

Descriptive and Comparative Studies of 1990 Urban Extent Data for the New York Metropolitan Region

Ann-Margaret Esnard and Yizhao Yang

Abstract: Regional planners and decision makers are teaming up with data providers and urban growth modelers in the ongoing effort to map historic extents of urbanization, forecast future extents, and predict impacts on the local environment. Spatial data layers that are nationally consistent across space and time are valuable input variables for some predictive models, but can be of limited use because of varying definitions of urban land and methods of data capture. This article reports on a descriptive and comparative study of two national data sets for mapping the extent of 1990 urbanization for the New York-New Jersey-Connecticut metropolitan region: 1) urban land classes based on Multi Resolution Land Characteristics data, and 2) block group urbanized area data from the United States Census Bureau. As multiple data sets from the earth, natural and social sciences proliferate and are injected into models, it is imperative that researchers and practitioners conduct and share findings of such comparative studies as part of the emergent metadata culture.

Background

Two generations of decentralized growth have dramatically expanded the urban portion of the New York-New Jersey-Connecticut tri-state region. In the last 30 years, the amount of urban land has increased by more than 60% despite only a 13% increase in population. (Regional Plan Association, 1998)

The disparity between population growth and land consumption, spatially and temporally, has made research on urban sprawl an issue of increased notoriety. Furthermore, the direct impacts on the physical, ecological, natural, and cultural resources are compounded by the fiscal costs associated with building new infrastructure and public services in remote developments. Since 1999, a multidisciplinary team from the United States Geological Survey, the Regional Plan Association, and Cornell University's Department of City and Regional Planning have collaborated to create digital views ranging from the 1930s to 1990s for the tri-state metropolitan region. This collaborative project is part of the United States Geological Survey's Urban Dynamics Research Program¹ and builds on other similar efforts throughout the U.S. The study region includes five counties in Connecticut, nine counties in New York State, the five boroughs comprising New York City, and 14 counties in New Jersey (Table 1).

Urban Growth Modeling for the NY-NJ-CT Metropolitan Region

In general, urban dynamics research intends to evaluate, utilize and enhance predictive models that at a minimum would allow urban planners, local government officials and the general public to visualize future urbanization growth patterns and potential impacts on the local environment. One such model, the SLEUTH model, also called the Clarke Cellular Automata Urban Growth Model and the Clarke Urban Growth Model, has six data inputs: slope, land use, exclusion, urban, transportation, and hill shading (Clarke et al. 1997, U.S. Environmental Protection Agency 2000). An underlying assumption is that historic growth trends will continue and that the future can be projected based on these trends.

In the SLEUTH model, urban land² is defined as residential, commercial, mixed use, and industrial land uses (U.S. Environmental Protection Agency 2000) and for the 1930s to 1990s was generally compiled using aggregated, multiple source scanned graphics and digital vector and image data from historic maps, satellite imagery, aerial photos, Census statistics, and commerce records (Acevedo et al. 1996). At least four urban time periods are required, with one defined as the "seed" or start year from which to predict urban expansion. In the case of the study region, the seed year is 1990 and the main sources of data have been the U.S. Geological Survey, the Regional Planning Association, the New York Public Library, and the U.S. Census Bureau.

Table 1 Counties and boroughs in study region		
Connecticut	New York	New Jersey
Fairfield, Hartford, Litchfield, Middlesex, New Haven	<i>Boroughs:</i> Borough of Queens (Queens County), Borough of Manhattan (New York County), Borough of Brooklyn (Kings County), Borough of the Bronx (Bronx County), Borough of Staten Island (Richmond County). <i>Counties:</i> Dutchess, Nassau, Orange, Putnam, Rockland, Suffolk, Sullivan, Ulster, Westchester	Bergen, Essex, Hudson, Hunterdon, Mercer, Middlesex, Monmouth, Morris, Ocean, Passaic, Somerset, Sussex, Union, Warren

Descriptive and Comparative Data Studies as an Important First Step

In setting out to create the 1990 urban extent snapshot for the study region, a choice had to be made between two national data sets: 1) urban/developed land classes based on Multi Resolution Land Characterization (MRLC) data captured and classified by remote sensing and spectral procedures, and 2) Census block group³ urbanized area data based on a population density threshold.

Concerns that spectral classifications do not adequately identify urban extent, particularly in heterogeneous urban fringes (Vogelmann et al. 1998, Zhu et al. 2000), and that Census-defined urbanized areas and the proxy measure of population density do “not measure the structures on land and does not account for the commercial, industrial or transportation components of urban land use” (Hitt 1994, p. 14)⁴ are well founded.

The choice was made based on these and other reported differences between land cover and land use and the common practice of employing land cover to study physical processes and land use to study cultural and economic processes (Dobson 1993). The dilemma is that urbanization and sprawl, like many other phenomena, have physical, social, economic, and cultural dimensions.

The literature provides limited information on: 1) the correlation between the two data sets, particularly land-use intensities in MRLC data and population density in Census data, 2) whether combining them resolves some of the known limitations, and 3) the appropriateness of both data sets for multi-state regional analyses and urban growth modeling efforts.

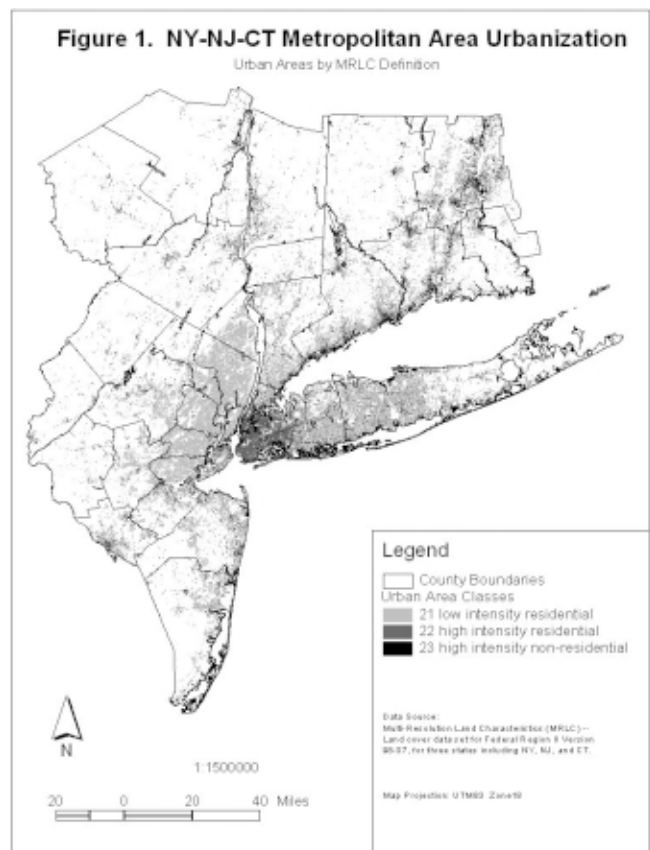
This article attempts to partially fill this information gap by reporting on the results of a descriptive and comparative study of the two national data sets previously mentioned. The impetus for these pre-modeling studies is perhaps best captured by a statement by Briassoulis:

Given all the constraints surrounding the proper use of theories and models, the elite who possess the requisite education and skills has an ethical obligation to guide the users of theories and models, i.e. the actual decision makers, to making wise use of them. It rests, therefore, with those individuals who “control” the available information to assure sensible and appropriate use (Briassoulis 2000, section 5.14 on website)

Urban Land in Multi Resolution Land Characterization data

MRLC is a generalized land cover (30-meter spatial resolution) data set created using a multiple-layer land-characteristics approach with Landsat thematic mapper as the main data source. Vogelmann et al. (1998) provide details of compilation procedures for U.S. Federal Region III, which includes the states of Pennsylvania, Virginia, Maryland, Delaware, and West Virginia. Similar procedures were used for Federal Region II, which includes the study region.

The MRLC data are broken down into 15 classes; however, for the purposes of this study, land cover Classes 21 (low-intensity residential), 22 (high-intensity residential) and 23 (high-intensity commercial/industrial/transportation) were selected to represent the extent of urban land (Figure 1). In general, these areas are char-



acterized by a higher percentage of construction materials (e.g., asphalt, concrete, and buildings). Undeveloped land that is completely surrounded by developed areas (e.g., cemeteries, golf courses, and urban parks) is not differentiated within these developed lands (<http://landcover.usgs.gov>, accessed May 2002).

Class 21: low-intensity residential – includes areas with a mixture of constructed materials and vegetation or other cover. Constructed materials account for 30-80% of the total area.

Class 22: high-intensity residential – includes heavily built-up urban centers where people reside. Constructed materials account for 80-100% of the total area.

Class 23: high-intensity commercial/industrial/transportation – includes all highly developed lands not classified as high-intensity residential.

According to Vogelmann et al. (1998, p. 51), “ancillary data was used to aid in the class labeling procedure and to split clusters into discrete land cover classes.” The ancillary data include digital elevation data, population Census information, defense meteorological satellite program city lights data, soil, prior land-use and land-cover data, leaf on/leaf off normalized-difference vegetative index, digital line graph data, and national wetlands inventory. According to Sohl (2000), the population Census data were primarily used to split urban forests (residential areas with heavy

Table 2. Aerial coverage (square miles) by land use intensity

	Class 21 Low Intensity Residential(sq. miles)	Class 22 High Intensity Residential(sq. miles)	Class 23 High Intensity Commercial/ Transportation/ Industrial(sq. miles)	Total (sq. mile)
Area of Urbanized BGs *1	2697	258	74	3029
Area of MRLC *2	2365	396	371	3132
Total developed land area in urbanized BGs *3	1883	226	52	2161
Area of MRLC and urbanized BGs spatially combined	3193	429	394	4016

*1 – An entire urbanized BG was classified as one type of developed land class (i.e. 21, 22 or 23) based on the majority developed land class in that urbanized BG.
 *2 – Area of the MRLC developed land class (21, 22 or 23) in the entire study region
 *3 - Actual developed area (21, 22 and 23) in the urbanized BG.-

tree cover) and real forest (forest land) that have similar vegetative signatures, and to differentiate Class 23 from Classes 21 and 22.

Urbanized Areas In Block Groups

Population density based on Census statistics is commonly used to present snapshots of the extent of urbanization and suburbanization. The Census Bureau defines urbanized areas as continuously built up areas with a population of 50,000 or more. Population density of at least 1000 people per square mile is one of the definitional criteria, the threshold value having remained the same since it was adopted for the 1960 Census (U.S. Bureau of Census Geographic Areas Reference Manual).

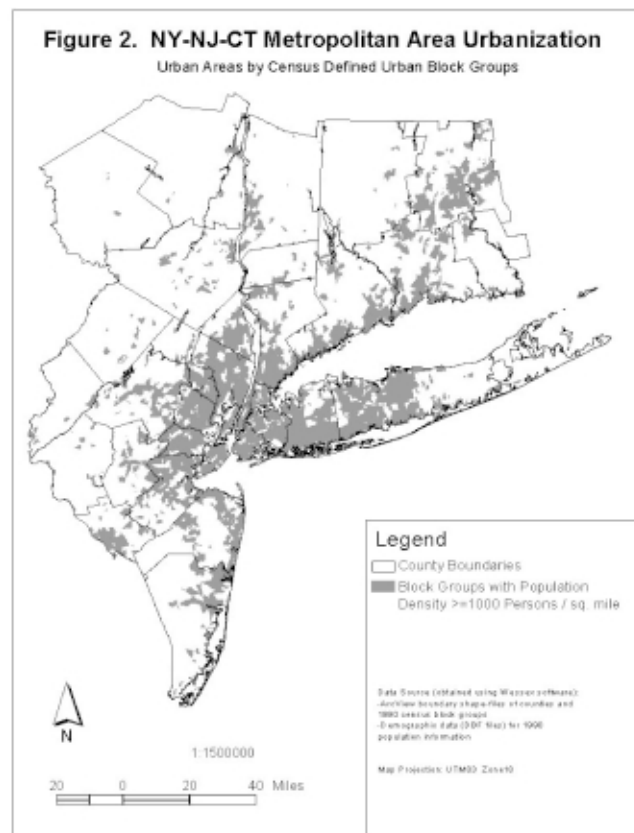
Block groups (BGs) were used to encode population density since they were the smallest Census summary level consistently available from multiple data providers⁵ and for the entire study region. Urbanized block groups were selected using the population density threshold of greater than 1000 people per square mile. The spatial extent of Census-defined urbanized areas is shown in Figure 2.

Observations and Interpretation

The overall match and correlation between land-use intensities and population densities were examined by summarizing aerial coverage and population data for: 1) urbanized BGs, 2) MRLC developed land⁶, 3) developed land in urbanized BGs, and 4) the spatial combination of urbanized BGs and MRLC developed land (See figure 3). The findings are presented in Tables 2 and 3.

Overall Match Between the Two Data Sets: A Focus on Aerial Coverage of Urbanized Block Groups and MRLC-Developed Land Classes

Class 23: MRLC captures more “high-intensity commercial/transportation/industrial” land (371 square miles) when compared to the aerial coverage of urbanized BGs (74 square miles) in this category (see Table 2). Two explanations are that urbanized BGs do not capture these non-residential land uses, and/or MRLC is



overestimating urban land as part of this developed land class. For example, a closer examination of the two data sets revealed that transportation routes spread into remote areas and industrial sites are picked up by MRLC.

It may be acceptable to identify such areas as “developed” in the sense that they are “man-made” as opposed to “natural;” however, such physical transformation may not be a sufficient qualification in some urban growth models for defining urbanized areas. Given the ambiguity in defining urbanization, it is necessary to review model assumptions and simulation rules. For example, in the case of the SLEUTH model, road-influenced growth is one of four types of growth being simulated (U.S. Environmental Protection Agency, 2000). Therefore, it is important to

Table 3 Urbanization extent in relation to population distributions among land-use intensities

	Type of Urbanized BGs	Population in Urbanized BGs	Percent of Population	Developed Land in Urbanized BGs (sq. miles)	Percent of Total Developed Land	Avg. Population Density	
NJ	Low-intensity Residential	5,041,379	95.48	783.47	96.43	10,540	
	High-intensity Residential	214,302	4.06	23.99	2.95	10,898	
	High-intensity Commercial/industry/transportation	24,415	0.46	4.98	0.61	6,177	
NY	Low-intensity Residential	4,930,933	44.41	747.45	80.79	17,404	
	High-intensity Residential	5,886,223	53.02	160.92	17.39	53,043	
	High-intensity Commercial/industry/transportation	285,751	2.57	16.78	1.81	33,448	
CT	Low-intensity Residential	1,401,965	69.59	351.9	83.16	4,093	
	High-intensity Residential	462,500	22.96	40.97	9.68	13,077	
	High-intensity Commercial/industry/transportation	150,031	7.45	30.31	7.16	6,065	
Entire Metro Region	Low-intensity Residential	11,374,277	61.83	1,882.82	87.14	9,629	
	High-intensity Residential	6,563,025	35.67	225.88	10.45	35,151	
	High-intensity Commercial/industry/transportation	460,197	2.50	52.07	2.41	15,614	
		Total population in urbanized BGs: 18,397,499					
		Total developed land in urbanized BGs: 2,161 sq. miles					

include data sets that capture transportation networks.

Class 22: Urbanized BGs capture less high-intensity residential land area. One explanation is that the averaging of population densities across entire BGs fails to specifically identify some high-intensity residential hot spots compared to the MRLC spectral classification.

Class 21: MRLC captures less low-intensity residential land area. Two explanations are the inability of MRLC to capture urban fringes due to the spatial resolution limit and/or the inclusion of the entire area of urbanized BGs despite the fact that development only took place in a small fraction of the BG.

Overall Correlation between the Two Data Sets: A Focus on Developed Land Area within Urbanized BGs

Class 21: There is a discrepancy in the aerial coverage of urbanized BGs in their entirety (2697 square miles) and the “developed land” in BGs (1883 square miles) – that is, only 70% of the land in all urban-

ized BGs in the study region identified by the population density criterion was captured in MRLC as developed. A covering percentage metric (the ratio of MRLC urban land area to the urbanized area of the block group) was used to examine this discrepancy more closely.

Results showed that there is a large variation between the extreme covering percentage values in each land-cover class, with minimum values as low as 1.27%, maximum values reaching 100%, and average values in the 73-85% range. More than 10% of the urbanized BGs had a covering percentage below 50% and only 2.5% were totally built up (i.e., 100% covering percentage) by 1990.

Urbanization Extent in Relation to Population Distributions among Land-Use Intensities: A Focus on Regional Variations

For all urbanized BGs in the study region in 1990, 62% of the total population were found in low-intensity residential BGs, but 87% of the total developed land was consumed for this intensity

of land use. This can be compared to the population distribution and land consumption in high-intensity residential BGs – 36% of the population consumed about 10% of the total developed land. This disproportional population distribution in relation to land consumption prompted a closer look at the intra-regional conformance or variation from this pattern.

In Table 3, the summary statistics for the New York counties and boroughs and the New Jersey counties in the study region show two contrasting patterns of urbanization:

- In New York, a much lower proportion of the population resided on land developed as low-intensity residential, compared to New Jersey.
- Comparing values of “percent of population” and “percent of total developed land,” land consumption is more directly proportional to population distribution among different classes of land-use intensity, in the New Jersey counties. In the case of New York, there are distinct differences in values among the low and high-intensity residential land uses. This contrast can be explained by taking a closer look at average population density, which shows lower and similar values across the land-use intensities in New Jersey, compared to that for New York.

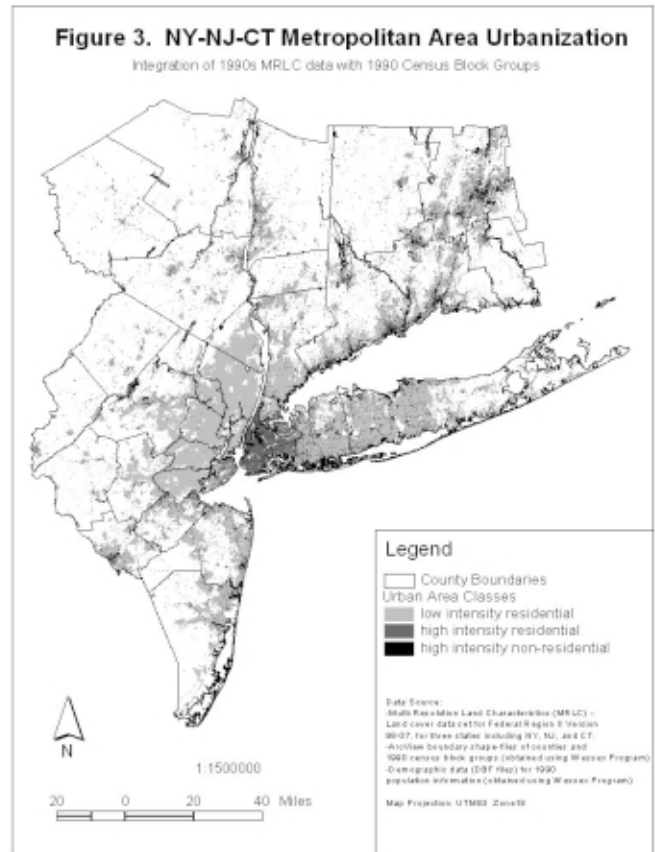
These observations highlight the obvious limitations of trying to generalize theories, trends, and patterns of urbanization patterns at the regional level and the hidden details in broadly categorized urban and non-urban land uses. There is also a need for other metrics such as employment densities, building types, and floor area ratio to explain some of the anomalies and discrepancies between population density, construction intensity, and covering percentage. For example, BGs with similar population densities and construction intensities could have different covering percentage of developed land due to differences in floor area ratio. A higher floor area ratio (such as high rises) can potentially reduce the horizontal urban extent.

Conclusions and Implications

This study confirmed the limitations of the two data sets that can be used to map the 1990 extent of urbanization for metropolitan regions. Overall, the advantage of combining the two data sets lies in the findings that MRLC can compensate for commercial/industrial and transportation mixed uses, while urbanized BGs can compensate for low-intensity residential land use. However, it is also important to understand model assumptions and simulation rules in deciding between individual or combined data sets.

One of the important messages/reminders was the need to understand regional data sets and inter-regional variations in concert with model results, given the differing concepts of urban land and different patterns of urbanization. Such “data studies” and the questions that they prompt provide requisite knowledge for model-output interpretation and signs of faulty output.

Beyond this specific case study, prevailing practices make these types of “data studies” relevant to geographic information system professionals, data providers, database designers, model develop-



ers, and policy makers. First, multiple data sets from the physical, earth, natural, and social sciences will continue to proliferate and be injected into models that have applicability to socio-economic-natural-demographic phenomena such as urbanization and intra-regional variation. Second, model users from multidisciplinary groups will continue to interpret commonly used terms such as “user-defined data sets” and to choose data based on levels of familiarity or widespread use in their fields of expertise.

It is imperative that researchers and practitioners conduct and share findings of descriptive and comparative studies on data sets as part of the emergent metadata culture. At the very least, it is a basic ethical obligation.

Authors

Dr. Ann-Margaret Esnard is Assistant Professor in the Department of City and Regional Planning at Cornell University. Dr. Esnard’s research and teaching interests are in the areas of land use and environmental planning, natural and environmental hazard assessment and mitigation, spatial analysis, and geographic information systems.

Corresponding address:

Dr. Ann-Margaret Esnard
219 West Sibley Hall
Cornell University
Ithaca, New York, 14853
ame7@cornell.edu

Ms. Yizhao Yang is a doctoral student in the Department of City and Regional Planning, Cornell University. Ms. Yang's research interests include residential location, land-use change analysis, and multi-disciplinary geographic information system applications.

Corresponding address:
Ms. Yizhao Yang
106 W. Sibley
Cornell University
Ithaca, New York, 14853
yy55@cornell.edu

Acknowledgments

The study in this article is one part of a collaborative project on Temporal Mapping in the NY-NJ-CT metropolitan region between the U.S. Geological Survey, the Regional Plan Association, and Cornell University. We wish to acknowledge the input of collaborators Robert Pirani of the Regional Planning Association and Dan Sechrist, Janet Tilley and Roger Barlow of the U.S. Geological Survey. Thanks also to Terry Sohl (affiliated with Raytheon ITSS Corp. at the time of this study) for providing detailed clarifications on questions related to the creation of MRLC data, and to the reviewers and editors for their helpful comments and guidance. The views expressed in the article do not necessarily reflect that of our collaborators.

References

- Acevedo, W., T.W. Foresman, and J.T. Buchanan, 1996, *Origins and Philosophy of Building a Temporal Database to Examine Human Transformation Processes*, <http://landcover.usgs.gov/urban/wma/pubs/asprswmap.html>, accessed May 2002
- Briassoulis, H., 2000, *Analysis of Land Use Change: Theoretical and Modeling Approaches* (West Virginia University: Regional Research Institute), <http://www.rri.wvu.edu/WebBook/Briassoulis/contents.htm>, accessed May 2002
- Clarke, K.C., L.J. Gaydos, and S. Hoppen, 1997, A Self-Modifying Cellular Automaton Model of Historical Urbanization in the San Francisco Bay Area. *Environment and Planning B*, 24, 247-261.
- Dobson, J.E., 1993, Land Cover Land Use Differences Distinct. *GIS World*, February, 21-22.
- Hitt, K.J., 1994, Refining 1970's Land Use Data with 1990 Population Data to Indicate New Residential Development. *U.S. Geological Survey Water-Resources Investigations Report 94-4250*, 15 pp, <http://water.usgs.gov/nawqa/WRI94-4250>, accessed May 2002

- Kaiser, E.J., D.R. Godschalk, and F.S. Chapin, 1995, *Urban Land Use Planning, Fourth Edition* (Urbana, IL: University of Illinois Press).
- Regional Plan Association, 1998, *Building a Metropolitan Greenward: New York-New Jersey-Connecticut Metropolitan Region* (New York: Regional Planning Association).
- Sohl, T., 2000, *Email Exchange with A.M. Esnard*, June 27.
- U.S. Bureau of Census, Geographic Areas Reference Manual, <http://www.census.gov/geo/www/garm.html>
- U.S. Environmental Protection Agency, 2000, *Projecting Land-Use Change: A Summary of Models for Assessing the Effects of Community Growth and Change on Land-Use Patterns. EPA/600/R-00/098* (Cincinnati, Ohio: U.S. Environmental Protection Agency, Office of Research and Development), 260 pp.
- Vogelmann, J.E., T. Sohl, and S.M. Howard, 1998, Regional Characterization of Land Cover Using Multiple Sources of Data. *Photogrammetric Engineering and Remote Sensing*, 64(1), 45-57.
- Zhu, Z., L. Yang, S.V. Stehnan, and R.C. Czaplewski, 2000, Accuracy Assessment for the U.S. Geological Survey Regional Land Cover Mapping Program: New York and New Jersey Region. *Photogrammetric Engineering and Remote Sensing*, 66(12), 1425-1435.

Endnotes

- ¹ The Urban Dynamics Research Program web site is <http://landcover.usgs.gov/urban/intro.html>, accessed May 2002
- ² Urban land is used in some contexts to mean built-up or developed land. In the land classification scheme described by Kaiser et al. (1995), developed land classes are characterized by lands currently developed for urban purposes with urban service available.
- ³ The U.S. Census Bureau defines a block group as a combination of Census blocks that is a subdivision of a Census tract or a Block Numbering Area (BNA). For the 1990 Census, enumeration guidelines specified an ideal size for a block group of 400 housing units, with a minimum of 250 and a maximum of 550 housing units and an average population of 1000 people. The guidelines further required that block group boundaries follow clearly visible features such as roads, rivers, and railroads.
- ⁴ In research by Hitt (1994), there was a further breakdown by housing density (e.g., low, medium, intermediate, high and very high density residential, and rural development) to facilitate the study of the impacts of point and non-point source pollution on water quality.
- ⁵ We also looked at TIGER files and Census attribute tables available from the U.S. Census Bureau's web site and from Cornell University Geographic Information Repository (<http://cugir.mannlib.cornell.edu/>).
- ⁶ A block group was designated as the same type of developed land to which the majority of Multi Resolution Land Characteristics grid cells in that block group belong.