

VEGETATION AND POPULATION DENSITY IN URBAN AND SUBURBAN AREAS IN THE U.S.A.

Francesca Pozzi
and
Christopher Small

Columbia University
Palisades, NY 10964, USA
fpozzi@ciesin.columbia.edu
small@ldeo.columbia.edu

*Presented at the Third International Symposium of Remote Sensing of Urban Areas
Istanbul, Turkey, June 2002*

ABSTRACT

Reflectance characteristics of human settlements influence energy fluxes in the physical environment as well as our ability to monitor urban growth with satellite-based sensors. Spectral heterogeneity at the scales of 10 to 30 meters complicate the traditional land cover classifications derived from moderate resolution satellite imagery in urban and suburban areas. In this study we consider population density and vegetation abundance as the principal demographic and physical characteristics in urban and suburban areas of the U.S.A. We investigate their relationship in the cities of Atlanta, Chicago, Los Angeles, New York, Phoenix and Seattle and compare the results with the USGS National Land Cover Dataset's urban classes. The multimodal population density distribution of the U.S.A. implies that it is possible to characterize suburban areas as those with population densities between 100 and 10,000 people/km², with rural areas at densities below 100 people/km² and urban areas at densities above 10,000 people/km². We find that maximum vegetation fraction diminishes with increasing population density over the full range of densities, but the spectral heterogeneity at pixel scales still results in a wide range of vegetation fractions in demographically urban and suburban areas. The resolution difference between census units and the Landsat sensor's field of view does not allow for a pixel scale spectral characterization of suburban land cover that is consistent with population density. A classification scheme based on spectral heterogeneity at multiple pixel scales, supplemented by auxiliary data sources, may provide a more accurate way to analyze urban growth.

1 INTRODUCTION

Urban growth in the U.S.A. is often characterized by expansion of suburban areas. The U.S. Census Bureau reports that in 1990 approximately 30% of the U.S. population lived in metropolitan areas of at least 5 million people (U.S. Census Bureau, 2001). It is also reported that between 1995 and 1996 more than 2 million people moved from U.S.A. cities and from non-metropolitan areas into the "suburbs" (Hansen, 1997). In the same report suburbs are defined as "all territory within Metropolitan Statistical Areas (MSA) but outside of a central city". Suburbs are most commonly known as "residential areas on the outskirts of a city or a large town" (Merriam-Webster Dictionary, 2002), and are often perceived as more vegetated areas, socially and economically dependent on large cities. Although such definitions are intuitive and easily understandable, there appears to be no consistent or formal characterization of suburban areas in terms of either physical or socioeconomic characteristics. If urban and suburban areas could be characterized by unique surface reflectance characteristics, urban growth could be efficiently quantified with optical sensors on operational satellites. Quantifying settlement patterns on the basis of reflectance properties would also facilitate physical models of regional environmental processes as well as ecological studies of the landscape.

In recent years, the relationship between population characteristics and environmental variables in urban areas has been increasingly explored, for a variety of purposes and applications. Some authors discuss cases of integration of land cover and population variables aimed at analyzing the correlation between them (Yuan et al, 1997; Radeloff et al., 2000; Chen, 2002) or at improving land cover classifications (Harris and Ventura, 1995; Vogelmann et al., 1998). To our knowledge, no study has yet been performed to examine the consistency of demographic and land cover characteristics of suburban areas across different physiographic environments.

In this study we investigate the question of whether suburban areas can be defined on the basis of demographic and physical characteristics, specifically population density and vegetation cover. Based on the observation that suburban areas are greener than urban centers, we attempt to quantify the extent to which suburban areas are vegetated in different U.S.A. cities. We consider suburban areas in terms of population density and of apparent spectral reflectance using Landsat data and quantify the relationship between the two variables. We also compare these results with the USGS National Land Cover Dataset (National Land Cover Characterization, 2001). We selected the Low Intensity Residential, High Intensity Residential and Commercial/Industrial/Transportation classes and cross-referenced the areal extents for these classes with vegetation fractions and urban classification estimates for the six cities included in this study.

2 DATA

The six cities we selected have very different characteristics, in terms of geographic location, spatial structure, physical environment and urban growth dynamics. Some of them are located in a temperate climate, both in a deciduous forest biome (New York, Chicago, Atlanta) and in an evergreen forest biome (Seattle) while others are located in an arid (Phoenix) or semi-arid climate (Los Angeles). We also included fast growing cities (Phoenix and Seattle), and cities that have experienced rapid growth in the past and now are characterized by large population (New York, Chicago, Los Angeles).

2.1 POPULATION DENSITY

We calculated population density, in persons/km², from the 1990 U.S. Census Bureau population counts at the block level, the lowest in the U.S. census structural hierarchy. These data are available separately as spatial data (Topologically Integrated Geographic Encoding and Referencing system-TIGER®) and tabular data (Summary Tape Files-STFs) for each county in the U.S.A. For each city we selected one or more counties containing the Centered Business District (CBD) and the surrounding suburbs. We then created a smaller subset concordant with Landsat coverage. The areas reported below refer to the extent of land and water within each subset, and are considered representative of urban and suburban settlements for each city.

Atlanta: 900 km²;

Chicago: 950 km²;

Los Angeles: 3100 km²;

New York: 2000 km²;

Phoenix: 4700 km²;

Seattle: 3200 km².

2.2 VEGETATION FRACTION

The spatial scale and the spectral variability of urban and suburban land cover pose serious problems for traditional image classification algorithms. In areas where the reflectance spectra of the land cover vary appreciably at scales comparable to, or smaller than, the Ground Instantaneous Field Of View (GIFOV) of most operational satellite sensors, the spectral reflectance of an individual pixel will generally not resemble the reflectance of a single land cover class but rather a mixture of the

reflectances of two or more classes present within the GIFOV. Because they are combinations of spectrally distinct land cover types, mixed pixels in urban areas are frequently misclassified as other land cover classes. Similarly, the definition of an “urban” spectral class will usually incorporate pixels of other non-urban classes.

Analysis of Landsat TM imagery suggests that the spectral reflectance of many urban areas can be described as linear mixing of three distinct spectral endmembers (Small, 2001a; Small, 2001b). If an urban area contains significant amounts of vegetation then the reflectance spectra measured by the sensor will be influenced by the reflectance characteristics of the vegetation. Macroscopic combinations of homogeneous “endmember” materials within the GIFOV produce a composite reflectance spectrum that can often be described as a linear combination of the spectra of the endmembers (Singer and McCord, 1979). If mixing between the endmember spectra is predominantly linear and the endmembers are known *a priori*, it may be possible to “unmix” individual pixels by estimating the fraction of each endmember in the composite reflectance of a mixed pixel (e.g. Adams et al., 1986; Boardman, 1989).

Principal component analysis of urban reflectance consistently yields eigenvalue distributions suggesting that the majority of scene variance is contained within a two-dimensional mixing plane (Small, 2002). The triangular distribution of pixels in the mixing space defined by the two principal components bears a similarity to the well known Tasseled Cap distribution discovered by Kauth and Thomas (Kauth and Thomas, 1976). The feature space distributions are similar in the sense that both contain a vegetation endmember that is distinct from a continuum of built surface between high and low albedo endmembers. Representing reflectance as simple mixtures of endmembers provides a consistent, verifiable and physically meaningful description of a wide variety of land covers.

The spectral endmembers determined for the areas investigated here correspond to low albedo (e.g. water, shadow, roofing), high albedo (e.g. cloud, sand, roofing) and vegetation. The strong visible absorption and infrared reflectance that is characteristic of vegetation is sufficiently distinct from the spectrally flat reflectance of the low and high albedo endmembers to allow the three components to be “unmixed” by inverting a simple three component linear mixing model (Small, 2001a). The result of the unmixing is a set of fraction images showing the areal percentages, given as fractions between 0 and 1, of each endmember present within each pixel. Analysis of Landsat, Ikonos and AVIRIS imagery of several urban/suburban areas shows that a three component linear mixing model provides stable, consistent estimates of vegetation fraction for both constrained and unconstrained inversions using three different endmember selection methods (Small, 2001b). Vegetation fraction estimates derived from Landsat TM data were validated with aerial vegetation fractions calculated from 2 m aerial photography and generally showed agreement to within 10% (Small, 2001a). The vegetation fraction estimates given here were derived from Landsat TM and validated with Ikonos MSI imagery.

2.3 LAND COVER

The National Land Cover Dataset (NLCD) was produced by the US Geological Service (USGS) as part of a cooperative project between the USGS and the US Environmental Protection Agency to develop a consistent land cover data layer for the conterminous U.S. based on 30-meter Landsat TM data (National Land Cover Characterization, 2001). The base dataset of the project was leaves-off TM data, nominal-1992 acquisitions. The classification procedure made use of ancillary datasets, including USGS 3-arcsecond Digital Terrain Elevation Data, U.S. Census population and housing data, and USGS Land Use and Land Cover data, among others.

The classification system used for NLCD is modified from the Anderson land use and land cover classification system. In the NLCD classification, “Developed” is defined as “Areas characterized by a high percentage (30 or greater) of constructed materials (e.g. asphalt, concrete, buildings, etc)”. We selected the three Level II classes included in the “Developed” class. The definition of the three

classes is reported below. For these, however, we could not find detailed information on the percent cover estimates in the original documentation (National Land Cover Characterization, 2001).

Low Intensity Residential (LIR): Includes areas with a mixture of constructed materials and vegetation. Constructed materials account for 30-80 percent of land cover. Vegetation may account for 20-70 percent of land cover. These areas most commonly include single-family housing units. Population densities will be lower than in high intensity residential areas.

High Intensity Residential (HIR): Includes highly developed areas where people reside in high numbers. Examples include apartment complexes and row houses. Vegetation accounts for less than 20 percent of land cover. Constructed materials account for 80-100 percent of land cover.

Commercial/Industrial/Transportation (CIT): Includes infrastructure (e.g. roads, railroads, etc) and all highly developed areas not classified as High Intensity Residential.

3 ANALYSIS AND RESULTS

A multimodal distribution of population density within the United States (Figure 1) suggests that it is possible to characterize rural, urban and suburban areas based on population density (Pozzi and Small, 2001). We consider suburban areas to be those with population densities between 100 and 10,000 people/km².

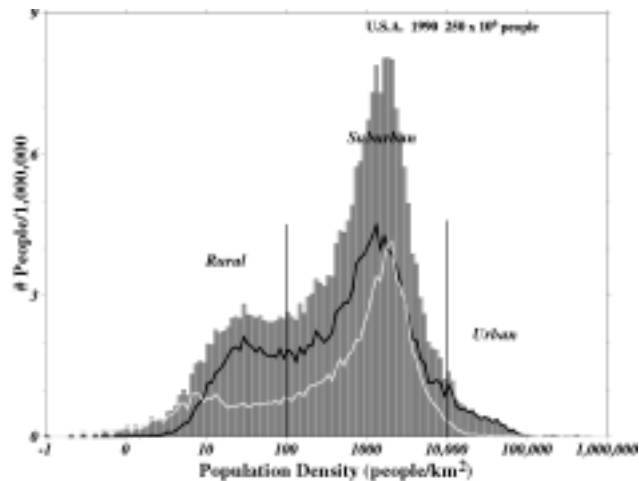


Figure 1. Histogram of the Population Density for the U.S., showing also the distribution for Eastern U.S. (East of the 90° W, black line) and Western U.S. (grey line).

To perform the study on the six cities, spatial and tabular data from the Census were initially aggregated based on the block numeric codes for each county. The resulting vector layers were then projected to UTM coordinates, rasterized to a 30 m grid and coregistered to the Landsat data. By combining population density and vegetation fraction, we produced an “Urban Classification” for each city (Figure 2). The same figure shows also their bivariate population distributions as a function of population density and vegetation fraction. We then summed the bivariate distributions to produce marginal distributions of people as functions of population density and vegetation fraction for each city. Finally, to compare these results with the three selected USGS NLCD classes, we produced distributions of areal extents of each class as a function of population density and vegetation fraction (Figure 3).

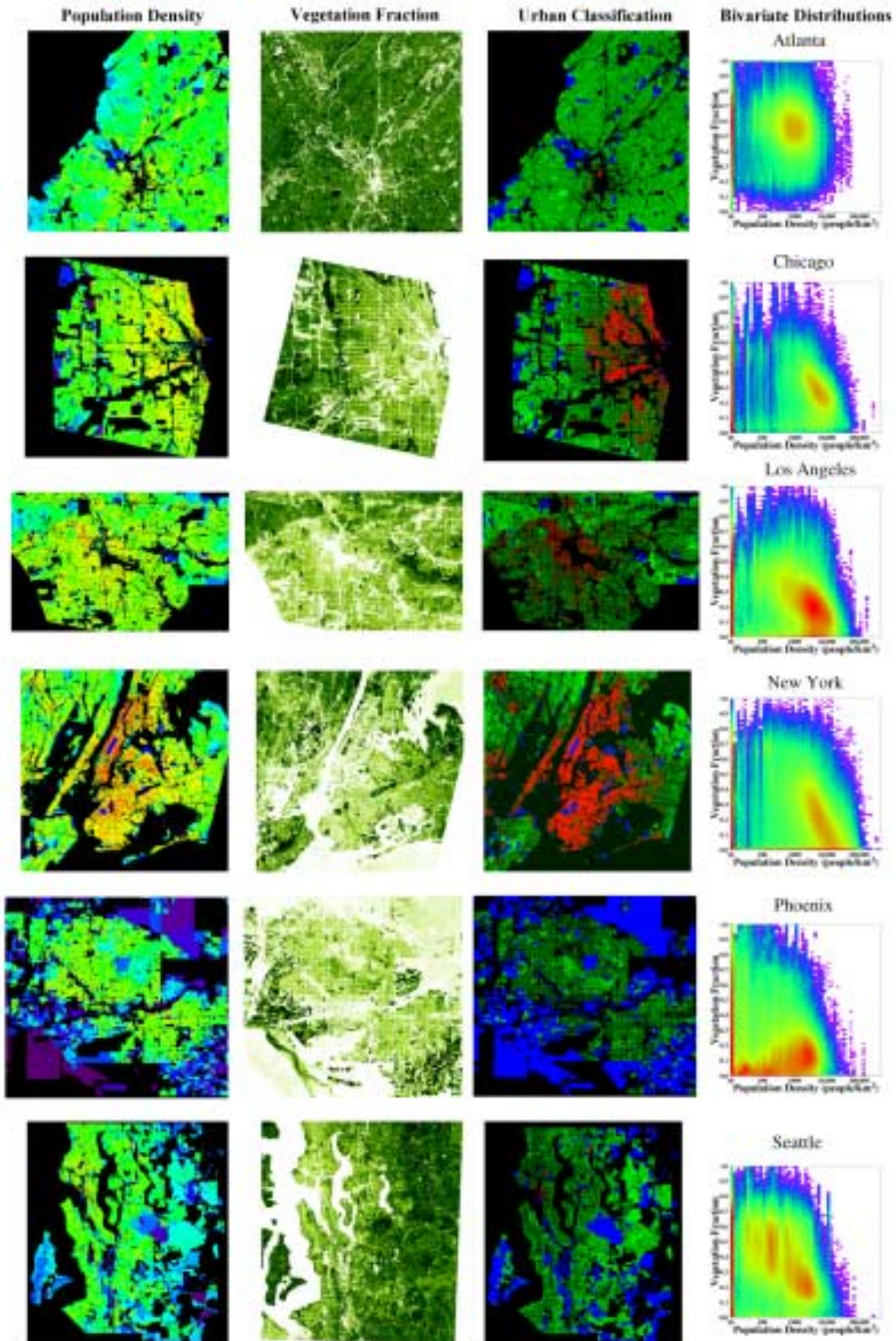


Figure 2. Spatial Distribution of Population and Vegetation.

In the Urban Classification, rural population densities are shown in blue, urban in red and suburban in green. Different shades of green correspond to different amounts of vegetation. Full resolution images are available at http://sedac.ciesin.columbia.edu/urban_rs.html

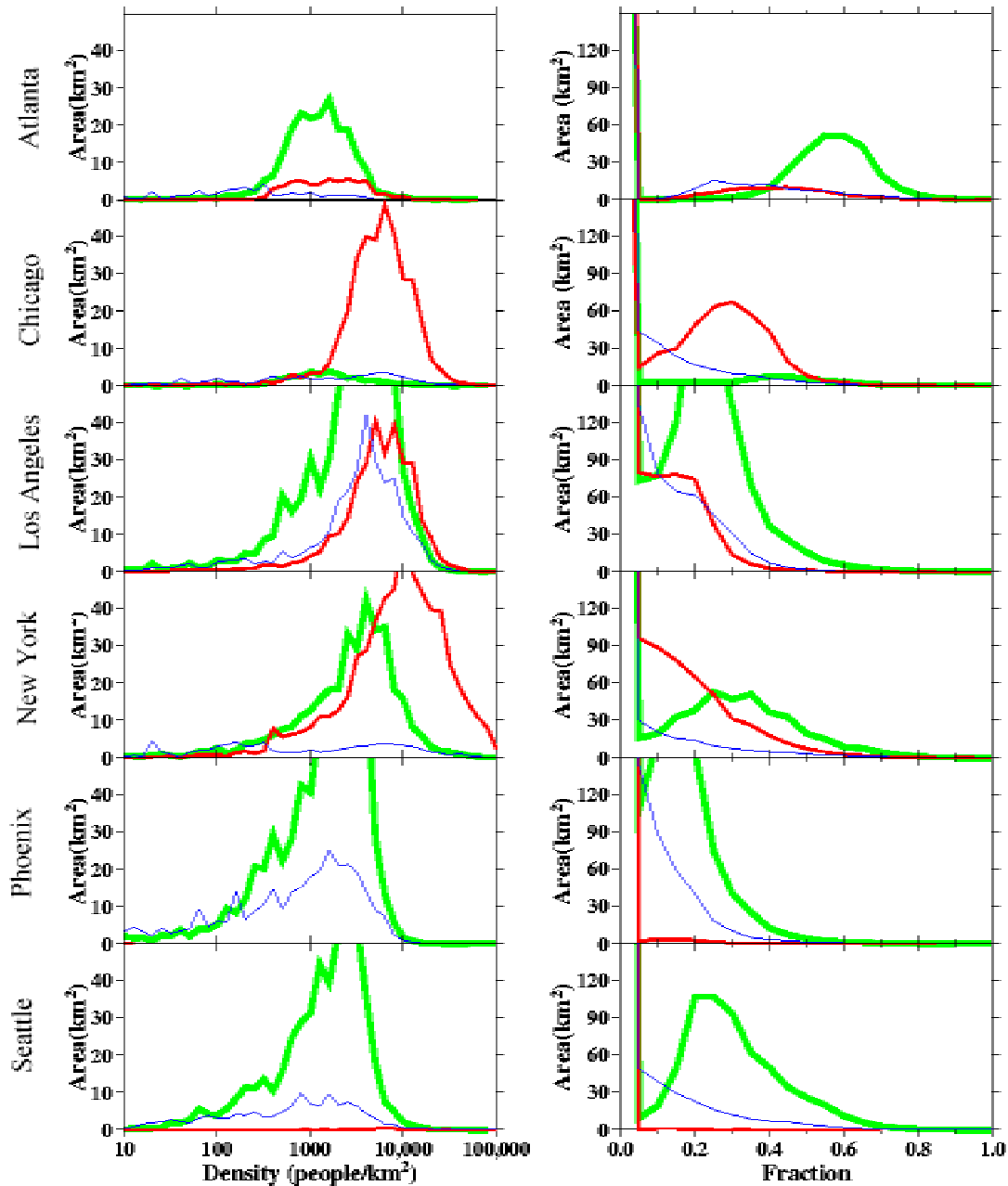


Figure 3. Distributions of areal extent of each “developed” USGS NLCD class as functions of population density and vegetation fraction. Green (thickest) lines represent LIR, red (thick) lines represent HIR and blue (thin) lines represent CIR.

4 DISCUSSION

As discussed in a previous work by the authors, there appears to be a pattern for suburban areas, both in terms of population density and vegetation fractions in the six cities considered (Pozzi and Small, 2001). The population density histograms show the suburban peak characteristic of the entire U.S.A. with Atlanta and New York at the extremes and the vegetation fraction histograms show a similar consistent pattern, with peaks varying between about 0.1 and 0.55. We find a consistent sub-linear

relationship, with vegetation fraction decreasing with increasing population density, for the largest cities (New York, Chicago and Los Angeles). However, the differences between the physiographic environments and the urban structures for the six cities are such that the peaks of the bivariate histograms are spread across a range of population densities and vegetation fractions (Figure 2).

Conversely, the distributions of areal extent of the three USGS NLCD classes as functions of population density and vegetation fraction do not seem to present a clear pattern among the six cities considered (Figure 3). In particular, the distribution of the Commercial/Industrial/Transportation class as a function of Population Density presents a clear peak corresponding to population densities of about 10,000 to 100,000 people/km² for Los Angeles, Phoenix and Seattle, while the High Intensity Residential has a peak corresponding to less than 100 people/km² in the cities of Phoenix and Seattle. While in the case of Los Angeles all the three classes present their peaks at population of about 1,000 to 10,000 people/km², New York and Atlanta seem to present a more logical pattern, with distributions of CIT decreasing and LIR and HIR increasing with increasing population densities.

The distributions of the three classes as a function of vegetation fraction seem to be a little more consistent, in that the Commercial/Industrial/Transportation class presents peaks at very low vegetation fractions and the Low Intensity Residential class distributions have peaks varying between 0.1 and 0.3, with the exception of Atlanta (0.6) and Chicago (almost no peak). Nonetheless, the ample range of vegetation fractions covered by the High and Low Intensity Residential distributions does not allow for unique consistent correlation between the USGS classification of developed areas and their correspondent spectral characteristics, in terms of vegetation cover.

The major difficulty associated with thematic classifications of urban areas that rely on moderate resolution imagery is the complexity of the urban landscape, which consists of a wide variety of land use classes. Unlike most other land cover classes, the urban/suburban mosaic is consistent only in its spectral heterogeneity at the scale of most operational satellite sensor GIFOVs (Small, 2001b). The resolution difference between census tracts and the Landsat sensor does not allow for a simple spectral characterization of suburban landcover. In the cities we investigated, the most consistent spectral characteristic of “demographically suburban” areas was related to the amount of vegetation cover. Large cities with high density urban cores show a distinct linear decrease in the modal vegetation fraction with increasing $\text{Log}_{10}(\text{population density})$ and generally maximum vegetation fraction diminishes with increasing population density. Nonetheless spectral heterogeneity still results in a wide range of vegetation fractions in demographically suburban areas and we find no evidence for a single consistent relationship between suburban population density and vegetation abundance in the U.S.A.

Co-analysis of population density, vegetation fraction and USGS urban classes lead us to conclude that it is not possible to consistently characterize urban and suburban areas in the U.S.A., based on the reflectance characteristics at the resolution scale of 10 to 50 meters. However, quantitative characterization of vegetation abundance in suburban areas provides a basis for comparison of the physical environments in which most Americans reside. Vegetation plays a major role in influencing the microclimate of the human habitat and understanding the relationship between vegetation cover and human activities in cities and their surroundings would be fundamental for any environmental urban studies. Such results may provide a basis not only for quantitative analysis of urban sprawl and other land use policy implications, but also for the development of a new type of classification schemes for urban areas, based on a “continuum” range of values for variables such as vegetation cover and broader demographic-based urban/suburban classes.

ACKNOWLEDGEMENTS

This work was supported by the Center for International Earth Science Information Network (CIESIN), through the NASA Socioeconomic Data and Applications Center (SEDAC), and the Columbia Earth Institute.

REFERENCES

- Adams, J.B. et al., 1986. Spectral Mixture Modeling: a New Analysis of Rock and Soil Types at the Viking Lander 1 Site. *Journal of Geophysical Research*, 91, pp.8089-8122.
- Boardman, J.W., 1989. Inversion of Imaging Spectrometry Data Using Singular Value Decomposition. *Proceedings of the 12th Canadian Symposium on Remote Sensing*, pp.2069-2072.
- Chen, K., 2002. An Approach to Linking Remotely Sensed Data and Areal Census Data. *International Journal of Remote Sensing*, 23(1), pp.37-48.
- Hansen, K.A., 1997. Geographical Mobility: March 1995 to March 1996. In: *Current Population Reports*, P20-497, U.S. Census Bureau.
- Harris, P.M. and Ventura, S.J., 1997. The Integration of Geographic Data with Remotely Sensed Imagery to Improve Classification in an Urban Area. *Photogrammetric Engineering and Remote Sensing*, 61(8), pp.993-998.
- Kauth, R.J. and Thomas, G.S., 1976. The Tasseled Cap – A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by Landsat. *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*, pp.4041-4051.
- Merriam-Webster Dictionary, 2002. Merriam-Webster. On-line at <http://www.m-w.com/home.htm>
- Pozzi, F. and Small, C., 2001. Exploratory Analysis of Suburban Land Cover and Population Density in the U.S.A.. *Proceedings of the IEEE Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, 8-9 November 2001, Rome, Italy, pp. 250-254
- Radeloff, V.C., et al., 2000. Exploring the Spatial Relationship Between Census and Land Cover Data. *Society and Natural Resources*, 13(6), pp.599-609
- Singer, R.B. and McCord, T.B., 1979. Large Scale Mixing of Bright and Dark Surface Materials and Implications for Analysis of Spectral Reflectance. *Journal of Geophysical Research*, pp.1835-1848.
- Small, C., 2001a. Estimation of Urban Vegetation Abundance by Spectral Mixture Analysis. *International Journal of Remote Sensing*, 22(7), pp.1305-1334.
- Small, C., 2001b. Multiresolution Analysis of Urban Reflectance. *Proceedings of the IEEE Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, 8-9 November 2001, Rome, Italy, pp.15-19
- Small, C., 2002. A Global Analysis of Urban Reflectance. This volume.
- U.S. Census Bureau, 2001. Population Change and Distribution 1990 to 2000. *Census 2000 Brief*.
- USGS, 2001. National Land Cover Characterization. <http://landcover.usgs.gov/natl/landcover.html> (December 2001).
- Vogelmann, J.E., Sohl, T. And Howard S.M., 1998. Regional Characterization of Land Cover Using Multiple Sources of Data. *Photogrammetric Engineering and Remote Sensing*, 64(1), pp.45-57.
- Yuan, Y., Smith, R.M. and Limp, W.F., 1997. Remodeling Census Population with Spatial Information from Landsat TM Imagery. *Computers, Environment and Urban Systems*, 21(3/4).