Cities & Satellite Imagery: Models for Regional Change

Christopher Frazier
Graduate Student Researcher
Department of Civil Engineering
The University of Texas at Austin
ECJ 6.9, Austin, Texas 78712
Email: stanforth@mail.utexas.edu

Kara M. Kockelman
(Corresponding Author)
Clare Boothe Luce Assistant Professor of Civil Engineering
The University of Texas at Austin
ECJ 6.9, Austin, Texas 78712,
Tel: (512) 471-4379
FAX: (512) 475-8744
Email: kkockelm@mail.utexas.edu

Submitted for Presentation at the 2004 Annual Meeting of the Transportation Research Board

Word Count: 4204 words + 8 tables & figures = 6204 word-equivalents

ABSTRACT

Cities are constantly evolving, complex systems; and modeling them, both theoretically and empirically, is a complicated task. In this paper, we develop a methodology to spatially model urban areas based on a grid system of data largely derived from satellite images. The work emphasizes spatial relationships between various geographic, land-use, and demographic variables characterizing fine zones across and around regions. It derives and combines land use cover data for the Austin, Texas, region from a panel of satellite images, cartographic maps and U.S. Census of Population data. A variety of spatial attributes, including land use mix, are computed, and several land use and demographic models are run

INTRODUCTION

Urban systems are intricate, multifaceted and constantly evolving. Their evolution is dictated by a large number of influences, including public policy, individual preferences and actions, the physical landscape, technology and history. All of these factors (and more) interact in myriad ways. Discerning how and why urban systems evolve is, from the start, an extremely difficult task.

There is great benefit to uncovering the dynamics underlying urban systems. Understanding the ways in which geographic, economic, demographic, political and other factors interact is of interest to transportation engineers and land use planners, economists as well as historians, policymakers and the public. Models that reliably track these interactions illuminate how policy impacts land use and travel patterns, welfare and development, congestion and air quality, and more.

The desire to understand the evolution of urban systems certainly is nothing new. Christaller (1954) introduced “central place theory” in the 1930s, hypothesizing that urban “centers” of varying size, economic power and function, would arrange themselves according to regular, geometric patterns. Allen (1997) and Sonis (2001) have applied this concept to models of both real and theoretical cities, with some success. However, in applying such a restrictive model to the real world, embedded structural features – rather than reality – can dominate and skew the application.

In order to “grow” urban systems, Weidlich (2000) applied purely theoretical concepts from synergetics, systems, and random utility theory. Though his cities appear realistic, they are purely theoretical, and the practical application of the work is not obvious. This problem is fundamentally the same as that with central places theory: the connection between theory and reality is tenuous. Care must be taken when attempting to apply theories, calibrated on the basis of empirical data, to actual urban systems.

Parker, et. al. (2003) discussed the wide range of many land-use/cover change (LUCC) models recently developed. They pointed out that, due to the complexity of the systems encompassing land-use/cover, no one existing model is of more use than others; thus, a wide range of models, from the theoretical to the empirical, are being investigated. In this paper, a closer connection between the real world and the model, as opposed to largely theoretical work, is sought. This parallels some recent models, developed for use by planning organizations for regional forecasting and policymaking. The regional models most similar to the work undertaken here are UrbanSim, What If?, and CUF2.

UrbanSim (Waddell 2002) micro-simulates the effects of location, land use, and policy decisions by households, workers, developers and policymakers on the land use patterns and rents across a region. Land use and development is modeled at the level of single parcels. Others are related to a relatively fine grid. Klosterman’s (1999) “What if?” model of land use assigns land uses to a set of homogeneous zones in a bottom-up fashion, derived from socioeconomic, geographic, transportation and zoning information. Landis and Zhang’s (1998) California Urban Futures 2 (CUF2) model employs multinomial models of land-use change per hectare (or other unit of observation) to predict future land use patterns. These models share qualities with the models pursued here. This work’s distinction lies in its acquisition and interpretation of data.

This paper introduces a framework to analyze urban growth, relying largely on land-cover data derived from satellite images. The following sections detail the data sets developed, along with model specifications and results for an Austin, Texas, application.

DATA DESCRIPTION

Satellite data offer excellent opportunities and considerable challenges. A serious and recurring problem for modeling urban systems has been the lack of panels[1] of spatially detailed data. Remote sensing, imaging technology, and geographical information systems are making accurate land cover maps far more accessible to the researcher, and to the public. In particular, global satellite imaging, initiated in the early 1970s, provides highly detailed images regularly. And image analysis software can classify these by various general categories. GIS software combines data maps of various types, dramatically facilitating spatial analysis.

The United States launched LandSat 1 in 1972. Passing over Austin every 18 days, this early satellite provides images with 79 m x 79 m pixel resolution. LandSat 4 was launched in 1982, and resulted in 185 km x 185 km images with 30 m x 30m resolution with a repeat orbit cycle of 16 days. 1999’s LandSat 7[2] has essentially identical orbit and image characteristics to LandSat 4, with the exception of a more extensive imaging system. This system works by scanning multiple passes (each representing one pixel) over an area and recording the reflectance of six distinct spectral bands[3] (Richards and Jia 2000). Some image distortion results from the satellite's motion during scanning, the fact that the scanner is effectively a point rather than a strip, and the overlap of individual scans. Pre-processing of the image corrects much of this, for viewing on a flat surface. When comparing pixels from the same image or across images, cloud cover, time of day and sun location (vis-à-vis the satellite) also can introduce errors.

The methods of rectifying a LandSat image are distortions themselves. And they do not guarantee image accuracy, for a given projection, when imported into a GIS or other image processing software. Thus, a large amount of post-processing may be necessary by the consumer in order to further correct image distortion. In the data set used here, the LandSat 7 image had to be matched with Digital Ortho Quarter Quadrangles (DOQQ's), which are rectified aerial photographs maintained by the U.S. Geological Survey.[4]

The land-cover[5] data used here was derived from a 48.5 km x 55.8 km, 30m x 30m resolution LandSat 7 image taken at 4:30 pm on September 4, 2000. It is of Austin, Texas, and the surrounding region (Trelogan 2002). Using a process called supervised classification (see, e.g., Richards and Jia 1999), planning professor Dr. Barbara Parmenter and students created a land-cover map from the original LandSat data. Using USGS topographic maps and DOQQ's as guides, they created a set of training data by classifying sections of the LandSat image. These were used to generate a set of decision rules by which the entire LandSat image was then classified. Spatial filtering also was performed, in order to remove residual noise from the processed map (Trelogan 2002).

Each 30m x 30m pixel's resulting land-cover data is classified into 9 categories: water, barren, forest/woodland, shrubland, herbaceous natural/semi-natural, herbaceous planted/cultivated, fallow, developed residential, and developed industrial/commercial/transportation. From these, secondary land-cover maps were developed, by combining different categories into single land-cover specifications. Figure 1 shows a secondary land-cover map using three classifications: developed residential, developed industrial/commercial/transportation and undeveloped.

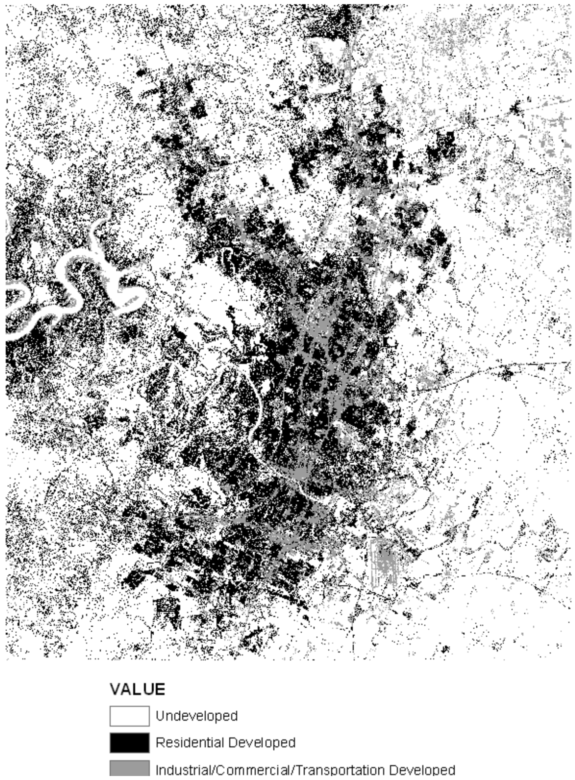

Figure 1. Secondary land-cover map of the Austin, Texas region distinguishing developed (both residential and industrial/commercial) from undeveloped land.

Using these data, one additional spatial statistic is calculated here. It is land mix, characterizing diversity in land cover. (Kockelman 1997) Mix is an index of adjacent pixels' dissimilarity; it measures the level of homogeneity between a central pixel's use type (X_0) and those of its neighbors (X_i). Here, the eight immediately surrounding pixels are considered (as in Figure 2). Mathematically, the index can be defined as follows:

$$\text{mix}(X_0) = \sum_{i=1}^{8} \frac{\delta_{X_0, X_i}}{8} \qquad (1)$$

$$\text{where } \delta_{X_0, X_i} = \begin{cases} 1 & \text{if } X_i \neq X_0 \\ 0 & \text{otherwise} \end{cases}$$

As an average measure of dissimilarity, the mix index ranges from 0 to 1, with a higher numerical value corresponding to greater distinctions between a pixel and its neighborhood. For the Austin map (containing just over 3 million pixels) this calculation results in an average mix value of 0.395, a standard deviation of 0.304, and a maximum mix of 1.

X_1	X_2	X_3
X_4	X_0	X_5
X_6	X_7	X_8

Figure 2. Reference diagram for calculation of mix statistic using pixel neighborhood.

In addition to the land cover data and its derived statistics, Census of Population data was used. Statistics from both the 100%-sample Census (SF1) and the 17% sample (SF3) were used. These include population, housing and travel variables. Data for Travis, Williamson, Bastrop and Hays Counties was collected so as to completely encompass the land-cover data region. Of course, the smallest areal unit for Census data is the block or block group, which typically encompass dozens of 30 m x 30 m pixel-based cells. For this reason, all data was transferred to a 300m x 300m resolution grid for analysis. To do this, the satellite land-cover data was first transferred from a raster to a vector ("shapefile") format[6]. The fraction of land cover by type in each of these 300m x 300m cells was computed. Entropy and mix statistics were averaged over the 100 pixels representing the new grid. Census data was transformed to this grid structure using an area-weighted average of each contributing block group's values.[7] This resulted in estimates of population, average household income and average vehicle ownership for each 300m x 300m grid cell.

DATA CAVEATS

As is the case with many data sets, there are opportunities for measurement errors in these data sets, particularly in the transformation of reflected light values to land use categories. There are many image processing steps required simply to get a satellite image into a usable form, and then several more to analyze and clean (filter) the image. Each can degrade the quality of the source image as well as the final product.

Furthermore, the supervised classification of the satellite image assumes that similar land-covers share distinctive spectral and visual characteristics. Abnormal or unusual land covers may be classified incorrectly. For example, the industrial/commercial/transportation land-cover type is predicated on presence of concrete or asphalt and larger building footprints (relative to residential and undeveloped areas). However, a residential area with many housing complexes may easily be coded as industrial/commercial/transportation, due to the presence of larger buildings and parking lots, and subsequent lack of yard-space.

Fortunately, errors caused by these shortcomings can be analyzed, in part, by a comparison of the resulting land-cover map with actual photographs (e.g., DOQQ's) or other, verified land-use information. And in this way, classification models developed from spectral information can be calibrated. While a qualitative, broad-scale analysis of the Austin map suggests accuracy, a rigorous, quantitative analysis is preferable.

Another potential source of error is in the transformation of the original raster data to a vector (shapefile) data set. It is not clear that the assumption of an equal area assigned to each raster data point is preserved in this transformation. Further analysis is required to see if errors are introduced when transferring to a larger grid scale. Lastly, the Census data is problematic in that it obscures within-block group (or tract) variation. This may be addressed, to some degree, by using land-cover data to inform the Census homogeneity assumptions for data assignment to grid cells. For example, one may be able to distinguish urban and rural portions within the same tract or block group, and/or interpolate a continuous spatial distribution of data (see Mennis 2003). The result may be a much better spatial representation of the Census data, particularly in peripheral tracts, where much of the land may be undeveloped.

MODEL SPECIFICATIONS

The models developed using this detailed data set fall into two broad categories: those that model land-cover as a response variable, and those that use land-cover as an explanatory variable.

Land-Cover Models

The first model developed splits the land-cover classifications into two categories: developed (residential and commercial/industrial/transportation) and undeveloped (all other classifications). A binary logistic formulation was used to estimate the percentage of developed land in each 300m x 300m grid cell:

$$p_i = \frac{\exp(\boldsymbol{\beta}\mathbf{X}_i)}{1+\exp(\boldsymbol{\beta}\mathbf{X}_i)} \tag{2}$$

where p_i = percentage of land developed in grid cell i

$\mathbf{X}_i$ = explanatory variables

$\boldsymbol{\beta}$ = coefficients to be estimated

Explanatory variables for this model include land use mix, population, average income, and distance to the CBD.

The next model uses the same binary formulation (as in Eq. 2) but with a response variable of residential development (i.e., the percentage of land in each grid cell developed residentially). The explanatory variables remain the same.

The final land-cover model is a multinomial model of the percentage of *each* type of land cover in the 300m x 300m cell. Four land cover designations are used: developed residential, developed commercial/industrial/transportation, undeveloped agriculture (herbaceous planted/cultivated & fallow), and undeveloped other. The specification is as follows:

$$p_{ij} = \frac{\exp(\boldsymbol{\beta}\mathbf{X}_{ij})}{\sum_j \exp(\boldsymbol{\beta}\mathbf{X}_{ij})} \tag{3}$$

where p_{ij} = percentage of land developed as type j in grid cell i

$\mathbf{X}_i$ = explanatory variables

$\boldsymbol{\beta}$ = coefficients to be estimated

Demographic Models

The models of Census information per 300m x 300m cell use land cover information as explanatory input. The first models population as a linear function of land-cover, land use mix, household income, and distance to CBD. To constrain the response variable (population) to positive values, the natural log form is used:

$$\ln(\text{population}_i) = \boldsymbol{\beta}\mathbf{X}_i + \varepsilon_i \tag{5}$$

The second demographic model examines average household vehicle ownership in terms of many of these same variables, and Equation 5 (replacing population with vehicle ownership).

ANALYTICAL RESULTS

The results of the percent developed/undeveloped binary logit model are presented in Table 1. As expected, the percentage of a grid cell that is developed increases as the population increases and decreases the farther the cell is from the CBD. The square roots of population and distance-to-CBD variables are used because the relative effect of an increase in these variables on the developed fraction is expected to "flatten out" as the values of these variables gets larger.

As the average household income increases, the developed fraction also increases. This may be due to higher land values in more dense and popular areas, requiring higher household incomes to cover home prices. As an area's land-cover dissimilarity (i.e., mix) increases, development is predicted to fall. This is probably an artifact of the mix definition used here, which relies on a variety of undeveloped land-cover categories, rather than a mix of more developed land uses.

Table 2 presents the results of the percent residential/not-residential binary logit model. Similar to the previous model, the percentage of a grid cell that is residential increases as population, income, and mix increase. As distance to the CBD increases, however, this percentage falls. This model's predictive power (or LRI) is lower than the first model's, suggesting that residential development is harder to predict than is total development, probably due to the additional specificity in such a measure.

The results of the four multinomial logit land-cover models are shown in Table 3. As expected, as population increases, the percentage of a grid cell that is residential or commercial increases, while the percentage that is agricultural decreases. Also, as the distance to the CBD increases, the percentage of a grid cell that is residential decreases and the percentage that is agricultural increases. Interestingly, the distance to the CBD has no statistically significant effect on the percent that is commercial. Again, the square roots of the population and distance to CBD variables are used to account for the "flattening out" of these variables effects as they increase.

As household income increases, the percentage of a grid cell that is residential increases (as noted earlier), while the percentages that are commercial and agricultural fall. This may be due to households not desiring to reside too close to commercial and agricultural land uses, though these certainly can hold many benefits. Finally, as land mix increases, the percentage that is residential increases, but the fractions that are commercial and agricultural decrease.

The results of the population regression model are presented in Table 4. As expected, the population per grid cell decreases the farther the grid cell lies from the CBD but rises with the percentage of developed land. Though the coefficient on the percentage of undeveloped land variable is also positive, it is smaller than that for the percentage of developed land, thus exerting a negative effect overall – unless the undeveloped land is converted from a water-covered area (rather than from a developed-land area). Lastly, as average mix increases, the population is predicted to fall, for reasons similar to those discussed earlier (for the percentage of area that is residentially developed).

The results of the model of average household vehicle ownership are shown in Table 5. As expected, as income and the distance to the CBD from a grid cell increases, the average number of vehicles per household increases. This is most likely due to a wealth effect and the fact that households farther from the center of the city will be more dependent on their vehicles. Interestingly, as land mix increases, so does the vehicle ownership, perhaps due to less connectivity between developed land uses requiring greater use of motorized modes. A scatter plot comparing the actual versus predicted average household vehicle ownership is shown in Figure 3. Clearly, mispredictions are prevalent for low-vehicle-owning areas; perhaps some key explanatory variables are missing to explain this model bias.

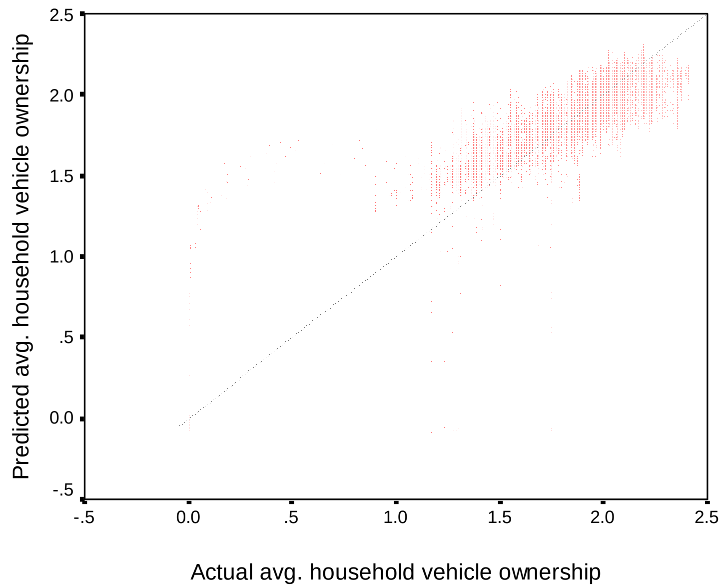

Figure 3. Comparison of actual versus predicted results for the ln(average household vehicle ownership) regression model. (The dark line is a 45^{o} reference line, representing a perfect fit.)

CONCLUSIONS AND EXTENSIONS

This paper develops methods for interpreting and modeling the texture and pattern of an urban region based on grid systems of data largely derived from satellite images. The methods emphasize land use patterns and relationships between various geographic, land-use, and demographic variables characterizing fine zones across and around the Austin, Texas region. The work combine land use cover data from satellite images and cartographic maps with U.S. Census of Population and network data. A variety of spatial attributes, including land use mix, are computed, and predictive models of land use and demographic models are specified.

The results of this work show that land-cover characteristics derived from satellite imagery can be used in modeling with good results. In the future, this work will be extended to include historical data so that growth can be tracked and future projections can be made. Also, more spatial and geographical information, including land-balance (entropy) and distance to major highway statistics.

The skies and their satellites hold a wealth of information for land use-transportation modelers. Though presently difficult to manage, due to size and complexity, satellite data sets, combined with traditional survey-based data sources, herald a new era in regional forecasting. This work is a step toward that future.

ACKNOWLEDGEMENTS

The authors wish to thank The University of Texas at Austin Research Internship program and the NSF CAREER Award program (Grant # 9984541) for providing financial support to this project. Additionally, we wish to express special thanks to Dr. Barbara Parmenter from the School of Architecture at The University of Texas, Austin for providing us with the land-cover data set as well as information concerning its derivation; and Dr. Melba Crawford, Director of the University's Center for Space Research and faculty member in the Department of Mechanical Engineering, for her helpful discussions concerning remotely-sensed data. We would also like to thank Michael Reyes for his help in preparing and analyzing the data.

ENDNOTES

[1] A panel data set consists of the same observational unit (or set of units), such as the Austin region, over a period of discrete time points.

[2] We focus here on LandSat 7 as it is that from which we derived our data. Images from other remote sensing satellites can be used and, though the discussions here are specific to LandSat 7, the general warnings and problems are common to all remote sensing systems (see Richards and Jia (1999) for more information).

[3] A seventh, thermal band is also recorded, but with 120m x 120m resolution.

[4] For information concerning the USGS Digital Orthophoto Program, one may visit http://mapping.usgs.gov/www/ndop/. DOQQ's may be ordered through this site, though many are available free through individual state's websites. For example, Texas' DOQQ files are available through the Texas Natural Resources Information System (http://www.tnris.state.tx.us/DigitalData/doqs.htm).

[5] Though this data is often called land use, technically it constitutes land-cover, since it is derived from the spectral (surface) qualities of the data, and its accuracy, with respect the "use" of the land, has not been verified.

[6] Raster data consists of discrete points corresponding to a grid, whereas vector data essentially consists of lines and polygons which do not necessarily have an underlying structure (such as a grid). (See Richards and Jia (1999) for a more detailed definition.)

[7] All census data was collected at the block group level, with the exception of "average number of vehicles per household," which could only be ascertained at the tract level. The transfer of this variable to the grid was identical to that of other Census statistics except tracts, rather than block groups, were used as the spatial reference.

REFERENCES

Allen, Peter M. *Cities and Regions as Self-Organizing Systems: Models of Complexity*. Gordon and Breach, Amsterdam. 1997.

Christaller, W., *Die zentralen Orte in Suddeutschland*. Gustav Fischer, 1933. Trans. C. Baskin, Bureau of Population and Urban Research, University of Virginia, 1954.

Fujita, Masahisa. *Urban Economic Theory: Land Use and City Size*. Cambridge U.P. 1989.

Herold, Martin. "Comparison of Land Use Models with Focus on Spatial and Temporal Frameworks and Data Issues." Presentation at Land-Use Change Modeling Seminar at The University of California, Santa Barbara, 2001. (URL: http://www.geog.ucsb.edu/~kclarke/ucime/Helens-Sem/seminar2001/student-com-pres/mherold.pdf Accessed July 29, 2003)

Jaynes, E.T. *Probability Theory: The Logic of Science*. "Fragmentary Edition.1994." (URL: http://omega.albany.edu:8008/JaynesBook. Accessed July 15, 2003.)

Klosterman, R.E. "What if?: Collaborative Planning Support System," *Environment and Planning B: Planning and Design*, 26. 1999, pp. 393 - 408.

Kockelman, Kara M. "Travel Behavior as a Function of Accessibility, Land Use Mixing, and Land Balance: Evidence from the San Francisco Bay Area," *Transportation Research Record 1607*. 1997, pp. 117 - 125.

Landis, J. and M. Zhang. "The Second Generation of the California Urban Futures Model: Part 1: Model Logic and Theory," *Environment and Planning B*, 30. 1998, pp. 657 - 666.

Mennis, Jeremy. "Generating Surface Models of Population Using Dasymetric Mapping," *The Professional Geographer*, 55:1. 2003, pp. 31 - 42.

Parker, Dawn C., Steven M. Manson, Marco A. Janssen, Matthew J. Hoffmann and Peter Deadman. "Multi-Agent Systems for the Simulation of Land-Use and Land-Cover Change: A Review," *Annals of the Association of American Geographers*, 93:2. 2003, pp. 314 - 317.

Richards, John A., and Xiaping Jia. *Remote Sensing Digital Image Analysis: An Introduction*. Springer-Verlag, Berlin. 1999.

Sonis, Michael. "Complication and Complexity in Dynamics of Linear Systems in Economic Geography and Regional Science." Working paper/manuscript presented at 2001 North American Meetings of the Regional Sciences Association International, Charleston, South Carolina.

Trelogan, Jessica. Html file titled "tm00_blob Metadata." (Data description for student interpreted land-use/land-cover data for Travis County.) 2002.

Weidlich, Wolfgang. *Sociodynamics: A Systematic Approach to Mathematical Modeling in the Social Sciences.* Harwood Academic, Amsterdam. 2000.

Waddell, Paul. "UrbanSim: Modeling Urban Development for Land Use, Transportation, and Environmental Planning," *The Journal of the American Planning Association*, 68:3. 2002, pp. 297 - 314.

List of Tables and Figures

Table 1. Results for Percent Developed Binomial Logit Model.

Explanatory Variables	***β***	***t-statistic***	***p-value***
Constant specific to % of land developed	-2.144	-22.746	0.0000
Square root (Population)	0.278	57.168	0.0000
Square root (Distance to CBD)	-0.101	-6.988	0.0000
Average mix	0.103	8.812	0.0000
Average household income	2.89E-06	5.653	0.0000
Log-likelihood	-15384.9180		
Log-likelihood constants	-18517.7900		
Log-likelihood ratio index	0.1692		
# of observations	29945.0000		

Note: Base case is % of land not developed.

Table 2. Results for Percent Residential Binomial Logit Model.

Explanatory Variables	***β***	***t-statistic***	***p-value***
Constant specific to % residential	-3.265	-30.793	0.0000
Square root (Population)	0.224	48.166	0.0000
Square root (Distance to CBD (km))	-0.0820	-5.199	0.0000
Average mix	0.1090	8.590	0.0000
Average household income	1.27E-05	23.219	0.0000
Log-likelihood	-13464.4953		
Log-likelihood constants	-15546.2635		
Log-likelihood ratio index	0.1339		
# of observations	29945		

Note: % of land not residential is the base case.

Table 3. Results for Multinomial Logit Land-Cover Model.

Explanatory Variables	***β***	***t-statistic***	***p-value***
Constants			
% of land residential	-2.84	-26.236	0.0000
% of land commercial	-1.89	-19.785	0.0000
% of land agriculture	-0.820	-6.621	0.0000
Square root (Population) (residential)	0.287	53.825	0.0000
Square root (Population) (commercial)	0.235	42.382	0.0000
Square root (Population) (agricultural)	-0.0663	-6.878	0.0000
Square root (Distance to CBD (km)) (residential)	-0.0678	-4.249	0.0000
Square root (Distance to CBD (km)) (agricultural)	0.216	10.883	0.0000
Average mix (residential)	0.085	6.364	0.0000
Average mix (commercial)	-1.82E-05	2.545	0.0109
Average mix (agricultural)	-0.203	-14.099	0.0000
Average household income (residential)	7.59E-06	13.274	0.0000
Average household income (commercial)	-1.82E-06	-18.851	0.0000
Average household income (agricultural)	-1.54E-05	-16.605	0.0000
Log-likelihood	-29897.7664		
Log-likelihood constants	-33784.5022		
Log-likelihood ratio	0.1150		
# of observations	29945		

Notes: % of land non-developed and non-agriculture is base case.
The identifiers in parentheses after the explanatory variables denote which choice(s) they are specific to.

Table 4. Results for ln (Population) Regression Model.

Explanatory Variables	***β***	***t-statistic***	***p-value***
Constant	4.203	67.449	0.0000
ln (Distance to CBD (km))	-0.846	-75.289	0.0000
% of land developed	2.587	46.305	0.0000
% of land not developed (not including water areas)	0.38	7.044	0.0000
Average mix	-0.061	-14.101	0.0000
R^2	0.532		
# of observations	29945		

Table 5. Results for Log (Available vehicles per household) Regression Model.

Explanatory Variables	***β***	***t-statistic***	***p-value***
Constant	-0.064	-11.593	0.0000
Distance to CBD (km)	0.003	80.500	0.0000
ln (% of land developed)	-0.061	-12.256	0.0000
ln (% of land not developed (not including water areas))	0.027	5.090	0.0000
Average mix	0.005	16.523	0.0000
Log (Average household income)	0.096	242.309	0.0000
R^2	0.734		
# of observations	29945		