Techniques for Mapping Suburban Sprawl

Jeanne Epstein, Karen Payne, and Elizabeth Kramer

Abstract

The pervasive problems generated by urban sprawl have prompted us to examine methods for delineating the extent of suburban land cover in Georgia. This paper assesses the advantages and disadvantages of two different methods of mapping suburban neighborhoods: traditional unsupervised classification of Landsat 5 TM data and a newly devised procedure for editing and buffering road coverages. We conclude that, while the amount of time required to edit and buffer road coverages is significantly higher than that for traditional remote sensing techniques, the improved thematic accuracy, spatial contiguity, and potential future uses of the resulting dataset justifies its use in a state-wide mapping program.

Introduction

Recently, Georgia has garnered national attention because its urban areas are experiencing some of the highest growth rates in the country. New subdivisions are appearing around all of Georgia's major cities at an unprecedented pace. Many municipalities are ill-equipped to handle problems associated with this growth: burgeoning traffic problems, air and water pollution, and loss of greenspace for farms and recreation. In this context, accurate information on the current extent of urban areas is needed for documenting growth, making policy decisions, and improving land-use planning (Bullard and Johnson, 1999; Gross, 2000; Knickerbocker, 2000; Jacobson, 2001), and is a required parameter for predictive land-use modeling.

As part of the Georgia Gap Analysis Program (GA-GAP) and the Georgia Land-Use Trends Analysis Program (GLUT), we generated a 1998 land-cover map of the state. In our protocol, we paid particular attention to methods for mapping suburban sprawl and created the first state-wide coverage of suburban sprawl in Georgia using readily available data: a 1:12,000-scale road coverage map, 1993 aerial photos, and Landsat Thematic Mapper (TM) data.

Existing remote sensing methods for mapping urban sprawl have focused on either traditional maximum-likelihood classifiers (Toll, 1985) or a combination of change-detection algorithms using multi-temporal imagery to delineate new urban areas. In particular, vector analysis (Haack *et al.*, 1987), image differencing (Jensen and Toll, 1982; Griffiths, 1988), the Normalized Difference Vegetation Index (NDVI) (Howarth and Boasson, 1983; Masek *et al.*, 2000) and Principle Components Analysis (Yeh and Li, 2001), have all been used to evaluate expanding urban boundaries. Geographic information system (GIS) based approaches for delineating urban sprawl generally rely on visual inspection and digitization of high resolution aerial photos and may include post-classification analysis (Gordon, 1980; Jensen and Toll, 1982; Wu and Yeh, 1997). Predominantly, the emphasis in these studies is on total urban growth, undifferentiated from the type of single-family residential sprawl typified by American suburbs.

Our research is distinguished from similar studies in that first, it explores methods for estimating the location and extent of suburban areas, as distinct from the more easily determined commercial/industrial centers or other types of urban land cover. Second, our study is not multi-temporal. The methodology was not intended for forecasting future changes in land cover. Instead, we create a single image of the current location and extent of suburban sprawl in our study area. Finally, this paper compares a traditional remote sensing technique to a new GIS-based method which, to date, has not been reported in the literature. We examine the benefits and costs, in terms of processing time and map accuracy, of using traditional unsupervised classification procedures versus the use of ancillary data, in particular road networks, to delineate suburban sprawl.

Study Area

The study area covers 15,182 hectares (37,486 acres, or approximately one USGS quadrangle) in Columbia County, Georgia (see Figure 1). The area is part of the current sprawl associated with the city of Augusta, the second largest metropolitan area in the state. Augusta hosts approximately 204,000 residents in its metropolitan area, and, due to a natural barrier provided by the Savannah River, heavy urban development has occurred in three directions: to the north, west, and south of the city. The study area consists of primarily residential neighborhoods, interspersed with commercial and forested land.

Methods

Two methods of mapping low-density residential areas in Augusta are compared in this paper. Here, we define low-density residential (LDR) areas as single-family dwellings, irrespective of their lot size. The first method is a traditional remote sensing approach utilizing 1998 Landsat 5 Thematic Mapper data, while the second is GIS based and uses a road network coverage. We will refer to these as the remote sensing (RS) and GIS techniques, respectively.

For the RS technique, a Landsat TM 30-meter resolution image was subset using the 1993 National Land Cover Dataset (NLCD; Loveland and Shaw, 1996; Vogelmann *et al.*, 1998). The 1998 TM image was split into four data sets: (1) pixels that the NLCD classified as agriculture and mining areas, (2) urban areas, (3) water bodies and wetlands, and (4) the remaining areas consisting mostly of forests, pine plantations, and clear-cuts and fields succeeding to forest. Separating agricultural areas from urban areas is particularly important at this stage

J. Epstein and E. Kramer are with the Institute of Ecology, University of Georgia, Athens, GA 30602 (jepstein@arches.uga.edu)(lkramer@arches.uga.edu).

K. Payne is with the Department of Marine Sciences, University of Georgia, Athens, GA 30602 (kapayne@arches.uga.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 63, No. 9, September 2002, pp. 913–918.

0099-1112/02/6809–913$3.00/0

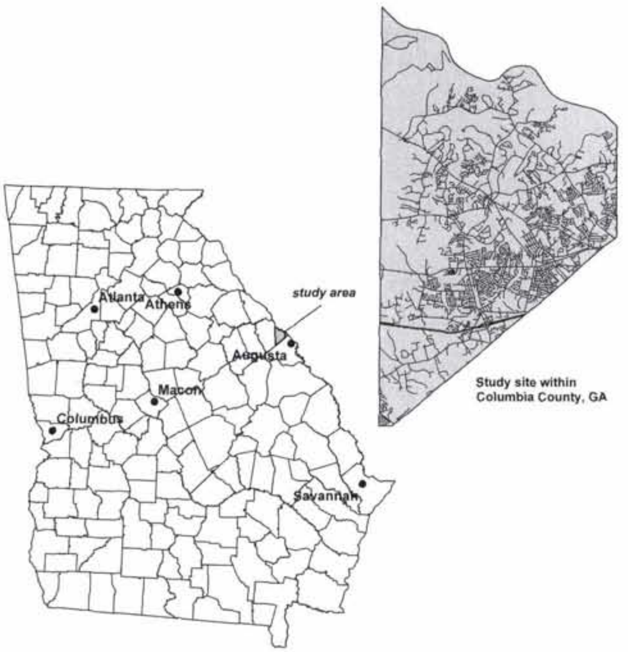

Figure 1. Study area shown in gray along with the road network in Columbia County, Georgia.

because past studies have demonstrated its effectiveness in reducing commission errors in classified imagery (Griffiths, 1988; Masek *et al.*, 2000).

In each of the four subsets, roads, railroad tracks, and utility corridors were removed. This was performed using statewide coverages of each land-cover type mapped at a scale of 1:12,000 (Georgia Department of Transportation, 1997). Each of these coverages was converted to a 30-meter grid and removed from the TM data.

The TM subsets were then subjected to an unsupervised classification, resulting in 25 to 50 clusters. The three subsets previously classified as "non-urban" in the NLCD data were interpreted using 1993 black-and-white digital orthophoto quarter-quadrangles (DOQQs) with a 1-meter resolution, and urbanized areas were identified, separated, and added to the general urban mask. The fourth subset classified as urban in the NLCD, in conjunction with the urban pixels from interpretation of the three other unsupervised classifications, was then masked out from the TM data and again subjected to an unsupervised classification and interpretation using the orthophotos. The four pieces were then added back together, along with gridded road, rail, and utility coverages.

In the second method, we edited vector road data to identify likely LDR areas. We first removed larger roads from the study area that were obviously not suburban in character. Many of these, such as interstates and state highways, could potentially be removed using the feature attribute from the coverage. Unfortunately, the majority of the roads in the coverage did not have an attribute that was useful for this technique. To distinguish these road types, we overlaid the road vector on the TM image or DOQQs and selected and removed them. New suburban roads not in the road coverage but visible in the TM data were digitized and added to the coverage. Our experience with aerial photo interpretation indicates that suburban streets in Georgia often have a particular character to them. They tend to be short and curvy, rather than straight, clustered in developments, and often contain a series of cul-de-sacs. This characterization of suburban streets may be a generalization that is limited to states surveyed by Metes and Bounds, and not the Public Land Survey System (PLSS). The resulting suburban road coverage was then buffered by 45 meters on each side, which, when gridded using a 30-meter cell-size resolution, became a swath three pixels wide. Later, the center pixel was overwritten by the one-pixel-wide road coverage, creating a grid of roads embedded in LDR. The buffered road coverage was then subtracted from the NLCD urban area mask, and the remaining urban areas were run through an unsupervised classification. The buffered areas were then rejoined with the interpreted urban areas, and roads were overlaid on top. The methodologies for both the RS method and the GIS method are outlined in Figure 2. The images resulting from both methods, presented in Plate 1, were compared and assessed according to the amount of time required to produce the image and the thematic and spatial accuracy of the image.

Results

The assessment method consisted of both a time trial and an accuracy assessment of the images resulting from both the RS and the GIS techniques. The tests were run on a Dell Precision 420 Workstation, with 512 Megabytes of RDRAM and an 800 MHz Pentium 3 Processor. The entire test area was mapped using the procedure described in Payne *et al.* (in prep.), but the urban section was mapped twice - once using each method. The urban portion of the RS method required 1 hour and 14 minutes of processing time, which included both creating the clusters and interpreting them. The GIS approach required significantly more time, 3 hours 42 minutes to complete.

To assess thematic accuracy, we created maps composed of LDR and non-LDR areas, and roads overlaid for navigating, shown in Plates 2a and 2b. Using a stratified random sampling scheme, we field checked 62 points in the study area. The location of the sampling points is presented in Figure 3. The error matrix for both methods is provided in Tables 1 and 2.

In the field, 26 of the 62 random points were actually LDR. For the RS method, only 11 of the 26 LDR points were mapped as LDR, returning a user's accuracy of 42.3 percent for that class and 72.6 percent for the binary map overall. Using the GIS method, 20 of the 26 points were accurately mapped as LDR, returning a user's accuracy of 76.9 percent, and 88.8 percent for the binary map overall.

Discussion

Our experience has shown that deriving accurate information on urban extents can be difficult. In rural areas where there is little urban and suburban land, it is possible to achieve adequate estimates of the extent and location of urban classes using standard unsupervised classification techniques. Here, urban areas are small enough that confounding problems are either inconsequential or easily remedied through manual alteration while examining higher resolution photos of the same area. However, in more urbanized areas, obtaining an accurate image of the location and extent of urban sprawl can be problematic. This is largely due to the problem of mixed signature pixels, where single 30-meter pixels are composed of different land-cover types. Because suburban areas can include structures, lawns, trees, and concrete or asphalt, there is often confusion between low-density residential areas and other classes that contain those components. In unsupervised classifications of densely suburban metropolitan regions, small problems become magnified: trees on lawns are confused with several forest classes; grassy areas are common in pasture, recreation, and institutional classes; and pavement is common to high-density residential and commercial/industrial areas.

In this project, the temporal differences between the DOQQ's and the TM image contribute to classification errors. The rate of suburban growth in many metropolitan areas in

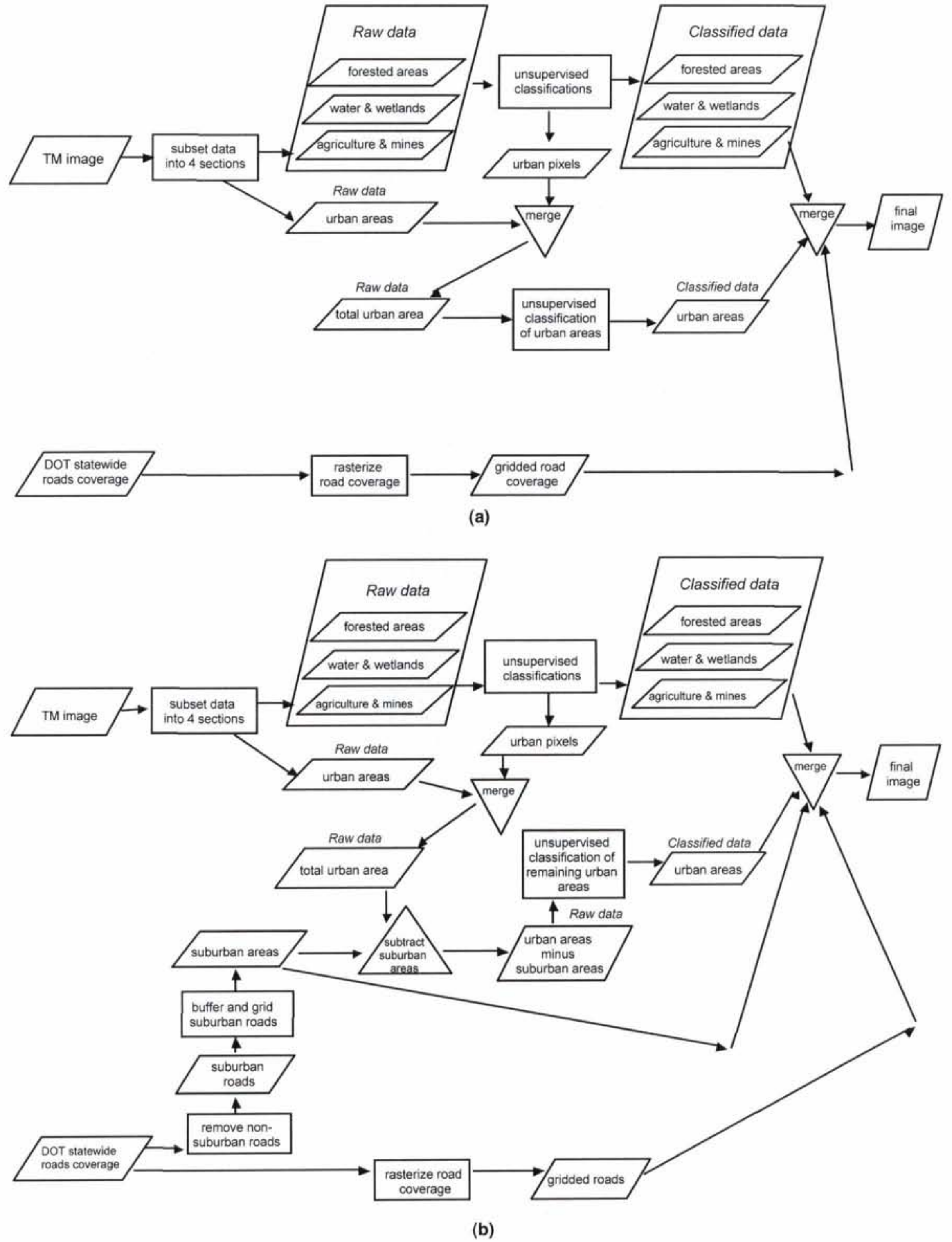

Figure 2. Methodology of the (a) remote sensing (unsupervised classification) approach and (b) the GIS (buffering) approach.

Georgia is rapid enough that what is depicted in the two sets of imagery and what is on the ground may be different. What was an abandoned field in 1993 could be cleared and undergoing construction in 1998, and appear to be an established subdivision in 2001. In addition, confusion between these three signatures in an unsupervised classification is often problematic, especially when using leaf-off imagery. This often results in an overestimation of clearcut and agriculture pixels in urban areas, or a scattering of urban pixels in agricultural areas.

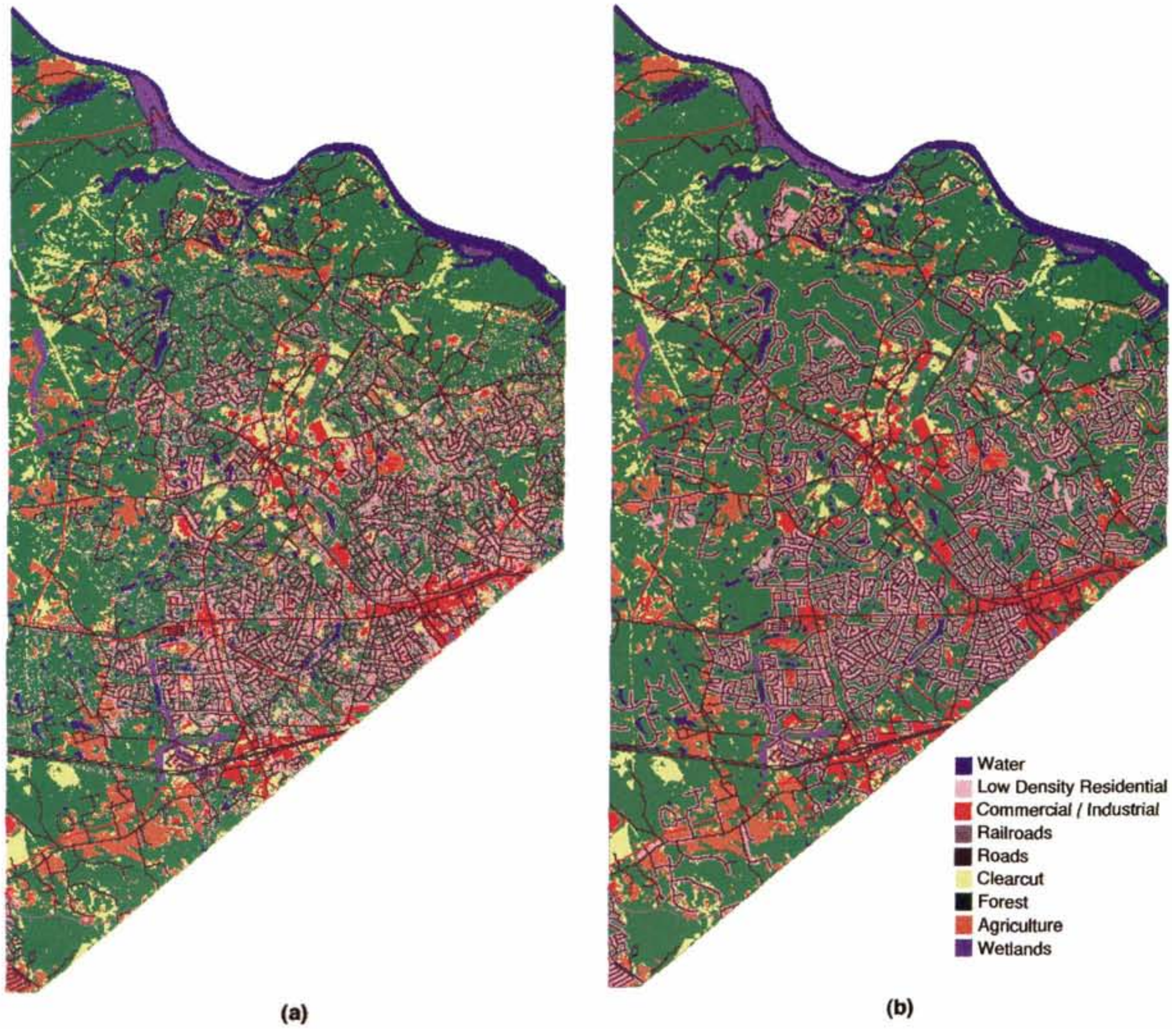

Plate 1. Results of mapping using (a) the RS technique and (b) the GIS technique. Pink represents low-density residential areas.

Overall, both the extent and location of residential areas were more accurate in the map produced using the GIS method. The increased accuracy occurred both within residential and non-residential areas. Housing in suburban areas was more consistently represented, and forested areas between older subdivisions were left intact. This could be due to the fact that older subdivisions tend to have larger lots that were not included in the suburban buffer, leaving forested areas available for classification in the TM image. In addition, the RS method did not represent the subdivisions well, because many roads through subdivisions were only partially enclosed by LDR or were canopied. Also, while the format of an unsupervised classification does not provide an operator the opportunity to use location as a factor in decision-making, context is inherent in the GIS method. Roads in new subdivisions that were added in the GIS method did not show up well in the RS method. Finally, in forested areas, the RS method alone produced a scattering of inaccurate LDR pixels, as shown in Plate 2. These stray pixels are absent from the results of the GIS method, boosting the accuracy for classes confused with LDR areas.

To assess the problem of pixel scatter in the RS method, we could use a low-pass filter to "eliminate" single scattered pixels of LDR into their surrounding classes (see, for example, Toll (1985)). However, Plate 2c demonstrates that this would still leave much of the LDR areas unmapped, because they were confused with other classes (in particular forests, clear-cuts, and commercial/industrial areas) in the unsupervised classification.

Not surprisingly, the cost of this improved accuracy was processing time. While it is undeniable that the time required by the GIS method is greater, the time necessary to edit the road coverages decreases as processors become more adept at manipulating the road database. Also, inexperienced operators get better results with less training using the GIS method because it requires minimal interpretation skills.

Still, there are several difficulties with the GIS method. First, it requires an accurate coverage of road networks, and maintaining an updated road coverage in a state whose residential areas are expanding daily can be a difficult task. This is illustrated by the fact that many of the streets in newer developments were not appended to the roads database. Every attempt was made to digitize these roads, although occasionally some were missed.

Second, a judgment call must be made for each feature to determine if it qualifies as an LDR road. Questions arise such as: At what point does a road become LDR in the transition from

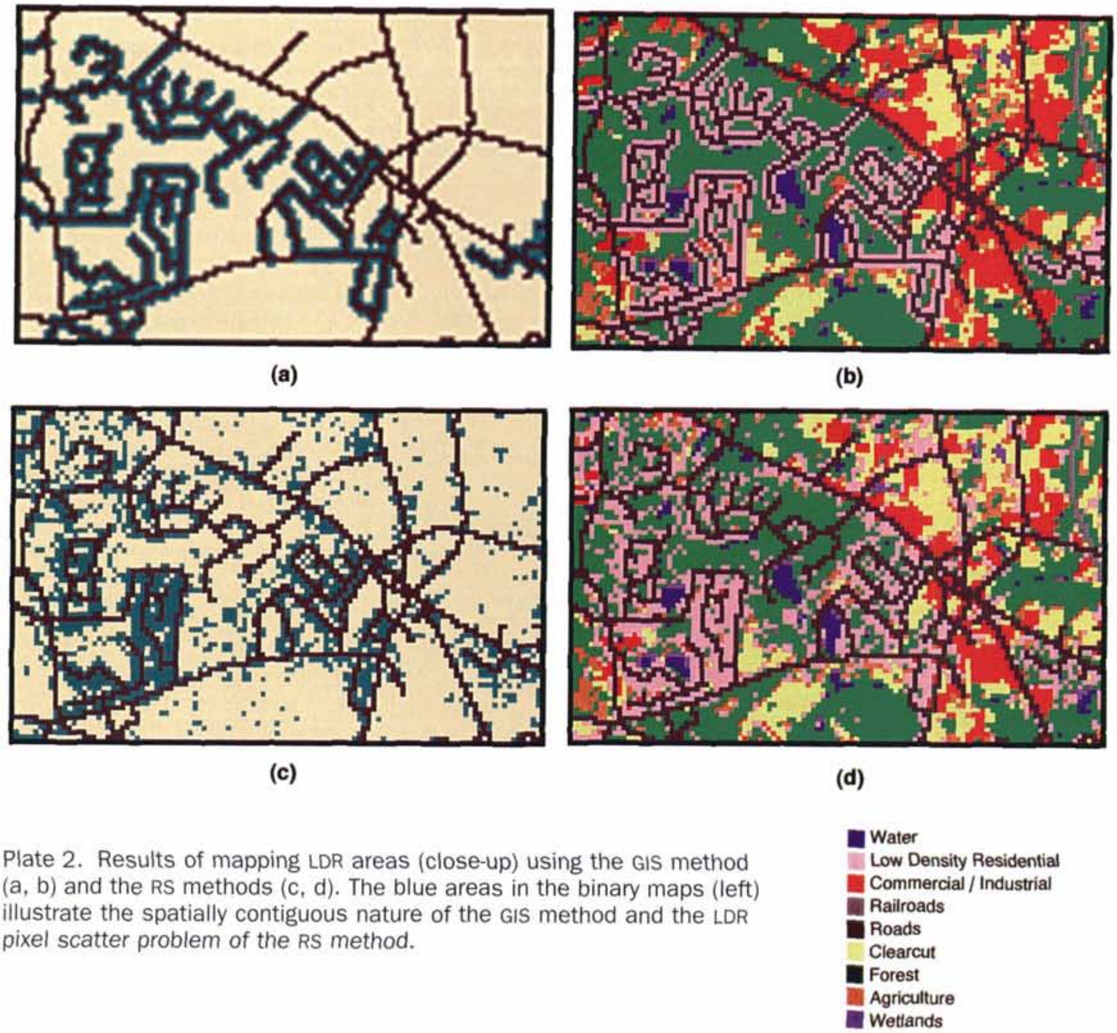

Plate 2. Results of mapping LDR areas (close-up) using the GIS method (a, b) and the RS methods (c, d). The blue areas in the binary maps (left) illustrate the spatially contiguous nature of the GIS method and the LDR pixel scatter problem of the RS method.

Table 1. Error Matrix for Remote Sensing Mapping Method

	RS	MAP Non-LDR	LDR	Total	User's Accuracy
Field	Non-LDR	**34**	2	36	94.44
	LDR	15	**11**	26	42.31
	Total	49	13	**62**	
	Producer's Accuracy	69.39	84.62		**72.58**

Table 2. Error Matrix for GIS Mapping Method

	GIS	MAP Non-LDR	LDR	Total	User's Accuracy
Field	Non-LDR	**35**	1	36	97.22
	LDR	6	**20**	26	76.92
	Total	41	21	**62**	
	Producer's Accuracy	85.37	95.28		**88.71**

rural to more urban? Should streets lined on only one side with houses be included in the coverage of LDR roads if both sides will be classed as LDR in the resulting database? Also, the GIS method assumes that all LDR areas have a width of only one pixel. While this is often true in more densely populated areas, in rural areas, and on the outskirts of cities, single-family homes often have large lots. Despite these drawbacks, we believe the level of improved accuracy achieved using the GIS method make the increased time expenditure worthwhile for a database of this nature.

Conclusion

Due to the nature, extent, and impact of suburban growth in Georgia, we are of the opinion that the time and effort required for the GIS method is well spent. We found that the GIS method results in a more accurate map and that individual operators have more control over the accuracy than with the RS method. Moreover, the resulting suburban map is more consistent and spatially contiguous. For these reasons, we decided to use this method in a state-wide land-cover mapping project for Georgia (Payne *et al.*, in prep.). A summary of the advantages and disadvantages of both methods is presented in Table 3. Areas for further research in the field of mapping suburban sprawl could include comparing the time and accuracy of the GIS method with other methods available such as fuzzy classification, and the use of neural networks and knowledge-based systems.

Acknowledgments

Funding for the GAP and GLUT projects was provided by the USGS Gap Analysis Program, the Turner Foundation, and the

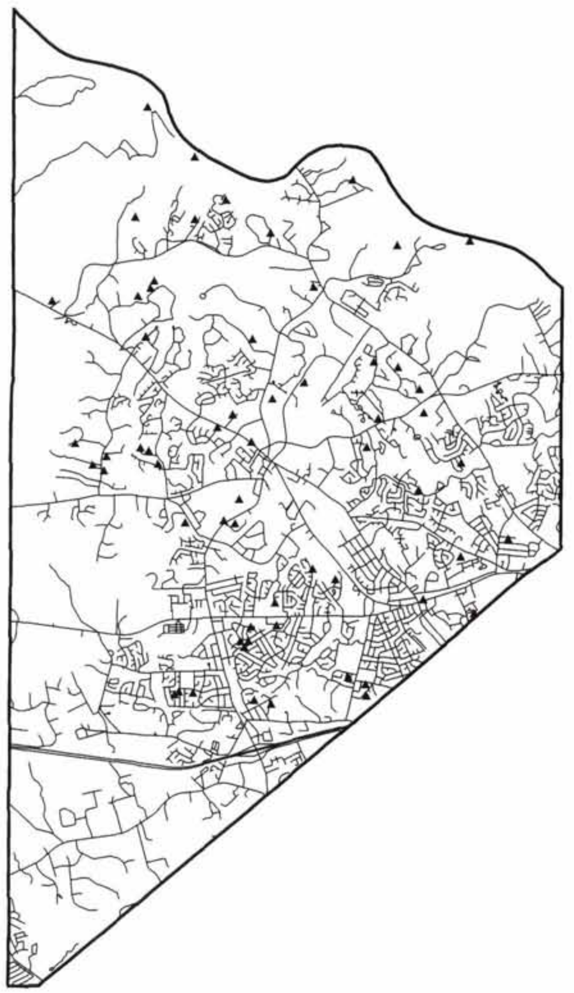

Figure 3. Location of sampling points (shown as black triangles) in the study area.

TABLE 3. SUMMARY OF ADVANTAGES AND DISADVANTAGES OF THE RS AND GIS METHODS OF DELINEATING SUBURBAN SPRAWL

Method	Advantages	Disadvantages
RS	Faster.	Accuracy suffers in LDR and other classes.
	One step process.	Results rely on experience level of operator.
	No ancillary data required.	
GIS	Improved accuracy in LDR - both location and extent.	More time intensive
	Improved accuracy in other classes.	Requires fairly recent photo-corrected road coverage.
	Inexperienced operators get more consistent results with less training.	Assumes all LDR has a width of one pixel.

Sapelo Foundation. The authors thank Silas Mathis and Matt Elliott for their very helpful manuscript edits and Kevin Samples for his technical assistance. Thanks are also due to the anonymous *PE&RS* reviewers for their constructive comments.

References

Bullard, R.D., and G.S. Johnson, 1999. Atlanta megasprawl, *Forum for Applied Research & Public Policy*, 14(3):17–24.

Georgia Department of Transportation, 1997. *Georgia DLG-F Roads and Highways, 1:12,000*, URL: http://www.gis.state.ga.us/.

Gordon, S.I., 1980. Utilizing LANDSAT imagery to monitor land-use change: A case study in Ohio, *Remote Sensing of Environment*, 9:189–196.

Griffiths, G.H., 1988. Monitoring urban change from Landsat TM and SPOT satellite imagery by image differencing, *Proceedings of IGARSS 1988*, 13–16 September, Edinburgh, Scotland, pp. 493–497.

Gross, J., 2000. Urban sprawl threatens the solitude and fragile lands of Georgia's State Parks, *New York Times*, 09 May, 149(51502):16.

Haack, B., N. Bryant, and S. Adams, 1987. An assessment of Landsat MSS and TM data for urban and near-urban land-cover digital classification, *Remote Sensing of Environment*, 21:201–213.

Howarth, P.J., and E. Boasson, 1983. Landsat digital enhancements for change detection in urban environments, *Remote Sensing of Environment*, 13:149–160.

Jacobson, L., 2001. Lawsuit accuses small business administration of promoting sprawl, *Planning*, 67(1):28.

Jensen, J.R., and D.L. Toll, 1982. Detecting residential land-use development at the urban fringe, *Photogrammetric Engineering & Remote Sensing*, 48(4):629–643.

Johnson, M.P., 2001. Environmental impacts of urban sprawl: A survey of the literature and proposed research agenda, *Environment & Planning A*, 33(4):717–736.

Knickerbocker, B., 2000. Forget crime--but please fix the traffic, *Christian Science Monitor*, 92(59):3.

Loveland, T.R., and D.M. Shaw, 1996. Multiresolution land characterization: Building collaborative partnerships, Gap Analysis: A Landscape Approach to Biodiversity Planning, *Proceedings of ASPRS/GAP*, 27 February – 02 March, Charlotte, North Carolina, pp. 83–89.

Masek, J.G., F.E. Lindsay, and S.N. Goward, 2000. Dynamics of urban growth in the Washington DC metropolitan area, 1973–1996, from Landsat observations, *International Journal of Remote Sensing*, 21(18):3473–3486.

Payne, K., K. Samples, J. Epstein, A. Ostrander, J.W. Lee, J.P. Schmidt, S. Mathes, M. Elliott, J. Nackone, S. Sand, F. Hay, M. Merrill, M. Gobaldi, M. Higgins, J. Howell, and L. Kramer, in prep. Multi-source data integration in a statewide land-cover mapping program: An example from Georgia.

Toll, D.L., 1985. Landsat-4 Thematic Mapper scene characteristics of a suburban and rural area, *Photogrammetric Engineering & Remote Sensing*, 51(9):1471–1482.

Vogelmann, J.E., T. Sohl, and S.M. Howard, 1998. Regional characterization of land cover using multiple sources of data, *Photogrammetric Engineering & Remote Sensing*, 64(1):45–57.

Wu, F., and A.G. Yeh, 1997. Changing spatial distribution and determinants of land development in Chinese cities in the transition from a centrally planned economy to a socialist market economy: A case study of Guangzhou, *Urban Studies*, 34(11):1851–1880.

Yeh, A.G., and X. Li, 2001. Measurement and monitoring of urban sprawl in a rapidly growing region using entropy, *Photogrammetric Engineering & Remote Sensing*, 67(1):83–90.

(Received 21 August 2001; accepted 27 September 2001; revised 04 March 2002)