Population Estimation Using Remotely Sensed Data and Regression Analysis

Addison Larson – Intro to Remote Sensing Final Project

PROJECT PURPOSE

This project uses remotely sensed data and statistical modeling to estimate the population of Midland County, Texas. While the U.S. Census Bureau conducts a decennial Census, remote sensing information may be employed to quickly and cheaply estimate the population of urban areas in the interim. This project explores two statistical models for population estimation — a linear model and a spatial autoregressive model — and looks briefly at model fit.

QUESTIONS

How accurately can land cover classifications estimate population? How can we improve population estimation through land cover classification and regression analysis?

AREA OF STUDY AND DATA

The study area is located at 32°0'N 102°6'W, including the City of Midland and surrounding Midland County. The City of Midland is the county seat of Midland County, and had a population of 111,147 in 2010.

This study draws upon two major data sources from the year 2010: data from the U.S. Census Bureau, and Landsat 5 imagery.

Census data includes 2010 decennial population counts by census tract and TIGER/LINE files providing official boundaries of the county and census tracts under study.

The Landsat imagery (path 30, row 38) was acquired on February 17, 2010, under clear-sky conditions.

FIGURES

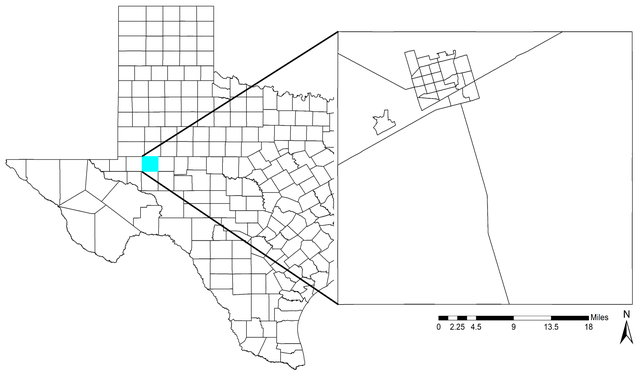


Fig. 1: Study area of Midland County, inset.

Fig. 2: Impervious.

Fig. 3: Cropland.

Fig.4: Arid.

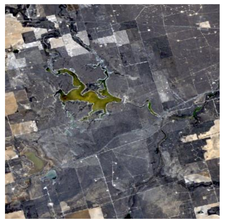


Fig.5: Water.

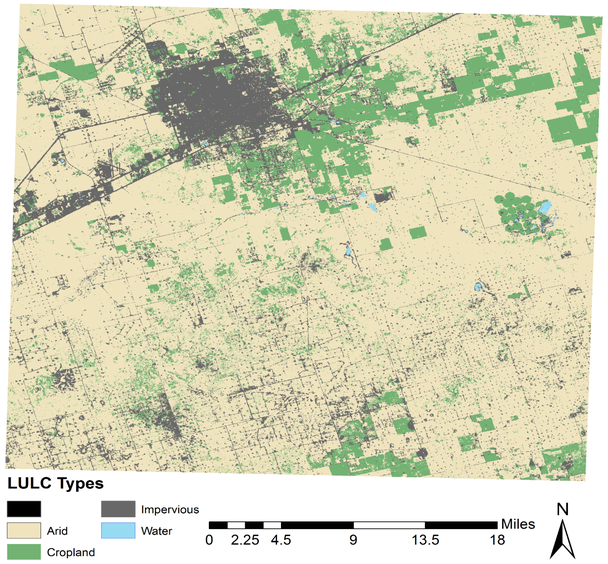


Fig. 6: Classified image of study area.

Fig. 7: Natural color image of study area.

METHODOLOGY

Midland County is divided into 27 Census tracts with areas ranging from 1.62 km^2 to 1125.34 km^2. Landsat 5 imagery of the study area was classified into 4 land cover types: arid, cropland, water, and impervious surfaces (using bands 7, 4, and 2 as red, green, and blue, respectively). After the classification, 45107 random points of land cover information were extracted at a density, on average, of 1670.63 points per census tract and 19.31 points per km^2. Point information was spatially aggregated by census tract and regression analysis was performed to estimate population by land cover types.

RESULTS

Linear Model

Variable	Coef.	SE	p-value
Arid	-1.698	0.577	0.008***
Cropland	9.699	2.093	0.0001***
Impervious	8.823	3.504	0.0195**
Water	-293.915	79.105	0.001***
Adj. R^2	0.6168		
F	11.46		

SAR Model

Variable	Coef.	SE	p-value
Arid	-0.771	0.439	0.079
Cropland	9.826	1.526	0.0001***
Impervious	2.545	2.367	0.282
Water	-392.252	66.915	0.0001***

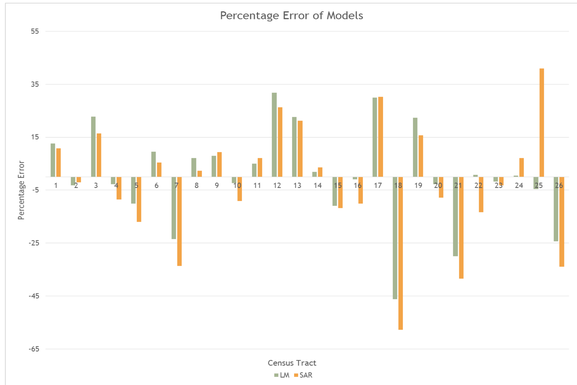
** and *** denote significance at α = 0.05 and 0.01, respectively.

Spatial Autocorrelation

Moran's *I* indicates a slightly positive but statistically insignificant correlation, with value *I* = -0.00836 and *E(I)* = -0.03842.

Percentage Error of Models

The overall percentage error of the linear model and SAR model are 0.425% and -1.939%, respectively; however, this masks large swings in model accuracy by census tract.



REFERENCES

Li, G. (2008). *Integration of remote sensing and census data for land -use and land cover classification and population estimation in Indianapolis, Indiana* (Doctoral Dissertation). Retrieved from ProQuest Dissertations Publishing. (3305420)

Weng, Q. (2010). *Remote Sensing and GIS Integration: Theories, Methods, and Applications.* New York: McGraw-Hill.