographic data, land use, remotely sensed data, spatial econometrics

## Do Roads Cause Deforestation?

## Using Satellite Images in Econometric Analysis of Land Use

It has been alleged that building new roads in developing countries exacerbates the rate of deforestation by making access to forested areas less costly. Anecdotal evidence is sometimes presented to support this allegation. For example, as the Brazilian road network expanded into the Amazon rain forest, deforestation increased dramatically. However, more general empirical evidence of this effect is scarce because data on changes in forest area are notoriously bad or lacking.

Underlying this concern is an implicit model that relates the category of use of a particular parcel (e.g., forest, crop, desert, urban) to the parcel's resource endowment and location. This paper presents a conceptual model of the determinants of land use with roots in the work of von Thunen and based on a model developed in Chomitz and Gray. To address the data constraint, the paper also shows how data on forest areas, and more generally on land use, can be extracted from satellite images and combined with other geographic information to estimate the model. The empirical results show how the effects of changes in human activity can be simulated.

Satellite images greatly increase the resolution of spatial data on land use, but provide no new data on socioeconomic variables. The challenge is to develop empirical models that incorporate rich spatial geophysical data (observations every 30-150 meters square; file sizes of up to 80 megabytes per variable) with sparse socioeconomic data. This paper provides such a model and econometric estimates for a location in central Mexico.

**A Model of the Determinants of Land Use**

A given parcel of land has a variety of potentially productive resources associated with it. These resources might be vegetative (timber, productive soil), mineral, or locational (e.g., proximity to a harbor). We assume these resources will be used if it is profitable to do so. The choice of a particular land use is made by comparing the net present value of the profitability of all possible land uses. If we assume that a given land use has a single marketed product, the net present value of the return to that land use, its net present rent ($R_{hl}$) at time T, is given by:[1]

(1) **Error!**

where P is the output price, Q is the quantity of output, **C** is a vector of input costs, **X** is a vector of inputs under operator control and i is the discount rate, all for each land use *h* at location *l* at time *t*. The land use chosen for the parcel has the highest $R_{hlT}$.

We assume a Cobb-Douglas production function with an index of parcel-specific geophysical factors (G) such as soil type and elevation that affect productivity:

$$(2) \qquad Q=G_{1ji}\prod X_i^{\alpha_i} \qquad 0<\alpha_i<1;\ 0<\sum_i \alpha_i<1$$

For this analysis, we have only cross-section data, meaning that we have no data on changes in prices or temporal changes in land use. However, we observe changes in land use across space. To take advantage of this fact, we construct proxies for location-specific P and C, following Chomitz and Gray, based on cost-of-access measures. These proxies are:

$$(3) \qquad \begin{aligned} P_{hl} &= \exp\left[\gamma_{0h}+\gamma_{1hl}D_{hl}\right] \\ C_{hl} &= \exp\left[\delta_{0h}+\delta_{1hl}D_{hl}\right] \end{aligned}$$

$D_{hl}$ - a measure of least cost access to parcel *h* for land use *l*[2].

We expect that location-specific output price falls as access cost increases ($\gamma_{1hl} < 0$) and location-specific input cost rises ($\delta_{1hl} > 0$) for land uses with marketed output (see Chomitz and Gray for further discussion). However, for some land uses with no marketed output (for example, idle land) the effect of cost of access might be zero.

Combining (1) to (3), taking advantage of the fact that most terms are not time-dependent, and adding an error term, we arrive at equation (4)[3]:

$$(4) \qquad \ln\left(R_{hlt}(\_\_)\right) = \eta_{0h} + \sum_i \eta_{1i} D_i + \eta_{2h} \ln G_h + \eta_{3h} \ln i_l + u_{hl} = V_h N_l + u_{hl}$$

Given standard hypotheses about the underlying parameters ($0 < \alpha < 1$, $\gamma < 0$, $\delta > 0$) we expect $\boldsymbol{\eta_0}$ and the $\boldsymbol{\eta_1}$ vector are negative, $\eta_2$ to be positive for G ordered so that an increase in G increases output per unit of land, and $\eta_3$ - 1[4]. Parcel h will be devoted to land use k if $R_{hkT} > R_{hlT}$, $\forall\, l \neq k$. If the u are Weibull distributed and uncorrelated across land uses, then (4) is equivalent to a multinomial logit model where:

$$(5) \qquad Prob_{hl} = \frac{e^{V_l B_h}}{\sum_j e^{V_l B_j}}$$

For estimation, the **V** vector consists of three sets of variables; G – site-specific geophysical variables, C – cost-of-access variables, and S – spatial effects geophysical variables.

**Data Sources and Data Manipulation**

The land use data in this paper are derived from a satellite image of Central Mexico taken by the multispectral sensor (MSS) of Landsat satellite 3 on 4/21/73[5]. The region is characterized by high plains and even higher mountains. The average elevation is 2,350 meters and the highest point is 3,850 meters.

At the time the image was taken, only irrigated fields had any crop cover. Much of the land is devoted to dryland agriculture and it is quite common to see a small field on a hillside next to an uncultivated area of scrub brush and stones. Parcels on hillsides are less likely to be cultivated. Goat grazing and fuel wood gathering are common in the hills. In the ejido land tenure system in Mexico, cultivated areas are associated with a specific individual, while grazing and forest areas are more likely to be communal lands with relatively open access allowed for members of the local ejido.

For this analysis, land use data collection at the location of the image was not possible. Hence, land use must be identified from satellite images, other digital data, and qualitative analysis based on a brief visit to the location. Inspection of the false color version of the image (available in the web version of the paper) suggests that most irrigated cropped land is located in the northwest corner of the image (crop vegetation is a lighter red in a false color image). Forests (darker red in a false color image) are located primarily on the tops of high mountains and ridges, especially in the south. The location of several lakes is also clear. It is not possible to associate land use with other parts of the image with any degree of confidence.

The IDRISI software's unsupervised cluster analysis routine (IDRISI Technical Reference, Richards) was used to identify seven land use categories (Figure 1)[6]. Based on

standard remote sensing color interpretations and some knowledge of the region, we infer that the land use associated with cluster (category) 6 is primarily forest; cluster 7 is primarily irrigated cropped land[7]. Visual inspection plus knowledge of the region suggests that uncultivated land on hillsides is often in category 4 and cultivated dryland area is in category 2.

We use six geophysical and three socioeconomic variables. The geophysical variables are elevation, slope, soil, potential intensity of solar radiation, a dummy equal to 1 for north-facing slopes, and a dummy equal to 1 for flat pixels. A seventh geophysical variable designed to capture spatial effects is discussed below. The socioeconomic variables are cost of access to nearest road, large population center, and village.

Slope and aspect can be computed from elevation data provided in the NALC data set. Because this part of Mexico is north of the equator, north-facing slopes receive less solar radiation than south-facing slopes. Therefore, we converted the aspect data into a dummy variable for north-facing slopes. Slope and elevation data can be combined with information on the sun's elevation and azimuth to calculate the intensity of solar radiation on each location. This "shade" variable has a range from 0.3 to 1.3.

`The raw soils data (Bliss) were transformed into a discrete index (1 to 9) with one for least productive soils for annual crop production and nine for most productive soils (Eswaran)[8].

Three variables were included to measure different kinds of access – to nearest road, to nearest large population center and to nearest small population center or village. The locations of roads and population centers were derived from vectors in the Digital Chart of the World (Defense Mapping Agency). The IDRISI Costgrow module calculates the least cost route to the nearest feature (e.g., road) by traversing a friction surface. We constructed the friction surface

from the square root of the slope (in degrees); the larger the square root of the slope, the greater the cost of traversing a location[9]. Rivers and lakes were added as barriers in the friction surface. The road cost-of-access variable is most relevant for mechanical transport activities, say by truck. The village cost-of-access variable is most relevant for activities done by foot or animal traction. Examples include collecting fuel wood and cultivating small plots. In addition to the cost variables, we also tried the model with Euclidean distance. We report overall results for the two approaches.

**Spatial Econometric Issues**

An important econometric issue is how to deal with what Anselin terms "spatial effects" – spatial autocorrelation and spatial heterogeneity. Spatial autocorrelation is similar to time-wise dependence, but is complicated by the two dimensional and multidirectional nature of dependence in space. Spatial heterogeneity "is related to the lack of stability over space of behavioral or other relationships under study." (Anselin, p. 9)[10]

Both heteroskedastic errors and spatial dependence in the error term cause inefficient but asymptotically unbiased estimates of ß. Since the data set used in estimation has over 26,000 observations, the potential for these effects is ignored in this paper. However, the inclusion of spatially lagged variables can lead both to inefficiency and biased parameter estimates, because the lagged dependent variable is correlated with the error term.

The above discussion is based on analyses with a continuous left-hand-side (LHS) variable. The econometric theory of spatial effects with a qualitative LHS variable is still in its infancy. However, we expect a potential for biased results if we do not correct for spatial dependence. Moran's I is sometimes used to detect the existence of spatial effects (see Anselin,

p. 101). It is similar to a correlation coefficient with a range of -1 to 1 if W is row-standardized.

Using a queen's case weight matrix (eight neighbors are included), I is 0.910 for the full image (with over 2.5 million locations). For the sample used in estimation, I is 0.472. Whatever significance test is used, these values are significantly greater than zero, suggesting the likelihood of spatial dependence.

To correct for this potential bias, we combine two approaches. First, following Haining (pp. 131-133) we use a coding scheme, selecting a sample from the full data set so no two sites in the sample are neighbors (neighboring pixels in the sample are 10 pixels apart in the full data set). Second, we construct a variable based on neighbors in the original data set. Since the dependent variable is qualitative, we use a normalized vegetative index (NDVI) value for each neighbor. This approach assumes that the spatial effects are associated with vegetative cover. The proxy lag variable is defined as

(6)

$$Ylag_{0,0} = \frac{1}{8}\left(\left(\sum_{j=-1}^{1}\sum_{i=-1}^{1} NDVI_{i,j}\right) - NDVI_{0,0}\right)$$

$$NDVI_{i,j} \equiv \frac{B4_{i,j} - B2_{i,j}}{B4_{i,j} + B2_{i,j}}$$

$Bn_{i,j}$ *intensity value of band n at relative point i,j*

**Results**

In the first part of this section, we discuss predictive power, using alternate approaches. In the second, we simulate the effects on land use of changing access cost.

*Approaches to assessing predictive power*

Table 1 reports values for three pseudo $R^2$s for four permutations of the model. The first permutation, using access cost rather than euclidean distance and unlogged values of all geophysical variables, has the highest pseudo $R^2$ value and the smallest condition numbers (a measure of multicollinearity; values greater than 30 suggest existence of multicollinearity).[12] The remaining analysis is based on this permutation.

As Greene (1991, p. 477) points out, the multinomial logit model proliferates betas. Since we have seven land use categories and 11 right-hand side (RHS) variables, there are 77 betas. However, the multinomial logit model is underidentified unless the betas for one of the categories is known. It is standard practice to choose one category as the base and set its betas to zero (Greene, 1993, p. 666). This means that the estimated betas are relative to the betas in the base category. Hence it is not possible to give economic meaning to the estimated betas unless prior information is available about the betas of the base category. For example, we expect the true beta for the price of the output to be positive (an increase in the output price increases rent). However, if the beta for the base category is larger than the beta for another category, the estimated beta is negative.

We also estimate the marginal betas, that is, the effect of a one unit change in a RHS variable on the probability that a location is in a given category, and their asymptotic significance (see Greene, 1991, p. 478). Since the predicted probabilities are functions of the RHS variables, so are the marginal betas. Table 2 reports the marginal betas for categories 6 and 7, evaluated at the mean of the RHS variables for those categories (rather than the mean for all categories as is commonly done). In other words, we ask the hypothetical question for a

location with geophysical and socioeconomic characteristics equal to the mean of those for the forest (irrigated crop) locations: what is the effect of a change in the RHS variables on the probability that this location is forest (irrigated crop) area? For the most part, the effects on forest and irrigated crop areas are as expected. An increase in the cost of access to roads (RDCost) and villages (PNTCost) reduces the probability that a location with irrigated crops will remain cropped and increases the probability that a forested location will remain forested. The effect of access to large population centers is ambiguous in sign but does not differ significantly from zero for either category.

Another assessment of predictive power can be derived from the prediction matrix. Each location is assigned a predicted category (the category with the highest predicted probability). The matrix rows show the number of locations *actually* in a given category; its columns show the number of locations *predicted* to be in a given category. Diagonal elements are correct predictions. The prediction matrix is similar to a confusion matrix in the remote sensing literature which compares categories identified using a classification scheme with categories identified by ground observation (Richards, p. 272).

Several measures of classification accuracy derived from the confusion matrix have been proposed in this literature (Congalton, 1991 as reported in Richards, p. 272). We report four here – average, producer, and user accuracy, and kappa (κ) (table 3). "Average" accuracy ($A^a$) is the sum of the diagonals divided by the number of observations.The next two assessment statistics are for individual categories. "Producer" accuracy ($A^p$) (Congalton, 1991) is the probability that the actual value is predicted correctly and is the diagonal element of a row divided by the sum of the row values. "User" accuracy ($A^u$) is the probability that a prediction

is correct and is the diagonal element of a column divided by the sum of the column values. $A^a$ is essentially a weighted average of the two measures.

(7) $$A^a=\frac{\sum_{j=1}^{n} a_{jj}}{N};\ N=\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}$$

(8) $$A_i^p=\frac{a_{ii}}{\sum_{j=1}^{n} a_{ij}};\ A_i^u=\frac{a_{ii}}{\sum_{i=1}^{n} a_{ij}}$$

The final measure of accuracy from the remote sensing literature is kappa (κ):

(9) $$\kappa=\frac{N\sum_k x_{kk}-\sum_k x_{k+}x_{+k}}{N^2-\sum_k x_{k+}x_{+k}}$$

$x_{i+}=\sum_j x_{ij}$ (*i.e., the all columns for row i*)

$x_{+j}=\sum_i x_{ij}$ (*i.e., the all rows for column j*)

The overall indicators of accuracy (average and κ) are similar in value to the pseudo $R^2$ statistics reported earlier. The producer and user accuracy measures suggest that the model is much better at predicting some categories than others. The best performance is for categories six (forest) and seven (irrigated crop).

A drawback of the prediction matrix and related indicators of accuracy is that they do not show the "strength", or power, of the predictions. A pixel can be assigned to a category with

low predicted probability, if the probability values for the other categories are even smaller. To assess predictive power, two statistics based on probability values – the maximum value and the correlation with other probability values – can be used.

The highest probability value for a land use category gives an indication of the strength of the model in identifying locations in that category (table 4). The maximum value for category 1 is 0.657, and for categories 2 and 3 the highest values are less than 0.6. Categories 6 and 7 have high maximum values (larger than 0.99), and the high skewness and kurtosis values suggest a bimodal distribution of probabilities.

A correlation matrix for the probability values can be used to indicate the model's discriminatory power. If the classification probabilities of two categories have a large positive correlation coefficient across all data points, the model discriminates poorly between them. Large negative values suggest strong discrimination power. The results suggest that the model does not discriminate well between categories 1 and 3, 2 and 4, and 3 and 5 (correlation coefficients of 0.42, 0.59, and 0.69). However, categories six and seven are strongly identified (large negative correlation coefficients except with each other).

A qualitative approach to assessing predictive strength is to plot the probability values for each category as an image. Figure 2 shows the probability images for land use categories 1, 2, and 3 in the top row and categories 5, 6, and 7 in the bottom row. A probability of zero is black, a probability of one is white, and shades of grey indicate intermediate probabilities. It is evident from this image that the model gives especially strong predictions for locations in categories 6 and 7. Where a location is predicted to be in these categories, the probability value is high (close to pure white).

To summarize, whether using the pseudo $R^2$ measures, the prediction matrix and derivative statistics, or the strength of prediction measures using the probability values, the model's overall predictive power is moderately good. In addition, at a disaggregated level, the model predicts categories 6 and 7 well.

*Effects of increasing access cost*

The econometric results can be used to simulate the effect on land use of changing access costs. For illustration, we increase the cost-of-access variables at each location by the maximum in the sample (and two times maximum for cost of access to the nearest road) and reestimate the probability values and predicted values at each location (table 5)[13]. Figure 3 displays probability images for six land use categories (categories 1 to 3 and 5 to 7) when the cost of access to the nearest road is increased by the sample maximum.

We highlight the results that affect categories 6 and 7, the areas identified as forest and irrigated crops. For an average pixel in the whole sample, changing the road access cost increases the probability that it is forested (table 5). Using the prediction matrix approach to allocate pixels to categories, the change in road access increases forest pixels by 20% and decreases irrigated crop area by 16%. Increasing the access cost to large population centers has a similar but smaller effect on forest area. The effect on irrigated area is positive; the probability that an average pixel is irrigated goes up slightly as does the predicted number of irrigated pixels.

Increasing the cost of access to the nearest village has apparent perverse effects on forest area. For the average pixel in the whole sample, the increased cost actually *reduces* the probability that it is forested. However, if we examine just category 6 (forest) pixels, the

changed access to village has almost no effect on the probability of being forested (the probability drops from 0.681 to 0.671) and the number of predicted pixels drops by less than 9 percent. Furthermore, the cost-of-access change increases the probability that a cropped pixel will become forested (from 0.350 to 0.423). The predicted number of crop pixels drops by almost 50 percent. The most dramatic effect of increasing the cost of village access is in category 4 (table 5). The probability that a pixel is in this category, tentatively identified as uncultivated hillsides and scrub areas, increases dramatically for all categories. The number of pixels assigned to this category increases ten-fold.

Without direct observations on the ground, any interpretation of these results must be viewed with caution. Nonetheless, the results suggest that the development of roads in this location does have a significant effect on agricultural land use and deforestation. Roads seem to influence locations most near currently forested areas. Removing roads allows forests to "grow" back down mountainsides. Removing roads also reduces the probability that irrigated crop land will remain irrigated. Removing villages has little direct effect on forest pixels. Instead the effect appears to be to shift land use away from many categories to the category that represents idle land with natural vegetation.

**Conclusions**

This paper has shown how data from a satellite image and other geographical sources can be used in a model of land use determination. It presents new techniques for correcting for spatial dependence and new measures of the predictive power of multinomial logit models. For a region in central Mexico, the cross-section results suggest that road access to a location does affect land use. The locations identified as forest increase as road access becomes more

difficult. From a statistical perspective the model has moderate aggregate predictive power and strong predictive power for land use categories of interest – forested and irrigated crop areas.

These particular results should be viewed with caution since the simulations use values outside the sample set and the model is essentially static while land use change is a dynamic process. Finally, field observation is crucial to assessing the accuracy of the land use classifications. Despite these caveats, the results have reasonable interpretations. More importantly, the availability of techniques to turn satellite images into economic data opens a vast, unexploited data set for analysis of environmental issues.

## References


Anselin, L. (1988). Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers.

Bliss, Norman, EROS Data Center, personal communication. The soils data were digitized by the EROS Data Center from a map entitled "Carta Edafologica" produced by Estados Unidos Mexicanos Secretaria de Programacion y Prespuesto.

Chomitz, K. M., & Gray, D. A. (1995). Roads, Land, Markets and Deforestation: A Spatial Model of Land Use in Belize, Working Paper, World Bank Policy Research Department, Environment Infrastructure, and Agriculture Division.

Congalton, R. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. Remote Sensing of Environment, 37, 35-46.

Eswaran, Hari, U.S.D.A. soil scientist, personal communication.

Greene, W. H. (1991). Limdep Version 6.0 User's Manual and Reference Guide. Bellport: Econometric Software, Inc.

Greene, W. H. (1993). Econometric Analysis (2nd ed.). New York: Macmillan Publishing Company.

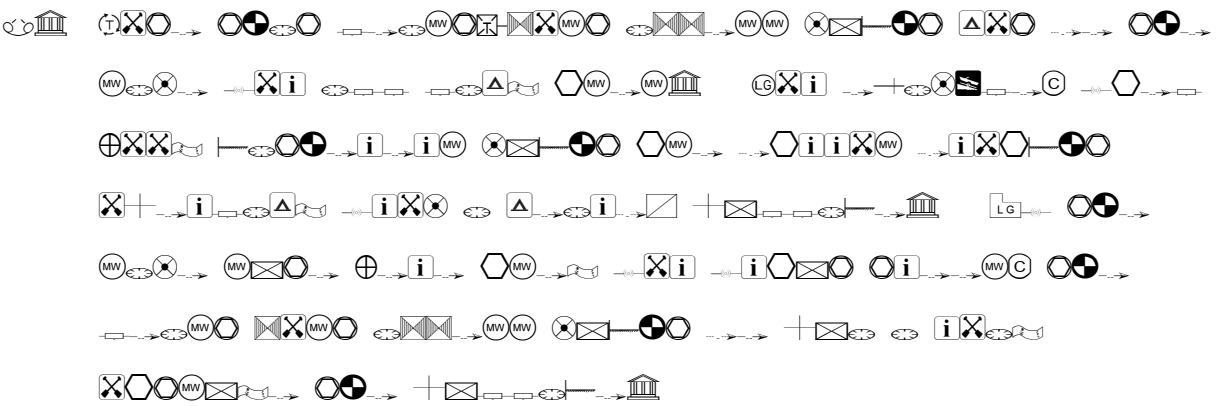
Haining, R. (1990). Spatial Data Analysis in the Social and Environmental Sciences. Cambridge: Cambridge University Press.

Maddala, G. S. (1983). Limited Dependent and Qualitative Variables in Econometrics. Cambridge: Cambridge University Press.

Richards, J. A. (1993). Remote Sensing Digital Image Analysis: An Introduction (2nd ed.). Berlin: Springer-Verlag.

Footnotes

1. This derivation follows Chomitz and Gray but introduces the time dimension explicitly.



3 A technical appendix with the derivation is available from the first author.

4. Note that the term that includes the discount rate, i, is not land use-specific. Because our proxy variables are not time-dependent, changes in the discount rate have an equal impact on all land uses. The discount rate is influenced by both the local interest rate and local land use rules. It is not uncommon for informal use rules to play a more important role in determining land use than official rules and regulations. If data on land tenure are available, they can act as a proxy for the location-specific effects of changing discount rates. For this analysis we have no proxies for the discount rate.

5. The image is from the path 27 row 46 data set of the North American Land Categorization (NALC) project of EPA and USGS. The NALC data consist of three MSS images and a digital elevation data set. The images were taken between the early 1970s and the early 1990s. Each image consists of four sets of data (called bands) corresponding to intensity of reflected light in the green, red, and two infrared frequencies. Possible values range from 0 to 127. The image and elevation data are georeferenced; every location (also called a pixel) has a corresponding elevation value

and latitude and longitude coordinates. Each pixel in a band corresponds to the average intensity of light from a location 60 meters on a side. The 1973 image was chosen for this analysis because it corresponds most closely to the road and population center data available in the Digital Chart of the World.

6. This routine uses a histogram peak technique. It identifies a "cluster" of pixels with similar reflectance values in three frequency bands. The cluster is a set of pixels around each peak. We allowed the software to determine the number of clusters, by including peaks such that 99% of the pixels are accounted for.

7. Some misclassification is evident. Some areas that appear forested in the false color image were classified in the same category as irrigated areas. This might be due to the confounding effect of shadows on the classification process. In addition, at least one lake was put into category 1 and another into category 2. All bodies of water in this area carry high sediment loads and large amounts of aquatic plant life which commonly results in misclassification.

8. The index was derived from a map where the number of discrete soil regions is much less than the number of pixels in the image. Therefore, the soils index is only a rough approximation of the actual productivity of the soil at a given location.

9. The choice of the square root of slope is arbitrary.

10. The general expression of spatial effects is as follows (Anselin, p 34):

$$y = \rho W_l y + X\beta + \varepsilon$$

$$\varepsilon = \lambda W_\varepsilon \varepsilon + \mu$$

$$\mu \sim N(0,\Omega)$$

The W's are spatial weight matrices with $W_l$ capturing the nature of spatial dependence of the dependent variable y, and $W_\varepsilon$ capturing spatial dependence in the error term. The diagonal elements of the error covariance matrix Ω are

$\Omega_{ii} = h_i(z\alpha)$ $\quad h_i > 0.$

z - exogenous variables that cause heteroskedasticity

If $\alpha = 0$, $h = \sigma^2$ (i.e., homoskedastic errors). It is usually assumed that $\Omega_{ij}=0$, $i \neq j$.

11. In a row-standardized weight matrix, $w^*_{ij}=\frac{w_{ij}}{\Sigma\Sigma_{i\ j} w_{ij}}$

12. Permutations using logs of selected geophysical variables have large condition numbers (larger than 300). Presumably, taking logs of a subset of the RHS variables reduces the variation in the X'X matrix.

13. This technique could also be used to estimate the effects of decreasing cost of access, for example, by building a new access road.