

# **Integrating Multisource Land Use and Land Cover Data**

By Bruce Wright, Mike Tait, Katherine Lins, Janet Crawford,  
Susan Benjamin, and Jesslyn Brown

Open-File Report 95-652

# Integrating Multisource Land Use and Land Cover Data

By Bruce Wright<sup>1</sup>, Mike Tait<sup>2</sup>, Katherine Lins<sup>1</sup>, Janet Crawford<sup>1</sup>, Susan Benjamin<sup>3</sup>, Jesslyn Brown<sup>4</sup>

## ABSTRACT

As part of the U.S. Geological Survey's (USGS) land use and land cover (LULC) program, the USGS in cooperation with the Environmental Systems Research Institute (ESRI) is collecting and integrating LULC data for a standard USGS 1:100,000-scale product. The LULC data collection techniques include interpreting spectrally clustered Landsat Thematic Mapper (TM) images; interpreting 1-meter resolution digital panchromatic orthophoto images; and, for comparison, aggregating locally available large-scale digital data of urban areas. The area selected is the Vancouver, WA-OR quadrangle, which has a mix of urban, rural agriculture, and forest land. Anticipated products include an integrated LULC prototype data set in a standard classification scheme referenced to the USGS digital line graph (DLG) data of the area and prototype software to develop digital LULC data sets.

This project will evaluate a draft standard LULC classification system developed by the USGS for use with various source material and collection techniques. Federal, State, and local governments, and private sector groups will have an opportunity to evaluate the resulting prototype software and data sets and to provide recommendations. It is anticipated that this joint research endeavor will increase future collaboration among interested organizations, public and private, for LULC data collection using common standards and tools.<sup>5</sup>

## INTRODUCTION

The USGS is in the process of developing a LULC mapping program to meet Federal and other organizational needs for nationally consistent LULC data. LULC data are an integral component of many natural resource applications, such as ecosystem analysis and management studies, water quality studies, and natural hazard impact and mitigation studies. Starting in 1973, the USGS developed techniques for collecting LULC data on a national scale (Anderson and others, 1976). A national LULC data base of the conterminous United States was produced for approximately 1975-85. These data, commonly referred to by the acronyms GIRAS (Geographic Information

---

<sup>1</sup>U.S. Geological Survey, 521 National Center, Reston, Virginia, 22092.

<sup>2</sup>Environmental Systems Research Institute, 380 New York Street, Redlands, California 92373-8100

<sup>3</sup>U.S. Geological Survey, EROS Data Center, Sioux Falls, South Dakota 57198

<sup>4</sup>Hughes STX Corp, EROS Data Center, Sioux Falls, South Dakota 57198

<sup>5</sup>Any use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government

Retrieval and Analysis System) or LUDA (Land Use Data Analysis program), are useful for historical LULC studies. The data have not been updated since initial compilation. To meet current requirements for LULC data, a follow-on inventory is necessary.

Many Federal agencies are developing land cover data bases to meet their particular needs. These include the Environmental Mapping and Analysis Program of the Environmental Protection Agency, the Coastwatch program of the National Oceanic and Atmospheric Administration, the National Water Quality Assessment Program of the USGS, the Gap Analysis Program of the National Biological Service, the Scientific Assessment and Strategy Team, a multiagency group evaluating the impacts of the 1993 flood on the Upper Mississippi River basin (Scientific Assessment and Strategy Team, 1994), and others. These groups all have needs for national or regional land cover data and some for land use data. The first four of these Federal programs have formed a consortium to purchase nationwide Landsat TM data sets from EOSAT company, and will be producing land cover data sets to meet their resource analysis needs (Shaw and others, 1993). They are working with the Multi-Resolution Land Characterization (MRLC) program at the Earth Resources Observation System (EROS) Data Center to purchase and process this data. A current, nationally consistent land cover data base could satisfy a large percentage of the needs of these organizations, but such a product does not exist. Several States (Georgia, Wisconsin, North Carolina, South Carolina, and Florida) are also developing statewide land cover data bases from satellite images.

Analysis techniques and data sources have changed since the initial compilation of the GIRAS data. Studies are underway at the USGS to make use of new digital techniques and data sources for collecting base cartographic data and LULC information. As part of this research effort, the USGS, in cooperation with the ESRI, is conducting a proof-of-concept pilot project to apply some of these techniques to develop more efficient methods for collecting LULC data. The goal of the proof-of-concept project is to evaluate LULC data collection methods and technologies and to develop cost and resource estimates for a national digital LULC data base program.

Many more digital sources are now available than when the earlier LULC collection program began. Aerial photos, similar to those used during the earlier GIRAS mapping program, are now available for some parts of the country in the form of digital orthophoto quadrangles (DOQ). DOQ's are digital representation of aerial photographs in which the distortions inherent in aerial photographs due to camera tilt and terrain relief are removed (U.S. Geological Survey, 1992b). Collection of data from these sources in a digital environment allows the separate steps of feature delineation and capture in a digital file to be combined, reducing the number of times each feature is manipulated. These source materials differentiate among manmade features, which can serve as indicators of the land use. However, the image files are very large and these sources may not be appropriate for large rural areas that do not require detailed source material to determine the primary vegetative cover. For this project, DOQ data were used to collect detailed vector LULC data only for developed areas where the DOQ level of detail is needed for interpretation of land use information.

Another data source that has become widely available since 1982 is the Landsat TM image, which has a spatial resolution of 28.5 meters. For most parts of the United States where there are extensive contiguous regions of a few selected land cover types, use of the very high resolution single

channel DOQ data is not appropriate for land cover collection, especially at a national scale. Land cover information extracted from TM data has the potential for complementing the LULC information extracted from the panchromatic DOQ data. TM data were used for this purpose in the proof-of-concept project. Other sources of LULC information include other Federal, State and local agency LULC maps, descriptive information, and ancillary data from which land cover information can be derived. Climatic, elevation, ecoregion boundaries, soils, surface geology, and other data types can provide land cover inferences. Land use information can also be inferred from sources such as agricultural and socioeconomic data. To test the feasibility of incorporating existing LULC with the other source data, locally produced data sets were obtained and evaluated.

## LULC CLASSIFICATION SCHEME

The classification scheme used for the earlier GIRAS program is a hierarchical system that allows flexibility for use at multiple levels, depending on the level of detail and scale required by the application (Anderson and others, 1976). Level I, the highest and the most general level of the Anderson scheme, includes major categories of LULC such as urban, agricultural, rangeland, and others (table 1). The more detailed level II was developed for mapping LULC at a national scale. In this scheme, some categories represent the predominant land use, whereas others only represent the land cover. For example, the level II categories under urban or built-up land are considered land uses (urban, commercial and services, and others), as are most agricultural uses, but the remaining level II categories are considered land cover information. This is a problem for users of the data who need information about the land cover in those areas mapped as land use. For example, the type of vegetation and density of manmade features in residential areas vary greatly by region in the United States.

**Table 1.** Anderson land use and land cover classification system

Level I	Description	Level II	Description
1	Urban or built-up land	11	Residential
		12	Commercial and services
		13	Industrial
		14	Transportation, communications and utilities
		15	Industrial and commercial complexes
		16	Mixed urban or built-up land
		17	Other urban or built-up land
2	Agricultural land	21	Cropland and pasture
		22	Orchards, groves, vineyards, nurseries, and ornamental horticultural areas
		23	Confined feeding operations
		24	Other agricultural land
3	Rangeland	31	Herbaceous

**Table 1. Anderson land use and land cover classification system**

Level I	Description	Level II	Description
		32	Shrub and brush rangeland
		33	Mixed rangeland
4	Forest land	41	Deciduous forest land
		42	Evergreen forest land
		43	Mixed forest land
5	Water	51	Streams and canals
		52	Lakes
		53	Reservoirs
		54	Bays and estuaries
6	Wetland	61	Forested wetland
		62	Nonforested wetland
7	Barren land	71	Dry salt flats
		72	Beaches
		73	Sandy areas other than beaches
		74	Bare exposed rock
		75	Strip mines, quarries, and gravel pits
		76	Transitional areas
		77	Mixed barren land
8	Tundra	81	Shrub and brush tundra
		82	Herbaceous tundra
		83	Bare ground tundra
		84	Wet tundra
		85	Mixed tundra
9	Perennial snow or ice	91	Perennial snowfields
		92	Glaciers

The USGS recently developed a LULC classification scheme (table 2), which is based on the Anderson system, but with a major change that addresses the problem of mixing LULC categories: LULC codes can overlap where there is developed land (U.S. Geological Survey, 1992a). Land cover information applies to all areas. Land use attributes are assigned where applicable. Unlike the Anderson system, or other LULC classification schemes currently in use today, each map feature can be described by more than one attribute. In residential areas, a code for residential use is applied as well as a code for the predominant land cover. Depending on the density of the development, the code will range from strictly manmade cover in the more densely developed areas to forested land in the less dense areas. This approach is very similar to that recently developed by the State of North Carolina (State of North Carolina, 1994), in which LULC are

addressed as separate layers that are combined with other layers for geographic analysis.

**Table 2.** USGS DLG land use and land cover codes, draft 4/94

LAND USE ATTRIBUTES	OPTIONAL DESCRIPTIVE CODES
<i>DEVELOPED LAND</i>	
340 0101 Residential	
340 0102 Commercial/light industrial	
340 0103 Heavy industrial	
340 0104 Transportation	
340 0105 Communications and utilities	
340 0106 Agricultural developed land	340 0610 Aquaculture site
340 0611 Confined feeding operation	
340 0107 Institutional	
340 0108 Entertainment and recreational	
340 0109 Extraction	
340 0110 Disposal	
340 0111 Memorial	
340 0112 Transitional	
340 0119 Developed land not classified	
LAND COVER ATTRIBUTES	DESCRIPTIVE CODES
<i>CROPLAND</i>	
340 0120 Orchards/groves/nurseries	340 0612 Irrigated
340 0121 Vines/bushes	
340 0122 Cropland	340 0612 Irrigated
340 0123 Cranberry bogs	
340 0129 Cultivated land not otherwise classified	340 0612 Irrigated
340 0130 Grassland	340 0603 Low level of management
	340 0604 Medium level of management
	340 0605 High level of management
<i>WOODY LAND</i>	
340 0140 Forest	340 0606 Deciduous
	340 0607 Evergreen
	340 0613 Clear cut
340 0608 Mangrove	
340 0141 Scrub/shrub	
<i>WATER</i>	
340 0150 Stream/river	340 0609 Intermittent

**Table 2.** USGS DLG land use and land cover codes, draft 4/94

LAND COVER ATTRIBUTES	OPTIONAL DESCRIPTIVE CODES
340 0151 Canal/ditch	
340 0152 Lake/pond	340 0614 Impounded
340 0609 Intermittent	
340 0153 Reservoir	340 0610 Aquaculture site
340 0154 Bay/estuary	
340 0156 Sea/ocean	
W340 0160 Wetland	
<i>EXPOSED LAND</i>	
340 0171 Rock	
340 0172 Boulders	
340 0173 Gravel	
340 0174 Sand	
340 0175 Mud	
340 0176 Evaporite deposits	
340 0179 Exposed land not classified	
340 0178 Manmade cover	340 0610 Aquaculture site
340 0180 Tundra	
<i>ICE</i>	
340 0190 Snowfield	
340 0191 Glacier	
340 0198 Unmapped area outside the United States	
340 0199 Unmapped area within the United States	

## STUDY AREA

As a test of this concept, a study area was chosen that represents a mix of land cover types and land use categories that are found in the United States within the bounds of a 1:100,000-scale quadrangle. However, realizing that one site will not fully represent the complete mix of possible land cover types found in the United States, an area was sought that contained a major metropolitan area, agricultural land, and a mix of land cover types, including deciduous and evergreen forests, and where the source material was readily available.

Much of the funding for the DOQ program in the past has come from users (for example, Natural Resources Conservation Service) needing this data source in agricultural areas. One of the few urban areas in which the data are available is the Portland/Vancouver metropolitan area (fig. 1).

Nearly the entire urban area was covered by DOQ data. In those developed areas for which there were no DOQ's available, in-house facilities were used to produce DOQ's of a similar quality that were acceptable for the 1:100,000-scale LULC compilation. Landsat TM data were available from the EROS Data Center for the entire study area. Locally produced LULC data also existed.

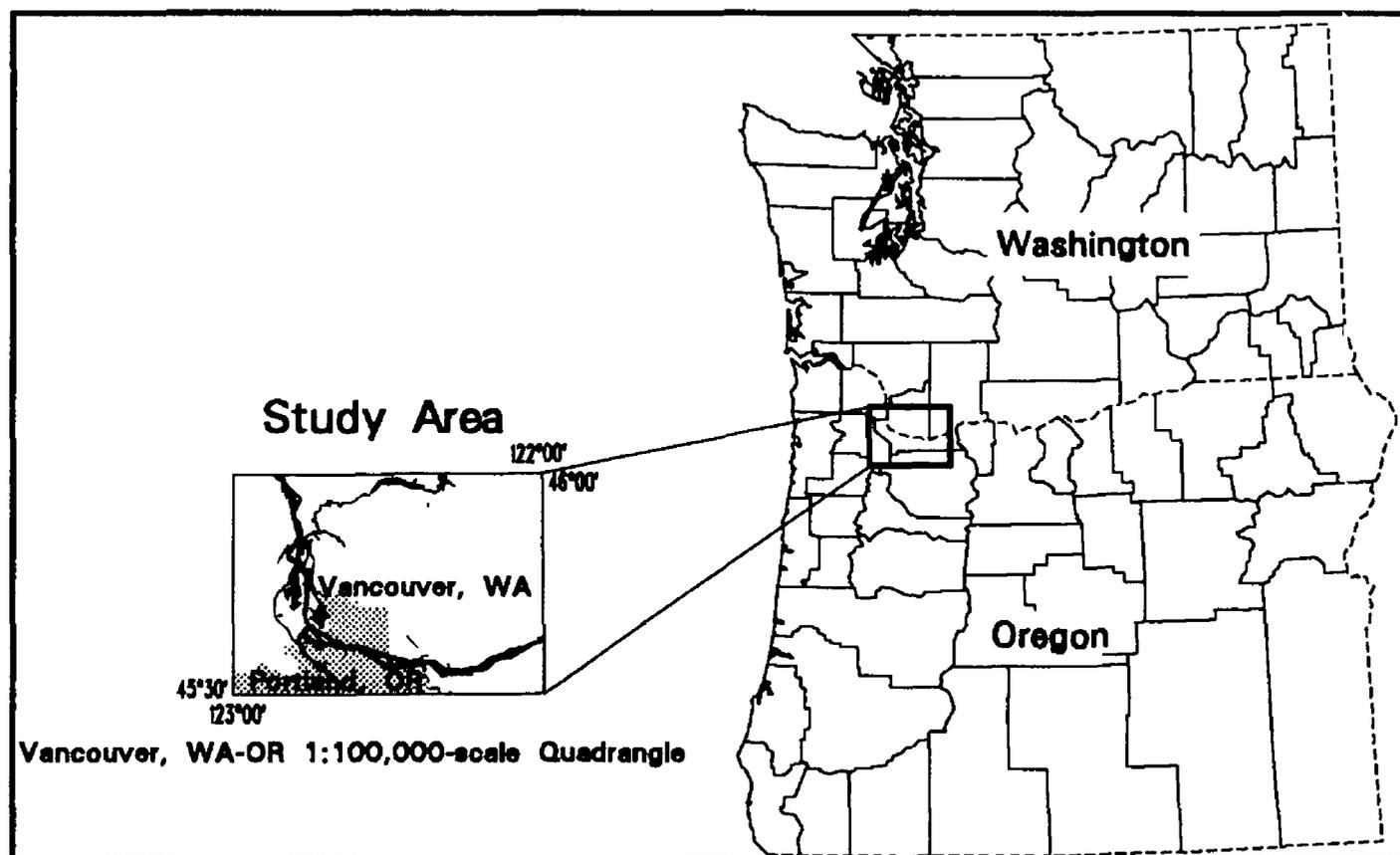


Figure 1. - The study area is the Vancouver, WA-OR 1:100,000-scale quadrangle that includes the Portland, Oreg. and Vancouver, Wash. metropolitan areas.

## DATA SOURCES

### Digital Orthophotoquads

The USGS is currently producing DOQ data for the conterminous United States (U.S. Geological Survey, 1990). Each DOQ covers a 3.75-minute area, one quarter of a standard series 7.5-minute quadrangle (1:24,000 scale). Approximately 3 percent of the DOQ's for the conterminous United States have been completed. The 1:40,000-scale air photos of the National Aerial Photography Program are the information source for the DOQ's produced by USGS (Light, 1993). Transparencies of each aerial photograph are scanned on a microdensitometer with a calibrated aperture, or other digitizing device. The digital image is rectified using camera-specific fiducial calibration values, fiducial point coordinates, ground and image coordinates of a few known control points, and a digital elevation model. The process removes image displacements due to terrain relief and aircraft pitch, roll, and yaw. The product is a spatially accurate raster image, with a 1-meter resolution. The image can then be used as a source for identifying natural and cultural features. At the USGS, DOQ's are the primary source for revision of 1:24,000- and 1:100,000-scale DIG features. The majority of the DOQ's being produced are single-band images of the black-and-white photographs. Where required by a State that is working cooperatively with the USGS, color-

infrared DOQ's are produced and can be entered into the data base.

### **Landsat Thematic Mapper (TM) Data**

Multispectral images of the U.S. sites have been acquired by the Landsat TM sensor since 1982. Images are acquired in 7 spectral bands with a nominal pixel resolution of 28.5 meters (with the exception of thermal-infrared band 6, which has a resolution of 120 meters). The TM sensor was designed to augment the capabilities of another Landsat instrument, the multispectral scanner, to measure and inventory Earth resources. In comparison, TM data have finer spatial resolution (28.5 vs. 57 meters/pixel), increased spectral resolution, and additional bands in the mid- and thermal-infrared region of the spectrum. These wavelength regions discriminate snow from clouds and aid in mapping vegetation and geologic formations. TM images are being used as sources for several statewide land cover mapping programs.

As stated above, the MRLC consortium of Federal agencies is purchasing TM scene coverage of the conterminous United States (Shaw and others, 1993). Images will have noise and distortions removed, be resampled to a UTM projection, and have terrain displacements reduced using elevation data. The multispectral images will also be spectrally clustered and available for interpretation in that form. For the clustering, the six spectral bands of TM data (excluding the thermal band) are compressed into a single-band image containing a large number of clusters (241) using an unsupervised clustering technique (Kelly and White, 1993). The TM scene (Path 46 Row 28) chosen for this study is one of the MRLC scenes provided by a consortium member (National Biological Service Gap Analysis Program). The image was resampled to a 25-meter UTM grid by EOSAT. It was acquired over the Vancouver/Portland area on July 7, 1991 and is cloud free.

### **Local Agency LULC Data**

Two major cities are located in the study area and form a large metropolitan region separated by the Columbia River: Vancouver, Wash. and Portland, Oreg. Vancouver is located in Clark County, which has collected land cover for the city by photointerpretation of aerial photos. Clark County land use information was produced from parcel data that was assigned land use codes based on the primary property type. In the Portland, Oreg. area, the regional government agency, Portland Metro, has collected detailed land use information from aerial photos, and land cover information in surrounding nonurban areas from TM data.

## **METHODS**

Three methods of collecting LULC data were investigated; these were (1) *low resolution source*: collection of land cover data from automated processing of Landsat TM data using the Spectrum classification program, developed by the Los Alamos National Laboratory (Kelly and White, 1993), (2) *high resolution source*: on-screen delineation of LULC information from DOQ using a

menu-driven interface to a commercial geographic information system (GIS) developed by the USGS, revision, product generation (RevPG), and (3) *local agency data*: conversion of existing local agency data. The results from these sources were then compared for a small study area, the Battle Ground, Washington 7.5-minute quadrangle (fig. 2).

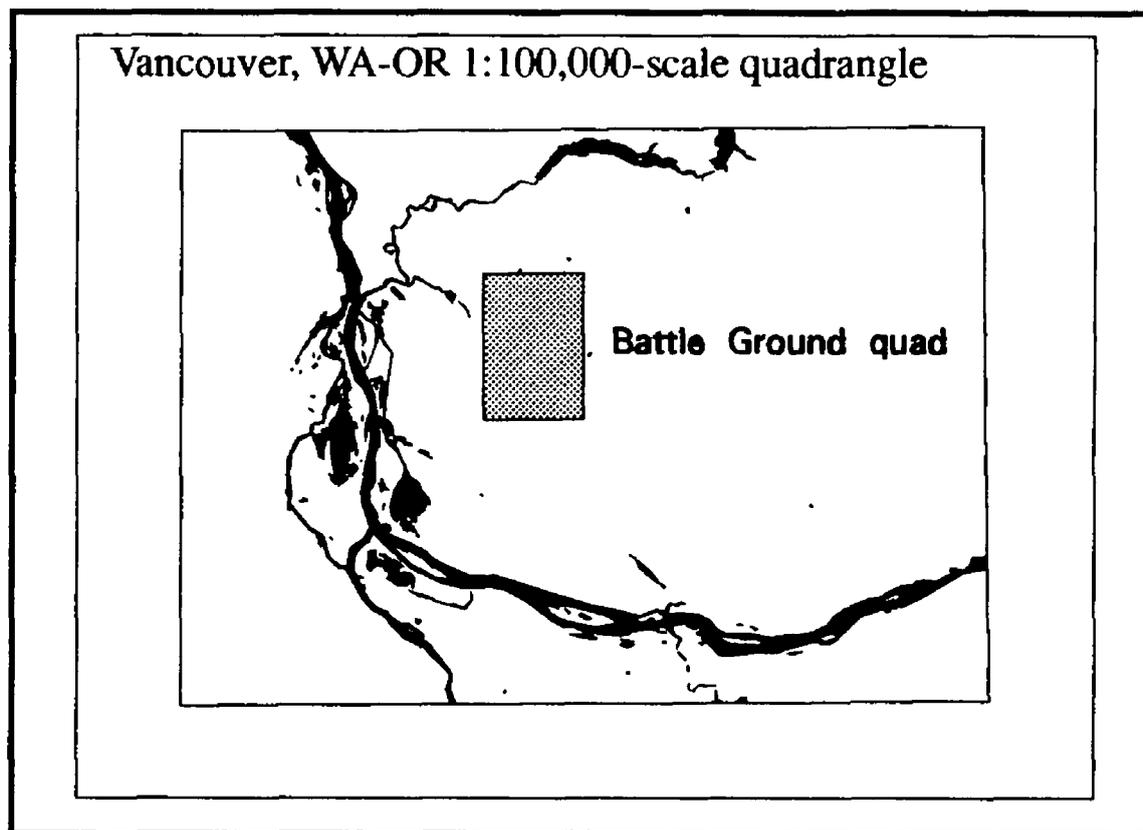


Figure 2. - Battle Ground, Wash. 7.5-minute sample test site.

### **Low resolution source (Landsat TM data)**

A subsection of the full TM scene was extracted for analysis that covered the Vancouver 100,000-scale quadrangle and extended at least 1 minute of latitude and longitude beyond the quadrangle boundary. This subsection is a rectangle of pixels 3,350 columns wide and 2,500 rows high that contains approximately one-ninth of the pixels in the full registered TM scene. Six bands were selected for analysis: only the thermal-infrared band (6) was excluded.

Using an unsupervised clustering method developed at Los Alamos National Laboratory (Kelly and White, 1993), the Vancouver TM image subsection was reduced to a single band of 240 cluster values. The clustering program uses random samples of pixels from the multispectral image to compute cluster centers and adjust them to best describe the image values in the samples. Statistics were computed for 240 clusters in 12 iterations, each with a random sample of more than 1 million pixels. The classification program assigns each image pixel to one of the 240 clusters, updates the cluster statistics to reflect the full range of pixels assigned, and writes an output image that contains the cluster statistics in its header. It took approximately 42 minutes to "project" each input pixel to one of the 240 clusters. The data were then ready to be interpreted using Spectrum.

Spectrum uses cluster statistics stored in the image header (cluster per-band means and the covariance matrix) to simulate the original multispectral image. As shown in an earlier study (Lillesand and Kiefer, 1994, p. 572), interpreters prefer using different combinations of TM image bands depending on the features of interest. Spectrum uses the cluster means to simulate the appearance

of different bands displayed in red, green, and blue color channels.

A team of three image interpreters analyzed the clustered image with Spectrum. Using small scale (1:65,000 and 1:40,000) color-infrared aerial photographs as a guide, clusters were assigned to land cover classes. These cluster assignments were checked throughout the 30- x 60-minute quadrangle area, then either retained or rejected. Several different sources were used to group the 240 clusters into 11 land cover classes. In an iterative fashion, photographs, maps of the area, Spectrum's statistical viewing tools (scatterplot, spectral reflectance curves), and existing GIS layers (hydrography, transportation, GIRAS), were used to assign each cluster to one land cover class. The TM interpreters started with the full set of DLG LULC codes (table 2). They found that only a subset of the codes could be identified in the TM image using these methods (table 3), and these codes were used as the final set of LULC attributes.

The TM image was generalized and converted from a raster image to a vector land use land cover GIS layer for integration with the DOQ-derived LULC and Clark County GIS information. Using ARC/INFO's GRID raster GIS, the 11 classes of the interpreted TM image were smoothed twice to blend isolated pixels into neighboring regions. The GRID boundary clean program then smoothed boundaries between regions. Each separate region of cells was identified and labeled. Any regions smaller than 5 acres were discarded and the surrounding pixels were used to fill in the gaps. This was also done for a 10-acre minimum. These GRID layers were vectorized to create 5- and 10-acre minimum mapping unit (MMU) polygon coverages with 21,229 and 10,417 polygons, respectively. Polygons of the 10-acre MMU coverage were attributed with the minor codes from the LULC standards. Some editing was performed to manually attribute lakes (minor2 = 152), forest clear-cuts (minor2 = 613), lava flows (minor1 = 171), and mines (minor1 = 109).

**Table 3.** Land cover classes applied to TM data

TM CODE	DESCRIPTION	Number of Clusters
1	Mixed residential cover	10
2	Cropland	19
3	Grassland	24
4	Forest (deciduous)	35
5	Forest (evergreen)	65
6	Scrub/shrub	21
7	Water	10
8	Wetlands	5
9	Exposed land	29
10	Manmade cover	15
11	Exposed land with vegetation	7

## **High resolution source- DOQ data**

In support of the revision of the 1:100,000-scale DLG data, a software production system was developed to allow photointerpreters to use the DOQ data as a background source upon which linear features such as roads and streams could be displayed. This system, RevPG, was developed using a commercial GIS, ARC/INFO. RevPG enables the compiler to utilize ARC/INFO commands to (1) make onscreen changes to DLG linework, (2) add appropriate DLG attribute codes (major and minor codes), (3) use onscreen symbology, (4) detect errors with semiautomated tools, (5) match edges with adjacent sections, and (6) assess check plots for data quality. The advantage of using RevPG is its user friendly interface between the robust and complex ARC/INFO GIS software and the human compiler. This interface enables the compiler to use GIS tools without extensive knowledge of the software upon which RevPG is based. RevPG has been modified to allow efficient collection of LULC data from the DOQ.

Previous LULC mapping techniques (manual compilation involving inking, scribing, scanning, vectorizing, and tagging) were slow and expensive. With onscreen digitizing, the earlier LULC collection techniques may be completely replaced. The goal of this phase of the project was to determine if RevPG software and black-and-white DOQ's could be used to collect LULC data at a reasonable cost. The final product was a 1:100,000-scale digital map.

Because TM collection was generally suitable only for the collection of the rural land cover features (see table 3), over 50 percent of the Vancouver WA-OR 1:100,000-scale quadrangle had to be collected using the high resolution mapping technique. The high resolution mapping technique allowed the collection of all of the urban LULC features displayed in table 2. Out of 128 DOQ's that cover a 1:100,000-scale map, 67 DOQ's were used to compile the LULC data. The qualification to use a DOQ for collection was that it had to contain at least 10 acres of urban features. Only black-and-white DOQ's were available for the study area. When the DOQ's were loaded onto the workstation, each image used approximately 48 megabytes of disk space.

The hydrography and road DLG's were converted to ARC coverages and used as a backdrop to obtain registration and alignment with the LULC features. The compilers digitized the LULC features as interpreted from the DOQ, using 7.5-minute quadrangles as an ancillary source. When features were delineated, they were coded with attributes that defined each polygon. While working in the RevPG environment, the compiler could immediately see both the symbol and attribute of each feature as it was coded.

The time to compile per DOQ varied depending upon the density of the urban features. If a DOQ contained high-density urban areas, the compiler spent an average of 4.7 hours collecting these data. In less dense, suburban type areas, an average of 3.2 hours was spent, and where the DOQ was almost entirely rural, the compiler usually completed each DOQ in an average of 1.9 hours. The total time to collect from all 67 DOQ's was 183 hours with an overall average of 2.7 hours per DOQ. The team consisted of individuals unfamiliar with computers, but were familiar with compiling LULC information. A project coordinator was responsible for work flow, file management, software and system problems, and general assistance on cartographic issues.

**Local agency data**

Detailed LULC data were available from Clark County (fig. 3), Washington

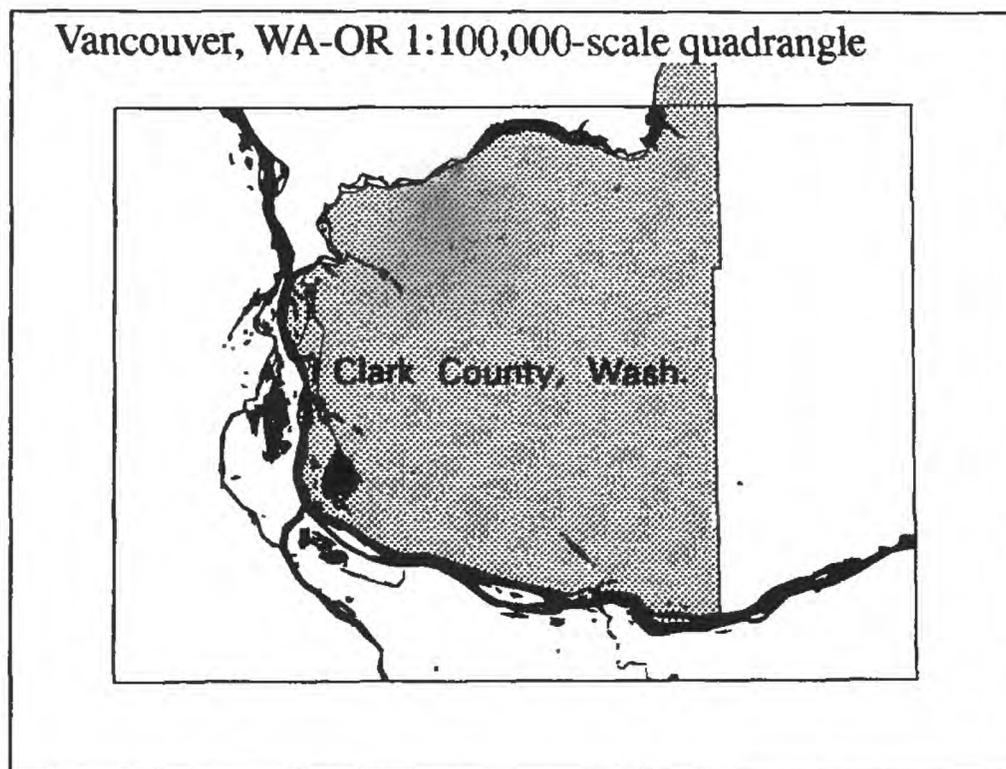


Figure 3. - Outline of Clark County, Wash.

Department of Assessment and GIS. A small subset of the data was chosen for comparison with the other two sources, and to evaluate the feasibility of converting the data to the USGS LULC classification scheme. The Clark County data were collected and stored as separate files. The land use data are in two forms; parcel-boundary data to which each parcel is assigned a land use code (property type code), and a file derived from the property-boundary data. The land use codes assigned to the property data are based on the most important use that exists in that parcel for tax purposes, and the codes are useful for small size parcels (table 4). The data are less useful for large parcels in which a small portion of the area is devoted to the primary use, and the remaining part of the parcel is either a less important use or predominantly unused, such as large rural residential parcels. Because of this, Clark County staff have created the derived land use file in which only those parcels below a minimum size (5 acres) are considered valid, and the remaining area is assigned a land cover class from the land cover file.

**Table 4.** Parcel-based land use codes, Clark County, Wash.

10	HOUSING UNITS, SINGLE FAMILY
11	Single family unit not sharing structure with other uses.
13	Single family unit sharing structure or premises with other major use.
14	Single family unit subsidiary to a "more important" use.
15	Nonresidential structure used as a single family dwelling.
16	Mobile home converted to permanent structure.
17	Single family condominium unit.

**Table 4.** Parcel-based land use codes, Clark County, Wash.

18	Single family cooperative housing unit.
19	Single family housing not elsewhere classified
...	...
990	UNUSED
991	Unused land cleared, etc.

The land cover information was collected by photointerpretation of large-scale aerial photos. Although very detailed, the classification scheme is not compatible with the USGS scheme. Land cover types are lumped into general classes, and percentage of cover of these classes is interpreted (table 5).

**Table 5.** Photointerpreted land cover classes, Clark County, Wash.

Land cover code	Description
1	Agriculture
2	Agriculture tilled
3	Mixed open
4	Open unused
5	Open recreational
6	Forest
7	Forest clear cut
8	Water
9	Less than 5% natural cover
10	Less than 33% open natural ground cover
11	Less than 33% mixed open natural cover
12	Less than 33% Forest natural ground cover
13	33-66% open natural ground cover
14	33-66% mixed open natural ground cover
15	33-66% forest natural ground cover
16	More than 66% open natural ground cover
17	More than 66% mixed open natural ground cover
18	More than 66% forest natural ground cover
19	Mixed recreational
20	Brush
21	Young forest

## RESULTS

### Classification of land cover from low resolution source

The TM processing was conducted in a short time by experienced image processing analysts. The interpretation of the clustered data is an intuitive process, and if proper ground truth information is available, then rapid delineation of land cover classes is possible. Due to limited time and resources, an accuracy assessment was not conducted, so the ability to make quantitative statements about the classified data is limited. A brief ground reconnaissance was conducted, and some of the areas of confusion were checked, such as areas of mixed deciduous and coniferous forest. The multispectral capabilities of the TM data allowed the interpreter to distinguish between these two vegetation classes more accurately than did the gray-scale DOQ.

The interpretation of the TM data for the study was done from a single image dated July 7, 1991. Agricultural fields that were in various stages of growth and cultivation were difficult to distinguish from other vegetation or exposed ground. The information from a second scene, acquired at a different time, such as early spring, would have greatly improved the classification of the agricultural areas. An additional scene could be accommodated by the Spectrum software by using additional bands as input to the clustering algorithm. Time constraints did not allow for a complete accuracy assessment, including collection of ground truth information. DOQ data used for the manual collection were used as a reference source for a preliminary accuracy assessment. Approximately 20 pixels per category were randomly selected from the TM classified data. These pixels were then converted to points based on the coordinates at the center of each pixel, and compared with the DOQ source. This error analysis pointed to problems in the classification system used for processing the TM data. In most cases, the exposed surface categories (9 & 11) were identified as cropland (2) in the reference source (table 3). These two categories could have been merged with the cropland category, however recent clearcut areas also appeared in this group.

Another problem class was residential areas (1), which were often categorized as either manmade surface (10) or exposed ground with vegetation (11). The errors with the residential areas and the cropland areas indicate the difficulties of inferring land use for spectral signatures.

An unexpected error appeared along the edges of several water bodies, where edge water pixels are classed as wetlands. Because there was a 2-year time difference between the TM data and the DOQ reference data, these areas may have been wetlands in the TM scene. Most likely, these mixed water/vegetation pixels are similar to wetland areas covered by vegetation. There was also a problem distinguishing many of the scrub/shrub areas from cropland. The TM cropland category includes pasture, and many pasture areas were classed as scrub/shrub.

Land uses were not mapped from the TM data, although the mixed residential areas corresponds closely to the higher density residential land use areas. Manmade cover within the urban areas was readily determined. The land cover, collected in areas that were mapped using the DOQ data were comparable, although the TM data provided a more detailed characterization of the cover. In the manual collection process, the dominant cover within a 10-acre area (up to 64 TM pixels) was

selected, whereas the TM data identified specific pixels of various cover types, as many as 64 pixels for the same 10-acre area. Instead of manually collecting both cover and use within these urban areas, as was initially proposed, a more efficient and improved method is to use the cover information from the TM data with only the use information from the DOQ. This allows the compiler to collect land use more rapidly without the added burden of determining the combination of LULC within 5- or 10-acre minimum mapping units.

### **Classification of land use and land cover from high resolution source.**

Compilation from the DOQ source presented several problems. The DOQ files are very large. Each file represents a quarter of a 7.5-minute quadrangle. There are thirty-two 7.5-minute quadrangles per 30- x 60-minute quadrangle so that there are 128 DOQ's per 30- x 60-minute quadrangle, approximately 6.4 gigabytes of data. Only a few of the files could be stored online. When a section was completed, those DOQ files were removed and new ones loaded, taking time away from the compilation process. The 1-m DOQ source provides a detailed look at the surface being mapped, allowing the collection of most land use information. However, because it contains a single band of gray-scale information, subtle differences in vegetation type are not easily distinguishable. The compilers had a difficult time distinguishing between deciduous and evergreen forest and shrub vegetation, between lawns and concrete in dense urban areas, as well as identifying the location and extent of wetlands.

The compilers doing the collection were unaccustomed to working in a digital environment. One benefit of using the menu-driven digital revision system, RevPG, was that the compilers became productive in a short time. Brief training sessions were held on the user interface, the terminology used for describing digital features, and the use of the mouse. Because they had compiled LULC data in an analog setting using the draft DLG standards, they were familiar with the application of the LULC codes. Within several days, the compilers began to collect the LULC features.

Because of the same constraints that applied to the TM processing, a thorough accuracy assessment was not conducted on the DLG LULC data collection. The points that were randomly selected from the TM categories were also checked against the DLG LULC data. The major problems with the DLG land cover data related to the inability to differentiate between deciduous and evergreen forest from the panchromatic DOQ data. The same DOQ data used for the compilation were also used for the accuracy assessment. Subtle differences in tone were apparent in the DOQ, and these related well to the deciduous and evergreen classes derived from the multispectral TM data. The compilers were not able to consistently make this distinction, mainly because they were not using the full resolution of the DOQ. A decision was made early in the process to limit the amount of zooming on features to speed the collection process.

The major problems occurred in dense residential areas. The land cover was consistently classed as high management grassland regardless of the residential density. In many cases, the combined area of rooftops, paved surfaces, and other manmade features exceeded the area of vegetative cover. There is a considerable difference between areas of very low density rural residential areas and high density urban residential areas with respect to evaporation, infiltration, runoff, and ther-

mal characteristics, which are important physical parameters used in many hydrologic models. The compilers need training to recognize the predominance of manmade features in an urban area to identify the areas accordingly.

A similar analysis was conducted on the land use attributes for these same areas. Many of the points selected did not occur in land use categories. Of those that did, the majority were in residential areas, and these were accurately delineated in the DLG data. Problem areas occurred where an area was labelled as residential, but the density of houses was below the minimum criteria for residential (4 houses per 10 acres). There also appeared to be problems in the separation of commercial from industrial, but this needs to be pursued further with ancillary sources.

### **Integration of local agency data**

The parcel-based data that was converted to the USGS LULC codes was compared with the DOQ-derived land use data (table 6). As expected, those larger parcels that were dominated by a

**Table 6.** Comparison of parcel-based and DOQ derived land use percentages for Battle Ground quadrangle

DLG Minor code	Parcel-based land use	DLG land use
101	51.71	17.81
102	0.60	1.01
103	0.08	
104	3.37	0.10
105	0.09	
106		0.08
107	1.08	
108	1.00	0.03
109	0.39	0.50
111	0.01	
122	6.05	40.34

single use of small spatial extent tended to overestimate the area of that use. This occurred most commonly with rural residential parcels. The total residential use in the parcel-based data

accounted for 51.71 percent of the land area vs. 17.81 percent in the DOQ derived data. (fig. 4).

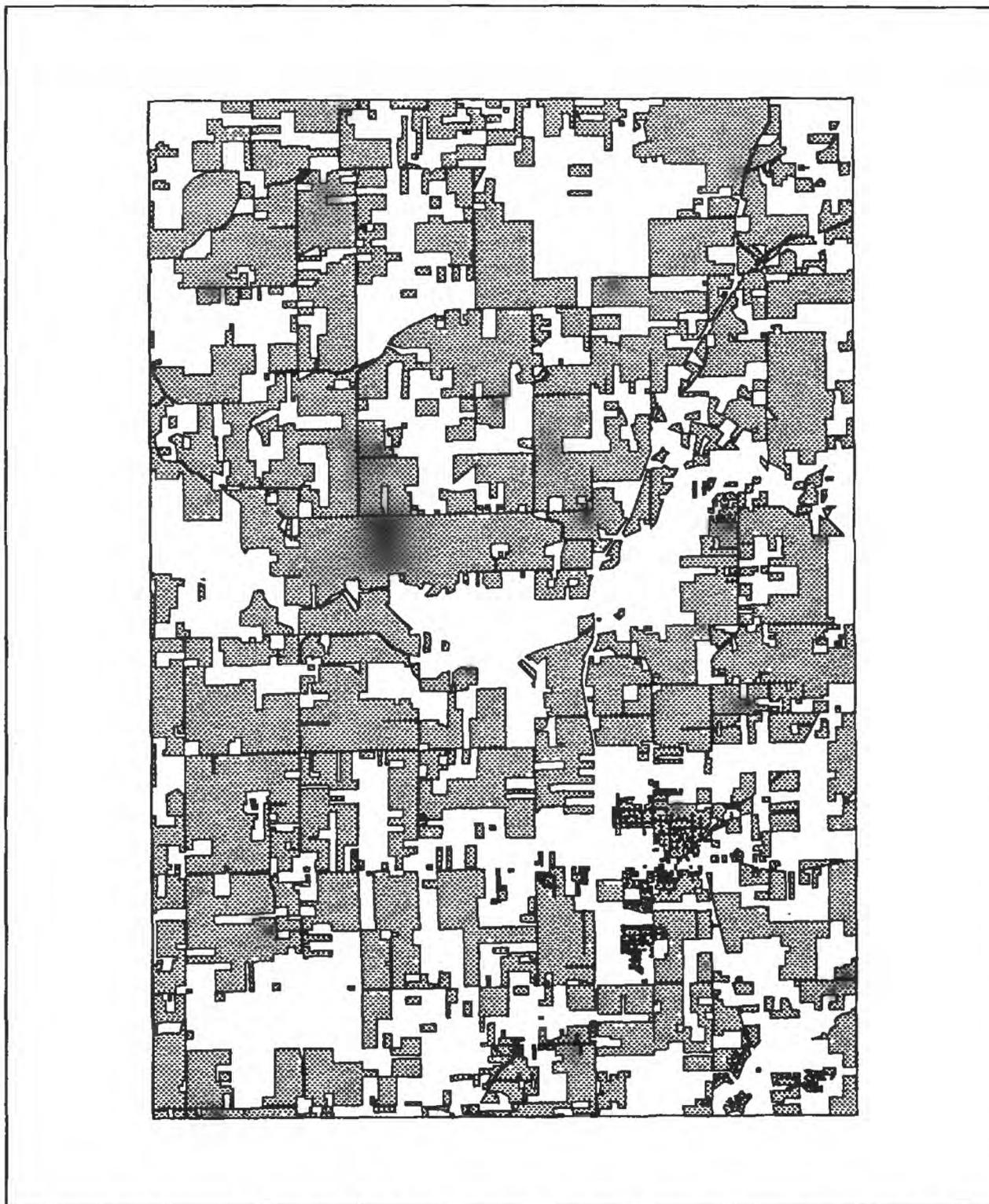


Figure 4. - Parcel-based land use data for Clark County, Wash., with 51.71 percent of the area identified as residential land.

As was done by Clark County for their derived land use map, the larger residential use polygons (greater than 5 acres) were removed from the file and the result was comparable to the USGS land

use (17.81 percent vs. 17.88 percent (fig. 5).



Figure 5. - Parcels less than 5 acres from parcel-based land use data (shaded in gray) compared with DLG land use data (cross hatched areas).

Another potential use of local agency data is as a supplemental source during the compilation process. It is often difficult for the photointerpreter to identify a particular land use within an urban area. From experience and with the use of some analog supplemental materials, for example USGS 7.5-minute quadrangles, most land uses can be identified. However, in those cases where there are questions, the digital data can be used as a reference source. In the digital environment of RevPG, a supplemental digital source such as the local land use data can be present on the system but not actively displayed. When there is a question about an area, the analyst can use the cursor to point to the area, query the local data, and display the land use description assigned by the local agency. In most cases, that information is enough to assign a DLG code to the delineated polygon.

## **Integration of results from low and high resolution data interpretation.**

The above methods produce several separate data sets, with both raster and vector data structures. One goal of the proof-of-concept was to determine the feasibility of merging one or more of the above data sources. An attempt was made to merge the raster TM-derived land cover data with the vector DOQ-derived LULC data. As described above, the TM LULC data were converted to a vector format and the vectors were then merged with the DOQ LULC data. The larger, more contiguous areas (grassland and forest land) increased in size at the expense of more isolated clumps of pixels (residential cover and exposed land with vegetation) due to the filtering and application of a minimum mapping criterium of 10 acres (table 7).

**Table 7.** Comparison of 25-m classified TM data and the resultant vectorized land cover data (percentages for Battle Ground quadrangle)

DLG minor code (cover)	DLG Minor code (descriptive)	Percent of quadrangle area (raster 25-m pixels)	Percent of quadrangle area (vector 10-acre mmu)	Class Description
101		1.72	0.35	Residential
122		17.04	11.01	Cropland
130		19.92	23.31	Grassland
140	606	20.70	28.33	Forest - Deciduous
140	607	9.76	11.55	Forest - Coniferous
141		16.22	12.97	Scrub/shrub
150		0.13	0.11	Water
160		0.33	0.08	Wetland
179		9.70	10.72	Exposed Land
178		0.73	0.74	Artificial surface
179		3.74	0.82	Exposed land w/vegetation

The TM-derived land cover polygons were then used to replace the photointerpreted land cover data in the DOQ-derived coverage. The land use polygons from the DOQ-derived data were digitally merged with these land cover polygons. There are several advantages to this method. The multispectral nature of the TM data allows a better differentiation between the deciduous and evergreen vegetation than the panchromatic DOQ data. The collection time is very rapid, and con-

tingent upon an accuracy assessment of both sources, may be more cost effective than collecting from DOQ. The photointerpreter is then left to concentrate on the collection of the land use information, which is not possible to derive from the TM data.

A disadvantage to this method is that the TM derived data are altered significantly after the above filtering and aggregation steps (table 7). A number of decisions were made regarding this process that affected the resultant boundaries. These decisions are better left to the user of the data so that the variables in the process can be controlled. Another disadvantage is that a large amount of information is lost in the conversion process of smoothing and aggregating the data. Although there are some problems involved in working with two different data models, raster and vector, many GIS's that are being used today have the capability of manipulating and displaying both models in the same system. Another consideration in using the merged outcome is that the data represent two different measures of accuracy, and this must be clearly specified in the metadata. Although it is feasible to merge the two, it was determined that the two sources be considered as separate but complimentary sources.

## CONCLUSIONS

An initial assumption was made that satellite data were sufficient for mapping land cover in most non urban areas, and that DOQ data will be used for collecting LULC data in the urban areas. Local agency data were evaluated for use as a direct source of urban land use or land cover information, or as a supplement to the collection of that data. This study shows that the TM data are not only sufficient for rural land cover information, but also for urban land cover information. Some land cover types require multitemporal images for proper delineation (agriculture), and others may not be identifiable from the TM data (clear cut forest land). A more powerful approach to deriving land cover information from TM data is being researched at the USGS EROS Data Center, in which ancillary data layers such as soils, elevation, slope, aspect, ecoregion boundaries, and others are collected and stored in a data base registered to the satellite data (Brown and others, 1993). Techniques developed for using a 1-km resolution data base and AVHRR data are being adapted for use at the regional scale using TM data. This flexible data base will allow the extraction of user defined land cover information. The TM based effort is initially being tested in the High Plains and will be expanded to other regions over the next several years.

Although it is feasible to produce vector files from the raster data, decisions that are made during the conversion process affect the final result and modify the original data. Thus, it is more appropriate to keep the data in the raster form and allow the users of the data to conduct the conversion from the original classified raster data to a vector format.

Both land use information and land cover types were mapped from the DOQ source, but if the TM classified land cover data is sufficient for the urban areas, only land use data need to be collected from the DOQ source, speeding the collection process for the urban areas. This assumes that both sources are available and used for collection in the same area. An advantage to this procedure is that percentages of land cover within land use polygons can be obtained as opposed to a single dominant land use, which is often useful for modeling purposes.

Local agency land cover data were not used as a primary source due to differences in classification schemes. Local land cover data are not as useful as the detailed land use data. The land use data were converted to the DLG coding scheme and compared favorably. However, given the wide diversity of classification schemes currently in use at the local and regional agency level, the most appropriate use for the local agency data is as a supplemental source to the compilation process.

## REFERENCES

- Anderson, J.R., Hardy, E.E., Roach, J.T., and Witmer, R.E., 1976, A Land Use and Land Cover Classification System for Use with Remote Sensor Data: U.S. Geological Survey Professional Paper 964, 28p.
- Brown, J.F., Loveland, T.R., Merchant, J.W., Reed, B.C., and Ohlen, D.O., 1993, Using multi-source data in global land-cover characterization: Concepts, requirements, and methods: *Photogrammetric Engineering and Remote Sensing* v. 59, no. 6, p. 977-987.
- Kelly, P.M. and White J.M., 1993, Preprocessing remotely sensed data for efficient analysis and classification, *Applications of Artificial Intelligence 1993: Knowledge-Based Systems in Aerospace and Industry, Proceedings: SPIE 1993*, p. 24-30.
- Light, D.L., 1993, The National Aerial Photography Program as a Geographic Information System Resource: *Photogrammetric Engineering and Remote Sensing* v. 59, no. 1, p. 61-65.
- Lillesand, T.M. and Keifer, R.W., 1994, *Remote Sensing and Image Interpretation* (3rd ed.): John Wiley, New York, 750 p.
- Scientific Assessment and Strategy Team, 1994, *A Blueprint for Change; Part V, Science for Floodplain Management into the 21st Century*: Washington, D.C., U.S. Government Printing Office, 272 p.
- Shaw, D.M., Field, D.W., Holm, T.M., Jennings, M.D., Sturdevant, J.A., Thelin, G.P., and Worthy, L.D., 1993, *An Innovative Partnership for National Environmental Assessment, Pecora 12, A Symposium on: Land Information from Space-Based Systems*.
- State of North Carolina, 1994, *A Standard Classification System for the Mapping of Land Use and Land Cover*: Center for Geographic Information and Analysis, 66 p.
- U.S. Geological Survey, 1990, *Forum on Orthophotography* NASA Goddard Space Flight Center, Greenbelt, Maryland, May 15, 1990, p. 25-27.
- \_\_\_\_\_, 1992a, *Draft Standards for Digital Line Graphs for Land Use and Land Cover: National Mapping Program Technical Instructions*.
- \_\_\_\_\_, 1992b, *Standards for Digital Orthophotos: National Mapping Program Technical Instructions*, 50 p.