Analysis of Urban Land Cover and Population Density in the United States

Francesca Pozzi and Christopher Small

Abstract
In this study we investigate the question of whether urban and suburban areas in the United States can be defined on the basis of demographic and/or physical characteristics, in particular, population density and vegetation abundance. 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 bimodal distribution of U.S. population density provides a demographic basis for distinguishing rural and suburban land use, while a distinct tail of high population densities (>10,000 people/km²) corresponds to high intensity urban residential cores. Results show that the maximum vegetation fraction diminishes with increasing population density, but the spectral heterogeneity at pixel scales still results in a wide range of vegetation fractions within demographically urban and suburban areas. None of the USGS residential classes show a strong correspondence to either vegetation fraction or population density. However, quantitative characterization of vegetation abundance provides a basis for comparison of the physical environments of suburban areas. We suggest that classification schemes based on spectral heterogeneity at multiple pixel scales, supplemented by auxiliary data sources, may provide a more accurate and self-consistent way to quantify urban land use and analyze urban growth than traditional thematic classification schemes.

Introduction
Changes in land use and corresponding changes in land cover have a direct impact on a variety of physical environmental and ecological processes (e.g., Turner et al., 1990; Meyer and Turner, 1990). Recognition of this fact has led to systematic efforts to map and monitor land use and land cover changes using remotely sensed observations (Skole, 1994; DeFries and Townshend, 1994; Singh, 1989; Vogelmann et al., 1998). Most land cover classification schemes are designed to map the spatial distribution of natural resources on the basis of their reflectance properties. This usually works well because many classes of naturally occurring land cover have characteristic reflectance properties that distinguish them from other classes. Urban and suburban areas, however, are generally defined on the basis of land use. These complex mosaics of land cover types generally do not have homogeneous reflectance properties at the scale of the moderate resolution (10 m to 100 m) imagery usually used for the classifications. As a result, land cover classifications used to generate land use classes often produce results that can be difficult to represent in terms of their physical properties. Attempts to overcome such difficulties by using fuzzy classification of suburban land cover are illustrated in Zhang and Foody (1998, 2001). Spectral heterogeneity is also a problem for physical and ecological process models that seek to use these classifications to parameterize land surface physical properties (Chen and Dudhia, 2001a,b; Walko et al., 2000; Zeng et al., 2000). Inconsistency between land cover and land use definitions also makes it difficult to use these classifications to quantify urban growth and human settlement patterns. The purpose of this study is to investigate the consistency of urban and suburban land use and land cover characteristics in the United States.

To our knowledge, no study has yet examined the consistency of residential land use and land cover characteristics in urban and suburban areas across different physiographic environments. In this study we investigate the question of whether urban and suburban areas in the U.S. can be defined on the basis of demographic and/or physical characteristics, specifically, population density and vegetation cover. Based on the observation that suburban areas in the U.S. are generally greener than urban centers, we attempt to quantify the extent to which vegetation abundance corresponds to population density in Atlanta, Chicago, Los Angeles, New York, Phoenix, and Seattle. We also explore the consistency of demographic and vegetation cover characteristics with respect to the USGS National Land Cover Dataset (NLCD, National Land Cover Characterization, 2001). From the complete NLCD, we selected the Low Intensity Residential, High Intensity Residential and Commercial-Industrial-Transportation classes and co-registered the areal extents for these classes with vegetation fractions and population density for the six cities included in this study. An objective of this analysis is to determine whether a remotely sensed physical characteristic, such as vegetation abundance, can be used to define land cover classes associated with specific residential land uses. Another objective is to quantify the relationship between population density and vegetation fraction and to compare each to existing land use classifications.

Francesca Pozzi is with the Center for International Earth Science Information Network, Columbia University, 61 Route 9W, Palisades, NY 10964 (fpozzi@ciesin.columbia.edu).

Christopher Small is with the Lamont-Doherty Earth Observatory, Columbia University, 61 Route 9W, Palisades, NY 10964 (small@LDEO.columbia.edu).

Photogrammetric Engineering & Remote Sensing
0099-1112/05/7106–0719/$3.00/© 2005 American Society for Photogrammetry and Remote Sensing
Background

In the United States, urban growth is often manifest as expansion of suburban areas. Urban growth and the concentration of people in urban areas have significant environmental and socioeconomic impacts at local, regional, and global levels. Policymakers and the public continue to raise concerns about the effects of urbanization on the landscape, such as loss of agricultural land and natural vegetation, increased traffic congestion, and associated degradation of air quality. Such detrimental impacts of urban sprawl are being increasingly incorporated into urban planning assessments. For instance, a number of “sprawl reports” have been compiled (Planning Commission Journal, 2001) and smart growth policies written and implemented. The U.S. Environmental Protection Agency recently supported the International City/County Management Association (ICMA) and the Smart Growth Network to produce “Getting to Smart Growth II: 100 More Policies for Implementation” (U.S. Environmental Protection Agency, 2005).

The U.S. Census Bureau reports that in 1990 approximately 30 percent of the U.S. population lived in metropolitan areas of at least five million people (U.S. Census Bureau, 2001). It is also reported that between 1995 and 1996 more than two million Americans moved from cities and non-metropolitan areas into the “suburbs”, where suburbs are defined as “all territory within Metropolitan Statistical Areas (MSAs) but outside of a central city” (Hansen, 1997). This is one of the definitions used by the U.S. Census Bureau. The Census provides definitions for Metropolitan Area, Metropolitan Statistical Area, Consolidated and Primary Metropolitan Statistical Area, and Central City. The Metropolitan areas (MA) in the U.S. are designated and defined by the federal office of Management and Budget based on a set of standards that are published in a Federal Register Notice. An example of the complexities associated with this type of definition is provided in the metadata itself (U. S. Census Bureau, 2005): “The territory, population, and housing units in MAs are referred to as “metropolitan.” The metropolitan category is subdivided into “inside central city” and “outside central city.” The territory, population, and housing units located outside territory designated “metropolitan” are referred to as “non-metropolitan.” The metropolitan and non-metropolitan classification cuts across the other hierarchies; for example, there is generally both urban and rural territory within both metropolitan and non-metropolitan areas. These definitions are based on administrative characteristics or population totals, and do not take into account physical and demographic characteristics.

In an attempt to characterize urban areas in a manner more useful to urban planners and researchers, in recent years a number of authors have examined the relationship between population characteristics and environmental variables in urban areas. Examples of these studies include those where integration of land cover, derived from remotely sensed physical characteristic can be used to represent land cover classes associated with specific residential land uses related to population density. In particular, we use vegetation abundance and population density, as described in the following paragraphs. Population density is extracted from the census at the block level, while vegetation abundance is derived from Landsat data at 30 m resolution. We rasterized the census data to 30 m pixels to match the vegetation data for several reasons; primarily because census blocks do not provide a consistent area of analysis, while the 30 m pixels of the Landsat sensor do. Moreover, the errors potentially introduced in the rasterization process are minimal compared to the loss of information that would happen if we used a single vegetation fraction for each block. The distribution of vegetation fraction within a census block is independent of the sensor’s pixel resolution, so by comparing the distribution of fractions within each block to the average density of that block, we retain all the information in the higher resolution satellite imagery.

Population Density

When assessing the impact of population on the surrounding environment, as in the case of urban growth, it is important that we think in terms of spatial distribution of human socioeconomic variables, such as population count, housing density, and land cover, and then discuss the potential of data integration to facilitate spatial analysis and decision-making processes at both local and regional scales (Yuan et al., 1997; Radeloff et al., 2000; Chen, 2002).

The urban landscape is characterized by a mixture of diverse material and land cover types. Such combination of elements produces a spectrally heterogeneous response that is more complex than the more spectrally homogenous reflectance properties of land cover types like forest and agriculture. This complicates the task of defining the urban and suburban landscape on the basis of physical characteristics because traditional supervised classification algorithms assume that each class is spectrally homogeneous and distinct from other classes. A more effective characterization of urban areas may provide a better understanding of the relationship between settlement patterns and land cover characteristics in cities and their suburbs, and thus facilitate quantitative analysis of urban sprawl and other land use policy. Quantifying settlement patterns on the basis of reflectance properties could also facilitate physical models of regional environmental processes, as well as ecological studies of the landscape.

Population Density and Vegetation Cover Analysis

To examine the consistency of urban and suburban land cover and demographic characteristics, we performed a comparative analysis for a sample of cities in the U.S. with very different geographic location, spatial structure, physical environment, and urban growth dynamics. Some of the cities are located in a temperate climate, both in a deciduous forest biome (New York, Chicago, and Atlanta) and in an evergreen forest biome (Seattle), while others are located in an arid (Phoenix) or semi-arid climate (Los Angeles). We also compared young, fast-growing cities (Atlanta, Phoenix, and Seattle), and cities that have experienced rapid growth in the past, and now are characterized by large stable populations (New York, Chicago, and Los Angeles).

One objective of this study is to determine whether it is possible to combine physical and demographic factors to characterize the urban and suburban environment in a quantitative manner. Since the characteristics of these six cities are very different, we do not expect to find a high level of correlation among them. We look at how a remotely-sensed physical characteristic can be used to represent land cover classes associated with specific residential land uses related to population density. In particular, we use vegetation abundance and population density, as described in the following paragraphs. Population density is extracted from the census at the block level, while vegetation abundance is derived from Landsat data at 30 m resolution. We rasterized the census data to 30 m pixels to match the vegetation data for several reasons; primarily because census blocks do not provide a consistent area of analysis, while the 30 m pixels of the Landsat sensor do. Moreover, the errors potentially introduced in the rasterization process are minimal compared to the loss of information that would happen if we used a single vegetation fraction for each block. The distribution of vegetation fraction within a census block is independent of the sensor’s pixel resolution, so by comparing the distribution of fractions within each block to the average density of that block, we retain all the information in the higher resolution satellite imagery.
population, not only absolute numbers. Naturally the total number of people living in a city, or region, has an impact on its surroundings, but it is the concentration of people in spatially localized areas that has the more significant environmental impacts both at the local and the global levels. This is why we use population density as the principal demographic characteristic of urban and suburban areas.

Population data were extracted from the 1990 U.S. Census Bureau as population counts at the block level and then converted into population density as persons/km². The block level is the lowest in the U.S. census structural hierarchy, and therefore allows analysis at relatively fine spatial scales. Census data are available separately as spatial data: Topologically Integrated Geographic Encoding and Referencing system TIGER®, and tabular data: Summary Tape Files (STF) for each county in the U.S. For each city we selected one or more counties containing what we would call “urban core” and the surrounding suburbs. We then created a smaller subset concordant with Landsat coverage. Below are reported the counties selected for each city and, in parenthesis, the areal extents of each subset. These subsets provide representative samples of urban and suburban settlements for each city.

- Atlanta: DeKalb and Fulton (900 km²);
- Chicago: Cook (950 km²);
- Los Angeles: Los Angeles (3,100 km²);
- New York Metropolitan Area: Bronx, Kings, New York, Queens, Richmond, Bergen, Essex, Hudson, Passaic, Nassau, Rockland, Westchester (2,000 km²);
- Phoenix: Maricopa (4,700 km²);
- Seattle: King (3,200 km²).

Spectral Mixing and 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 observed reflectance of an individual pixel will generally not resemble the reflectance of a single land cover class, but rather a mixture of the reflectance of two or more materials 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, 2005). 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 (Adams et al., 1986; Boardman, 1989; Smith et al., 1990).

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, 2005). 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 (water, shadow, and roofing), high albedo (sand and roofing) and vegetation. The strong visible absorption and high 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. Vegetation fraction estimates derived from Landsat TM data were validated with areal vegetation fractions calculated from 2 m aerial photography and generally showed agreement to within 10 percent (Small, 2001a). The vegetation fraction estimates given here were derived from Landsat TM as described by Small (2002). As in previous studies, the Landsat-derived estimates agree with the higher resolution measurements to within 10 percent (Small, 2004).

Analysis

Previous studies (Small, 2001b; Pozzi and Small, 2001) show that the distribution of population density of the United States has distinct modes corresponding to rural, urban and suburban areas. Figure 1 shows the multi-modal distribution of population density of the United States has distinct modes corresponding to rural, urban and suburban areas. Figure 1 shows the multi-modal distribution of population density in the United States. Based on this distribution, we consider suburban areas to be those with population densities between 100 to 10,000 people/km², and urban as of population densities greater than 10,000 people/km².
To perform the study on the six cities, spatial and tabular data from the 1990 U.S. 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 co-registered to the Landsat data. Plate 1 shows the bivariate population distributions as functions 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 (Figure 2).

The population density histograms show the suburban peak characteristic of the entire U.S. with Atlanta and New York at the extremes. The vegetation fraction histograms show a similar consistent pattern, with peaks varying between about 10 percent and 55 percent areal vegetation fraction (0.1 and 0.55). In the bivariate distributions, we find a consistent Log-linear relationship, with vegetation fraction decreasing with increasing Log10 of 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 (Plate 1).

Comparison With the National Land Cover Dataset

NLCD Description

The National Land Cover Dataset (NLCD) was produced as part of The National Land Cover Characterization project created in 1995 to support the original Multi-Resolution Land Characterization (MRLC) initiative, and fulfill the requirement to develop a nationally consistent land cover data set (National Land Cover Characterization, 2001). The NLCD is the result of a collaborative project between the USGS and the U.S. Environmental Protection Agency and is available for download on a state-by-state basis at http://landcover.usgs.gov/natlandcover.html (last date accessed: 11 March 2005). The dataset was compiled through unsupervised clustering of early 1990s Landsat TM data at 30 m resolution (Kelly and White, 1993) and augmented by a suite of other geospatial ancillary data sets, including topography, census, agricultural statistics, soil characteristics, other land cover maps, and wetlands data. Twenty-one classes of land cover were mapped using consistent procedures for the entire U.S. and based on a modified Anderson Level II Classification system. Details of the classification process are discussed in Vogelmann et al. (1998).

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.

Low Intensity Residential

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

High Intensity Residential

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

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

In the original documentation (National Land Cover Characterization, 2001), however, we could not find detailed information on the percent cover estimates, nor how to distinguish sub-classes within the Commercial/Industrial/Transportation class.

The accuracy assessment is done on all NLCD on a Federal Region basis following a revision cycle that incorporates feedback from MRLC Consortium partners and affiliated users. The accuracy assessments are conducted by private sector vendors under contract to the USEPA. A protocol has been established by the USGS and USEPA that incorporates a two-stage, geographically stratified cluster sampling plan (Steelman et al., 2000; Zhu et al., 2004) utilizing National Aerial Photography Program (NAPP) photographs as the sampling frame.

Analysis

In order to examine the correspondence between the three USGS NLCD “developed” classes and population density and vegetation cover for the six cities we produced distributions of population density and vegetation fraction for each class.

As Figure 3 illustrates, the distributions of population density and vegetation fraction for the three USGS/NLCD classes do not have a consistent pattern among the six cities considered. In particular, the distribution of population density for the Commercial/Industrial/Transportation class shows a clear peak corresponding to population densities of about 1,000 to 10,000 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 three classes have their peaks at population densities of about 1,000 to 10,000 people/km², New York and Atlanta show a pattern consistent with the assumption that the distribution of CIT will decrease while LIR and HIR will increase with increasing population densities.

The distributions of vegetation fraction for the three classes is more consistent, in that the Commercial/Industrial/Transportation class has 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 (0.5) and Chicago (almost no peak). Nonetheless, the wide range of vegetation fractions covered by the High and Low Intensity Residential classes does not result in a consistent correspondence between the USGS classification of developed areas and vegetation cover.

The considerable overlap in vegetation fraction among the developed classes would seem to preclude the use of the USGS classification as an input for physical process models in areas with significant fractions of urban or suburban land use. Mesoscale climate models routinely use lower resolution (1 km) land cover classifications as the basis for land surface boundary conditions, such as vegetative fraction, albedo, and roughness length by assigning representative parameter values to each class (Xiu and Pleim, 2000). The vegetation fraction, at 30 m resolution, could however represent a valuable input as a land surface model con-
Plate 1. Spatial Distribution of Population and Vegetation. In the Population Density maps, warmer colors correspond to higher density. In the Vegetation Fraction maps, darker shades of green correspond to higher vegetation fractions. The third set of maps shows the three demographic classes as identified in Figure 1, combined with vegetation fractions. 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/)

The constraint for models at resolutions of 1 km or higher. Given the spectral complexity of urban and suburban land cover, the USGS classification provides a reasonable thematic representation of the spatial distribution of these land use classes. However, we feel it is important to point out the physical inconsistencies in using land use categories as a
basis for physical properties associated with urban and suburban mosaics.

Conclusions
The purposes of this study were (a) to investigate the consistency of urban and suburban land cover, based on the relationship between population density and vegetation cover among different cities in the U.S., and (b) to examine the correspondence between the three USGS NLCD “developed” classes and population density and vegetation cover for the same cities. The objective was to quantify the variability of these land use categories among diverse American cities.

The first significant finding is that the block-level population density distribution in the U.S. may provide a demographic basis for distinguishing urban, suburban, and rural areas. A multimodal density distribution indicates that suburban areas can be characterized by moderate densities (100 to 10,000 people/km²). In larger cities like New York, Chicago, and Los Angeles, these suburban areas surround higher density urban cores, with population density values greater than 10,000 people/km², as well as sparsely populated city centers and commercial districts. In smaller cities, the high density core is absent, but the low density commercial districts are sometimes present.

The major difficulty associated with thematic classifications of urban areas that rely on moderate resolution imagery is the spectral complexity of the built landscape, which consists of a wide variety of land cover types. 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, 2005).

The spatial resolution of the Landsat sensor does not allow for a simple spectral characterization of suburban land-cover. We find that large cities with high density urban cores show a distinct linear decrease in the modal vegetation fraction with increasing Log₁₀(population density). Nonetheless spectral heterogeneity still results in a wide range of vegetation fractions in demographically suburban areas. We find no evidence for a single consistent relationship between population density and vegetation abundance in the U.S. The analysis of the USGS NLCD is consistent with these findings. At the scales of this study, comparable with the regional one for which the dataset was developed (Vogelmann et al., 1998), the areal extents of the three urban classes are not consistent with vegetation fraction in the cities considered.

Co-analysis of population density, vegetation fraction and USGS “developed” classes lead us to conclude that it is not feasible to consistently classify urban and suburban areas in the U.S. on the basis of moderate resolution (30 m) reflectance properties alone. 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. Such results may provide a basis for quantitative analysis of urban sprawl and other land use policy implications, as well as for the development of new classification schemes for urban areas. Classification schemes based on spectral heterogeneity at multiple pixel scales, supplemented by auxiliary data sources, may provide a more accurate way to analyze urban growth than current methods.

Acknowledgments

This research was supported by the Earth Institute at Columbia University, by the U.S. National Aeronautics and Space Administration (NASA) under contracts NAS8-98162 and NASA-93117 for operation of the Socioeconomic Data and Applications Center (SEDAC), and in part, by the U.S. Environmental Protection Agency’s Center for Environmental Research (NCER) STAR Program under Grant R-826733. The views expressed are those of the authors and not necessarily those of CIESIN, SEDAC, the Earth Institute, Columbia University, or NASA. Although the research described in this article has been funded in part by the U.S. Environmental Protection Agency, it has not been subjected to the Agency’s required peer and policy review, and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

References


PHOTOGRAMMETRIC ENGINEERING & REMOTE SENSING

June 2005 725


