Creating Urban Data Using AVHRR Thermal-Infrared Data

By Bruce E. Wright

Open-File Report 94-401

U.S. Department of the Interior
U.S. Geological Survey
National Mapping Division

Creating Urban Data Using AVHRR Thermal-Infrared Data

Abstract

Using 1-km advanced very high resolution radiometer (AVHRR) satellite data, relative temperature differences caused by thermal conductivity and inertia allow the separation of urban and nonurban land covers. The goal of this research is to investigate a method of creating a global urban data layer from recent satellite data. AHVRR data that were composited on a biweekly basis and distributed by the U.S. Geological Survey EROS Data Center in Sioux Falls, South Dakota, were used for the classification process. These composited images are based on the maximum normalized difference vegetation index (NDVI) of each pixel during a 2-week period. The resultant images are nearly cloud free and reduce the need for extensive preclassification.

In areas of the Western United States, where precipitation is low, the temperature difference between urban areas and rural areas is much less than in the more humid, well-vegetated Eastern United States. Because of this, the initial study was limited to the eastern half of the United States. In the East, the time of maximum temperature difference between the urban surfaces and the vegetated nonurban areas is the peak greenness period. Composite images of the Eastern United States for three 2-week periods were used for determining the urban areas. Two channels of thermal data (channels 3 and 4) and a composited NDVI image were classified using conventional image processing techniques. Although the overall accuracy is low compared with other large-scale urban area delineations, the results indicate the intensity of urban activities within large urban areas.

Introduction

Global mapping of the Earth's surface on a regular basis is becoming feasible using satellite sensors with moderate surface resolution and daily coverage. The advanced very high resolution radiometer (AVHRR) onboard the NOAA-11 satellite collects data in several spectral bands that have proved valuable for collecting global vegetation data with a resolution of 1-km (table 1).

Table 1. Spectral bandwidths of AVHRR data
[From Kidwell, 1991]

Channel	Wavelength (microns)
1	0.58-0.68
2	0.725-1.10
3	3.55-3.93
4	10.3-11.3
5	11.5-12.5

Data for the conterminous United States are being composited on a regular basis and distributed by the U.S. Geological Survey (USGS) EROS Data Center (EDC) (Eidenshink, 1992). AVHRR data are collected daily at the EDC from orbits that pass over the United States at approximately 1330 local standard time. During the growing season, biweekly scenes are produced from these data. The pixels with the minimum cloud cover for the 2-week period are selected and merged into a composite scene for each of the bands. Thus, one image could contain pixels from several orbits. This process provides data for detailed land cover mapping of the United States (Loveland and others, 1991). The vegetation is mapped using the normalized difference vegetation index (NDVI). The NDVI tracks the progression of plant growth during the growing season. Histograms of the NDVI over time for each pixel characterize the surface based on seasonally distinct land cover regions.

The NDVI is produced through combinations of the visible and the near-infrared portions of the electromagnetic spectrum (channels 1 and 2 respectively). Vegetation typically absorbs incoming radiation in the visible spectrum (channel 1) and reflects in the near-infrared spectrum (channel 2). The NDVI is useful in a variety of vegetation studies. The remaining three channels of AVHRR data (channels 3, 4, and 5) are not used for this vegetation mapping. These channels are the middle and far-infrared portions of the electromagnetic spectrum and allow an analyst to distinguish between areas on the ground that possess different thermal characteristics.

Background

Thermal differences between urban and rural areas have been noted by many researchers (Geiger, 1965; Oke, 1978; Landsberg, 1981). One of the earliest studies is a reference published by Howard, who compared temperature readings between downtown London and the surrounding countryside in 1820 (Landsberg, 1981). Known as the urban heat island effect, it is characterized by differences in near-surface air temperature between urban and surrounding rural areas. This effect is attributed to differences in the thermal properties and emissivity of land covers that exist in these two environments (asphalt, concrete, and roofing materials vs. forest and other vegetated rural areas). The low latent heat that is characteristic of urban areas exacerbates the differences. The well vegetated eastern suburban and rural areas contain higher levels of moisture than cities and, thus, can offset the potentially higher surface temperature through evaporative cooling. Impermeable urban surfaces such as roads, buildings, and parking lots do not retain water, which is lost as overland flow. Not only does the vegetation in rural areas contain a reserve of moisture, but the soil also contains a higher amount of moisture. Direct heat is also generated in cities because of heating, air conditioning, and transportation and to satisfy other needs of the people residing there.

Although much research has been conducted using thermal satellite data for studying the water and surface energy budgets of the Earth (see Pinker, 1990; Sellers and others, 1990), fewer studies have used these data for explicitly studying the urban environment. One of the earliest satellite-based studies of urban and rural temperature differences used a night scene of the TIROS (ITOS-1), which had a resolution of 7.4 km at nadir and a thermal channel that measured radiation in the 10.2-12.5 um region (Rao, 1972). Previous to that study, the thermal data produced by the TIROS satellite were used for cloud mapping and the study of ocean currents. In this study, large warm areas along the east coast related directly to the large urban corridor from New York to Washington, D.C.

Carlson and others (1977) investigated diurnal temperature variations in urban land in Los Angeles, California, using 1-km NOAA-3 very high resolution radiometer (VHRR) data. Also, an extensive investigation of urban and rural temperature differences in 11 metropolitan areas was conducted using 1-km NOAA-5 VHRR data in the eastern United States (Matson and others, 1978). This study demonstrated that major urban areas can be delineated using thermal data and that the intensity of the heat island effect was directly related to the density of urbanization. Comparisons to the U.S. Bureau of Census urbanized maps were noted. The differences between the two sources indicated these data could be used for urbanized area change detection studies. The Heat Capacity Mapping Mission satellite, with a spatial resolution of 500 m, was also used for studying the effect in the New York-New England area (Price, 1979). A difference in satellite-observed brightness temperatures of 10-15 °C between many larger cities and the surrounding rural areas was calculated.

Several studies have used AVHRR data for studying the urban heat island effect (Roth and others, 1989; Casalles and others, 1991; Gallo and others, 1993). The study by Gallo and others demonstrated a linear relationship between the differences in urban and rural NDVI and minimum temperatures.

To conduct an inventory of urban areas useful in global scale environmental models requires that the data source be capable of defining major urban areas and yet be a manageable size. As stated above, 1-km AVHRR data have been compiled for the conterminous United States for 14-day intervals since 1990. These data are available on CD-ROM's (U.S. Geological Survey, 1991). These data will soon be available globally. Data at this resolution is useful for depicting regional vegetation and temperature patterns. The conterminous United States is represented by 4,587 by 2,889 pixels. Using data with such a coarse spatial resolution limits the type of information that can be extracted for urban land cover studies. With a spatial resolution of 1-km, many small cities that are less than 2 km^2 may not be detected. Many suburban areas that contain a great deal of vegetation also may not appear as developed areas. The primary consideration of most land and atmosphere models is the predominant land cover, so this does not present a great problem. The resolution of the regional models is generally more than 1-km, so that finer resolution data would not be necessary. In those areas of extensive vegetation interspersed with housing, the vegetation cover is of primary interest. The urban core area, which has a noticeable heat island effect and is primarily covered with materials such as concrete and asphalt, is extensive enough to be detected. This is the area that most heavily influences the local climate and regional land and atmosphere interactions.

The 1-km AVHRR land cover characteristics data, compiled by the EDC (Loveland and others, 1991), do not contain an urban layer. A method of extracting extensive urban areas for inclusion into this data would be desirable. For the conterminous United States there are digital urban data layers available such as the Digital Chart of the World (U.S. Defense Mapping Agency, 1992), U.S. Bureau of Census urban areas (U.S. Bureau of Census, 1992) and USGS land use and land cover data (U.S. Geological Survey, 1986). For global applications, the Digital Chart of the World contains the most comprehensive urban layer. However, these data were compiled from nautical charts that vary in age, may not represent the current state of global urbanization, and do not indicate the intensity of urbanization within the broad urban boundaries. The goal of this research is to investigate a method of creating a global urban data layer from recent satellite data.

Methods

Data from the biweekly composite CD-ROM's distributed by the EDC for 1991 were used for this study. The same method of pixel selection, based on the maximum NDVI (derived from channels 1 and 2), was applied to the remaining thermal channels (channels 3, 4, and 5). Temperature conditions on the ground can vary greatly over a 2-week period. The maximum temperature may not relate to a maximum NDVI, and different results are possible if the compositing is based on a maximum temperature index. Because relative temperature differences between urban and rural areas are being investigated, thermal pixels should be extracted from the same period (Price, 1986). However, single date cloud-free images were not available. Because of the short interval used in the compositing process (14 days) and the cloud-free nature of the resultant composites, composited data were used for this study.

Assuming that the time of maximum vegetation cover and maximum summer heat result in maximum differences between urban and rural areas, those scenes available during the peak of the growing season were checked for visual quality using visualization software. Some of the composited scenes contained noticeable linear discontinuities that appeared to be a result of adjacent scene merging. These scenes were removed from consideration. A number of urban regions were extracted from the data for all 1991 periods in 256 by 256 windows, and these were viewed using animation software to quickly evaluate scene quality. Three periods were selected for further analysis, during the early, middle, and late parts of the growing season (March 26 - May 9, June 21 - July 4, and August 30 - September 12). Because the peak period varies seasonally from south to north, it is not possible to obtain one biweekly scene that represents the peak season for all regions.

The ability to discriminate urban from rural areas using maximum vegetation information is most appropriately applied to the well-vegetated, high-precipitation Eastern United States and the Pacific coast region. Vegetation patterns in the dry Central Plains and arid West are very different, and the urban vegetation patterns are often reversed. Irrigated residential areas can contain more vegetation than the surrounding rural areas, especially in such areas as Salt Lake City, Denver, and other Western cities. These vegetation differences are primarily caused by precipitation differences. Areas in the West that receive less than 20 inches of rain compose the less vegetated rangelands and desert areas. This region extends from north to south at approximately 100° W. longitude to the Pacific coast. The eastern boundary is also the approximate dividing line between the more populated and urbanized East and the more sparsely populated and rural Great Plains States. The range of average annual precipitation west of this line, which extends to the more humid coastal States, descends to less than 5 inches in Nevada, Utah, and Arizona and has had a major effect on the land cover and use within that region (Marschner, 1959).

Because of these factors, the conterminous U.S. image was subdivided at approximately 97° 30', and only the east portion was used. The final image contains 2,889 rows and 2,288 columns. Three bands of AVHRR data containing the maximum information regarding urban and rural differences were selected for unsupervised classification using off-the-shelf commercial image processing software. The raw data values received by the AVHRR satellite are converted to energy units onboard the satellite (Kidwell, 1991). The energy values are converted to brightness temperature in degrees Kelvin. These values are then converted to byte data using a scaling factor in which 202.5 is subtracted from the brightness temperature value and the difference is multiplied by 2 (U.S. Geological Survey, 1991). This allows data values to fall within the range of 0-255, thus to be represented by 8 bits.

Table 2 is a statistical summary of channels 3, 4, and 5 and the NDVI data. Thermal chan-

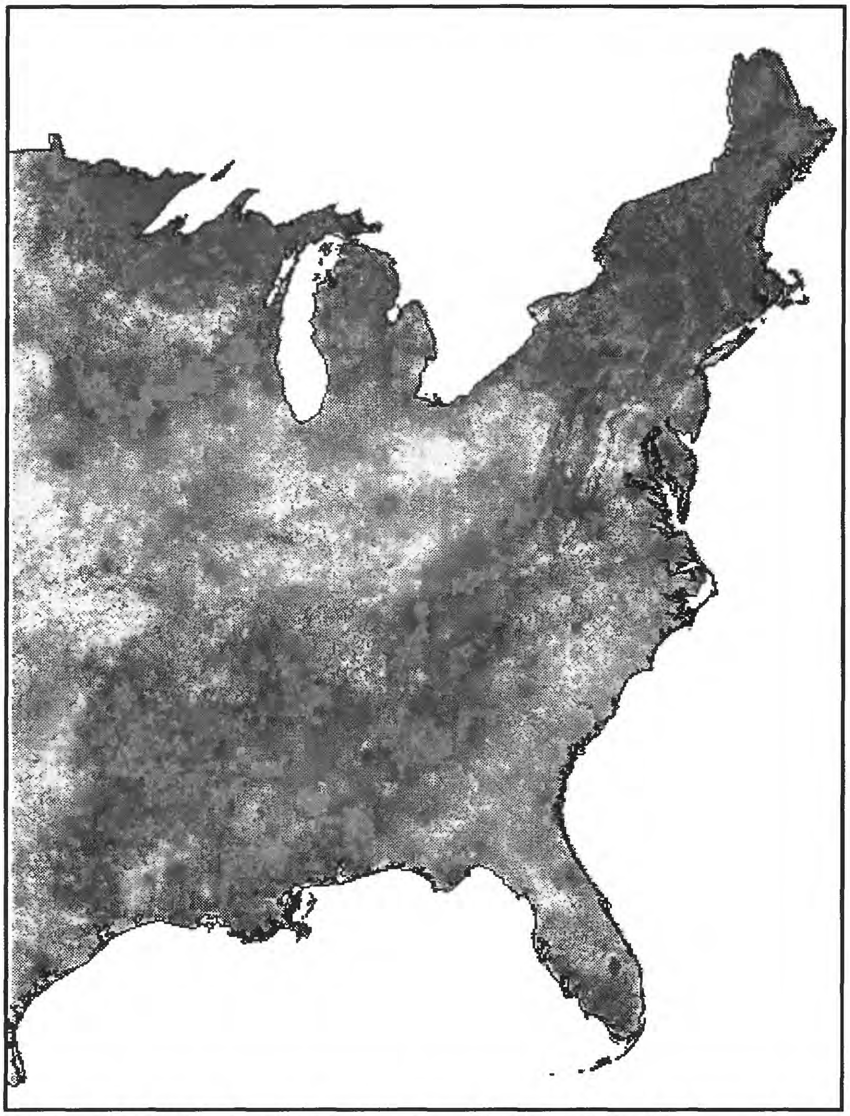

Figure 1. AVHRR composite image, thermal channel 4. Light areas are areas of highest surface temperature.

nels 4 and 5 are very highly correlated (correlation coefficient of 0.99986) . Channel 4 (fig. 1) was

Table 2. Statistical summary of AVHRR bands and 2 year maximum NDVI data

	Channel 3	Channel 4	Channel 5 (Not used)	2 yr. max. NDVI
Min.	109	7	6	102
Max.	238	217	212	177
Mean	199.07	189.97	185.31	160.89
Std. Dev.	9.25	10.77	10.89	7.76
Median	200	191	189	162

selected as the principal thermal channel because the urban characteristics represented in channels 4 and 5 should be nearly identical. Channel 3 consists of reflected and emitted radiation (of approximately equal magnitude (Becker and Li, 1992)) that is obtained from the middle-infrared portion of the electromagnetic spectrum in the 3-5 micron atmospheric window (3.55-3.93 microns). Because of the reflected component, it is not used for most terrestrial thermal applications. The peak of infrared emission of the Earth is approximately 9 microns, making data from the longer wavelength atmospheric window (8-14 microns) more appropriate. The fact that channel 3 does represent emitted and reflected radiation makes it very useful for this study in which both temperature and vegetation characteristics are used to distinguish urban areas.

A negative relationship is expected between the amount of vegetation present and the intensity of urban land use (fig. 2). Given this relationship the NDVI data were selected as the

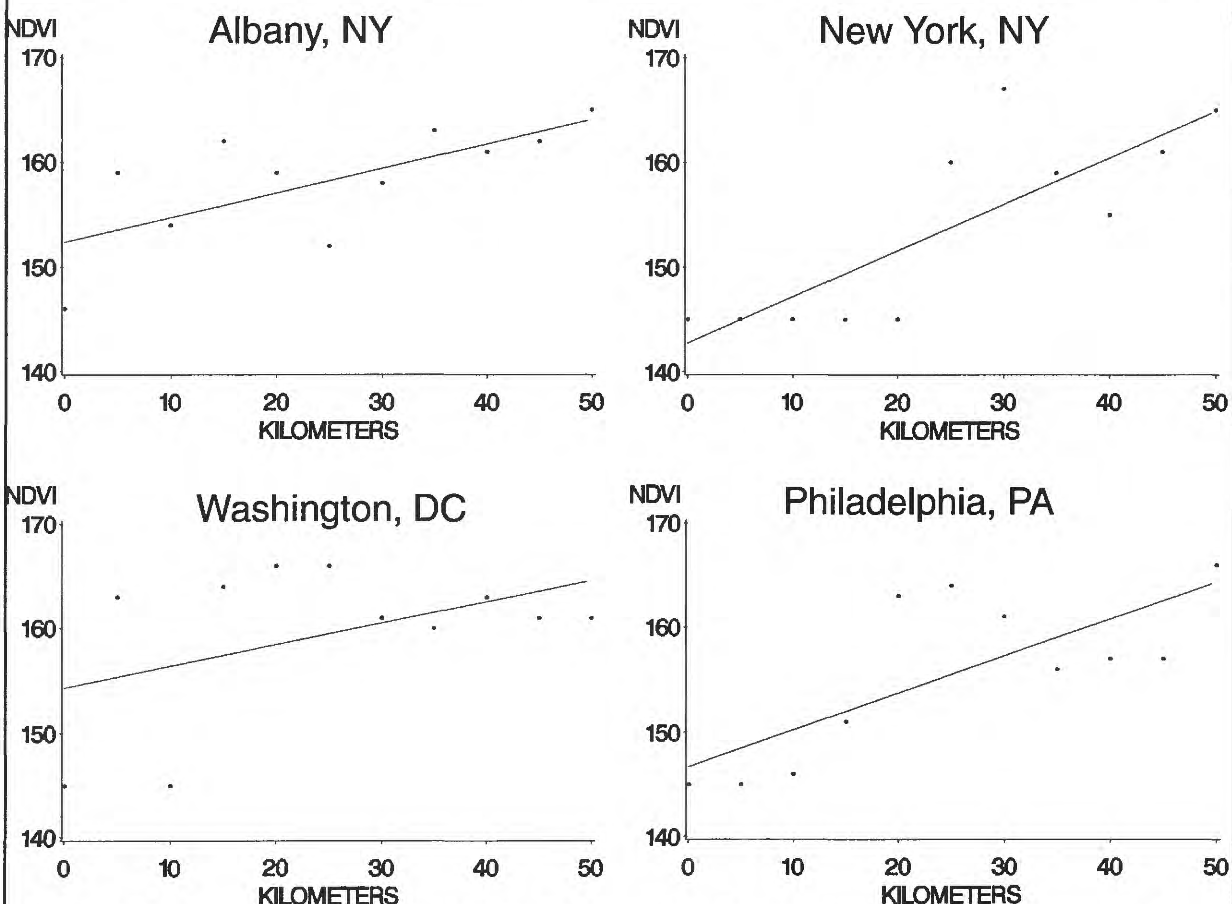

Figure 2. Relationship between the NDVI and the distance from the urban core. Pixel values were sampled at 5-km intervals along a 50-km profile directly west of the urban core.

third band for classification. Biweekly NDVI images for the three periods were available. However, as portions of the country become vegetated at different times, a maximum NDVI image was selected that contains the maximum NDVI for each pixel during a 2-year period, 1990-91

Figure 3. Two-year maximum NDVI image (1990-91). Light areas contain maximum vegetation. Urban areas, with less vegetation than surrounding rural areas, appear as darker areas.

(fig. 3). These three channels were then classified using the isodata clustering routine, which produced 100 unique clusters. The isodata technique is an automated clustering routine that iteratively clusters multiband data, beginning with arbitrary cluster centers that are adjusted so as to produce the maximum separation between the desired number of clusters (Richards, 1986).

Clustering was conducted on the three sets of image bands. The final clusters were identified by determining the values of the known urban areas based on ancillary data such as the Digital Chart of the World, Bureau of Census urban areas, and USGS land use and land cover data.

Results

The early-summer and mid-summer scenes contained a large number of pixels misclassified as urban. In most cases these were rural farmland. These errors are caused by the presence of plowed fields that have a high daytime temperature. Fields containing early season crops with canopies that do not fully cover the soil also have higher temperatures. The highest agreement between urban areas and the clusters was produced using the late-summer bands of channels 3 and 4 and the 2-year NDVI data (fig. 4). High late-summer temperatures lead to high surface tem-

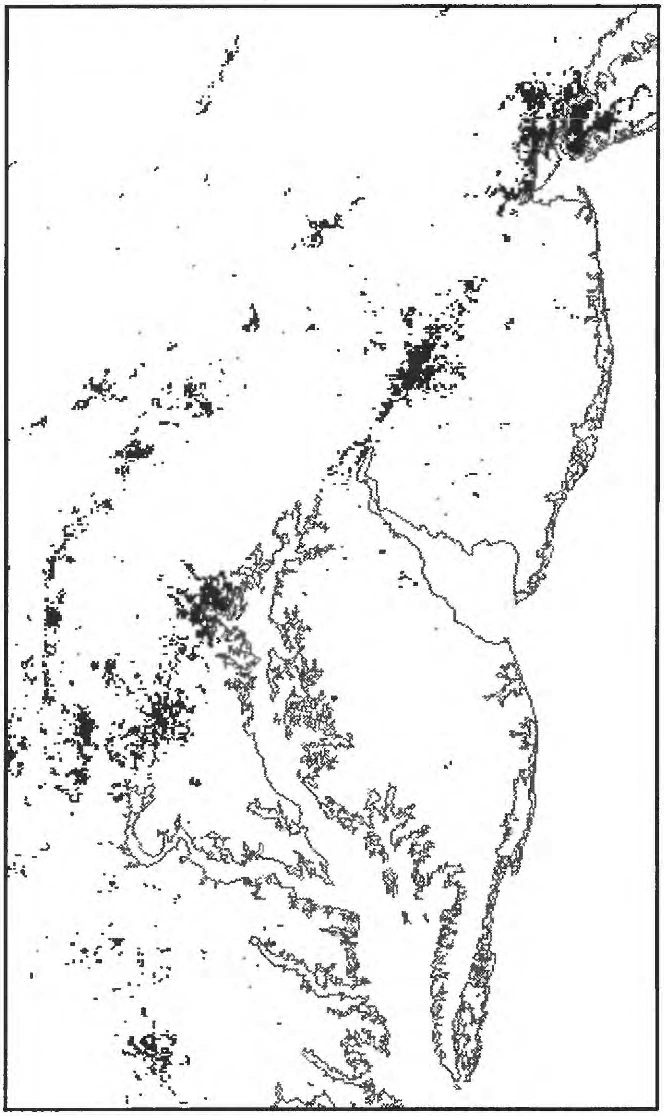

Figure 4. Results of unsupervised classification. Dark pixels are those areas identified as urban. Washington, Baltimore (center), and Philadelphia and New York (top) are the major urban areas.

peratures in urban areas, and a higher level of energy is produced by an increased use of air conditioning in offices and high-density residential areas. Although the vegetation in many of the southern areas is beyond the time of peak greenness, there is still heavy vegetation present, and there are fewer areas of bare fields than in the earlier scenes. To calculate the accuracy of the classes, the Census urban area map was used. These data were combined with the classification results to determine the number of pixels classified as urban that existed within the urban area map (table 3).

Table 3. Classified pixels located within the urban area map

Class	Urban pixels	Mean NDVI	Mean ch. 3	Mean ch. 4	Total pixels in class	Percent correct for class
95	1180	157.07	215.13	204.60	24121	4.89
96	1524	151.30	215.11	190.01	9006	16.92
97	8658	148.82	213.88	201.13	32754	26.43
98	150	163.90	217.33	205.69	11424	1.31
99	12733	134.88	218.42	200.61	16626	76.58
100	3703	150.41	220.29	206.06	15951	23.21
TOTAL	26768				109882	
					% correct =	24.36

Table 3 shows that class 99 was the most accurate (76.58 percent). This class represents the urban core in which there is relatively little vegetation present. This lack of vegetation is indicated by the low mean NDVI (134.88). The mean channel 3 value is comparatively high as well (218.42). In the channel 3 spectral window, the spectral response of vegetation is low, so that a high value in this channel indicates low vegetation content.

The low overall accuracy of 24.36 percent is a result of large agricultural areas in eastern Kansas and southeastern Nebraska, eastern Texas, central Florida, and scattered throughout east-central Illinois that were classed as urban because of high surface temperature. The areas in Nebraska and Kansas that are identified as urban exist in a region in which the peak of greenness is very early in the year (June 8 - June 21) compared with areas just to their east (August 17-August 30). Although the NDVI image compensates for the early peak greenness by containing the maximum vegetation regardless of time period, the lack of vegetation has a strong influence on the thermal values. In these areas, crops, primarily corn, soybeans, and some winter wheat, have been harvested or are senesced by late August. There is much bare soil with low thermal inertia and high daytime surface temperature. It is an agricultural area that receives a low amount of precipitation and is east of a region of irrigated cropland.

Other discrepancies are those cases in which urban areas or areas undergoing development were not identified as urban in the 1990 Census data. In central Florida, several large areas that are outside of the Census urban areas have undergone extensive anthropogenic change (that is, resi-

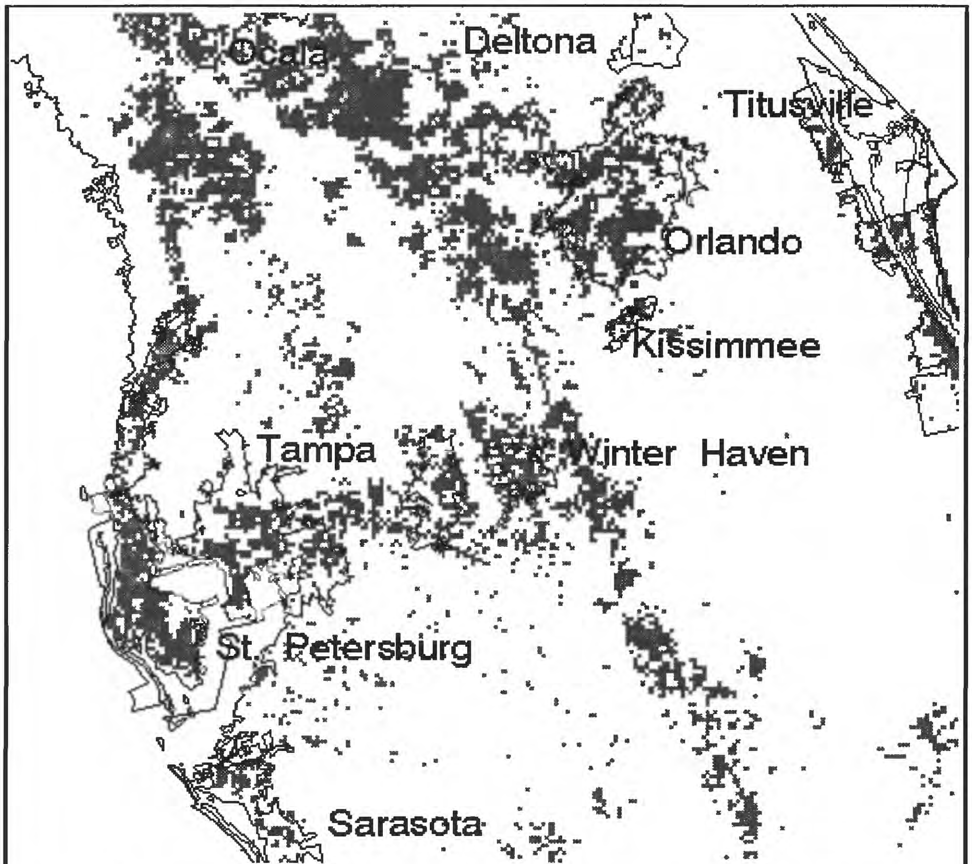

Figure 5. Central Florida urban areas. Census area is outlined in black. Classified pixels are in gray.

dential and commercial development) (fig. 5). Walt Disney World caused much development along U.S. Highway 27 in central Florida, from the Winter Haven area northwest to Ocala. This area was formerly an extensive region of orange groves. Much of this area has undergone extensive anthropogenic change that is not reflected in the other urban data sources.

Classification accuracies for the other two periods based on the Census boundaries are 16 percent for period 11 (June 21 - July 4) and 7 percent for period 7 (March 26 - May 9). If the agricultural areas in the west and the extensive transformed areas in Florida are excluded from the analysis, the accuracy increases. For the August 30-September 12 scene, the accuracy increases from 24 percent to 43 percent. Although still low compared with conventional land use and land cover classification, the comparison includes urban areas of extensive vegetation cover that were not classed as urban in the AVHRR classification but were classed as urban by the Bureau of Census. For biophysical modeling, an indication of the intensity of the urban use may be more important than a determination of the type of use, such as parks, golf courses, and low-density residential areas that are heavily vegetated.

Conclusion

Major urban areas in the northeastern United States were determined using a combination of NDVI and thermal-infrared AVHRR data. The overall classification accuracy for the Eastern States (24 percent) and the Northeastern States (43 percent), although low, is, in part, because of the use of Census data for comparison. Many areas of extensive vegetation cover are classified as "urban" in the Census data. In this study, those areas that were vegetated areas within the Census urban areas were not identified as urban. However, urban core areas with extensive anthropogenic change were identified as urban. Also identified were lower density urban areas that were partially vegetated. When combined with the higher density urban core areas, these data give an indication of the intensity of urban activity. The areas that exert the strongest influence on climate are those more developed, highly urbanized areas. Other errors were associated with extensive agricultural areas in which there is low precipitation.

This technique is useful for distinguishing highly urbanized areas from those less urbanized. The late summer period (August 30 - September 16) showed the highest percentage of urban areas when compared with the U.S. Census urban area map. In the Middle Atlantic and Northeastern States, this period ensures a maximum amount of vegetation cover and high urban surface temperatures.

References

Becker, F. and Li, Z.L., 1992, Temperature-independent thermal infrared spectral indices (TISI) and land surface temperature determined from space, in TERRA-1, in Mather, P.M., ed., Understanding the terrestrial environment - the role of earth observations from space: London, Washington, D.C., Taylor & Francis, p.185-201.

Carlson, T.N., Augustine, J.A., and Boland, F.E., 1977, Potential application of satellite temperature measurements in the analysis of land use over urban areas: Bulletin of the American Meteorological Society, v. 58, no. 12, p. 1301-1303.

Casalles, V., Garcia, M.J.L., Melia, J., and Cueva, A.J.P., 1991, Analysis of the heat-island effect of the city of Valencia, Spain, through air-temperature transects and NOAA satellite data: Theoretical and Applied Climatology, v. 43, no. 4, p. 195-203.

Eidenshink, J.C., 1992, The 1990 conterminous U.S. AVHRR data set: Photogrammetric Engineering and Remote Sensing, v. 58, no. 6, p. 809-813.

Gallo, K.P., McNab, A.L., Karl, T.R., Brown, J.F., Hood, J.J.,and Tarpley, J.D., 1993, The use of NOAA AVHRR data for assessment of the urban heat island effect: Journal of Applied Meteorology, v. 32, p. 899-908.

Geiger, R., 1965, The climate near the ground: Cambridge, Mass., Harvard University Press, 611p.

Kidwell, K.B., ed., 1991, NOAA polar orbiter data users guide (TIROS-N,NOAA-6, NOAA-7, NOAA-8, NOAA-9, NOAA-10, NOAA-11, and NOAA-12): Washington, D.C., National Ocaeanic and Atmospheric Administration.

Landsberg, H.E., 1981, The urban climate: International Geophysics Series, v. 28, New York, Academic Press, 275 p.

Loveland, T.R., Merchant, J.W., Ohlen, D.O., and Brown, J.F., 1991, Development of a land-cover characteristics database for the conterminous U.S.: Photogrammetric Engineering and Remote Sensing, v. 57, no. 11, p. 1453-1463.

Marschner, F.J., 1959, Land use and its patterns in the United States: Agriculture Handbook no. 153, U.S. Department of Agriculture, 277 p. with map insert.

Matson, M, McClain, E.P., McGinnis, D.F., and Pritchard, J.A., 1978, Satellite detection of urban heat islands: Monthly Weather Review, v. 106, p. 1725-1734.

Oke, T.R., 1978, Boundary layer climates: New York, John Wiley & Sons, 372 p.

Pinker, R.T., 1990, Satellites and our understanding of the surface energy balance: Palaeogeography, Palaeoclimatology, Palaeoecology (Global and Planetary Change Section), v. 82, p. 321-

342.
Price, J.C., 1979, Assessment of the urban heat island effect through the use of satellite data: Monthly Weather Review, v. 107, p. 1554-1557.
________ 1986, Remote sensing in the thermal infrared: Remote Sensing Reviews, v. 1, p. 187-196.
Rao, P.K., 1972, Remote sensing of urban "heat islands" from an environmental satellite: Bulletin of the American Meteorological Society, p. 647-648.
Richards, J.A., 1986, Remote sensing digital image analysis - an introduction: New York, Springer-Verlag, 281 p.
Roth, M., Oke, R.R., and Emery W.J., 1989, Satellite-derived urban heat islands from three coastal cities and the utilization of such data in urban climatology: International Journal of Remote Sensing, v. 10, no. 11, p. 1699-1720.
Sellers, P.J., Rasool, S.I., and Bolle, H.J., 1990, A review of satellite data algorithms for studies of the land surface: Bulletin of the American Meteorological Society, v. 71, no. 10, p. 1429-1447.
U.S. Bureau of Census, 1992, Tiger/line files supplemental [machine-readable data files]: Washington, D.C.
U.S. Defense Mapping Agency, 1992, Digital chart of the world [machine-readable data files]: Washington, D.C.
U.S. Geological Survey, 1973, Washington, D.C.,VA,MD, 1:250,000-scale land use map [machine-readable data file]: Washington, D.C.
__________________ 1986, Land use and land cover digital data from 1:250,000- and 1:100,000-scale maps, data users guide 4: National Mapping Program Technical Instruction, 36 p.
__________________ 1991, Conterminous U.S. AVHRR biweekly composite: U.S. Geological Survey CD-ROM set, AVHRR-9104, 5 discs.