CHAPTER 9

Assessing the Accuracy of Satellite-Derived Land-Cover Classification Using Historical Aerial Photography, Digital Orthophoto Quadrangles, and Airborne Video Data

Susan M. Skirvin, William G. Kepner, Stuart E. Marsh, Samuel E. Drake, John K. Maingi, Curtis M. Edmonds, Christopher J. Watts, and David R. Williams

CONTENTS

9.1 Introduction....116
9.2 Background....116
  9.2.1 Upper San Pedro Watershed Study Area....116
  9.2.2 Reference Data Sources for Accuracy Assessment....117
  9.2.3 Reporting Accuracy Assessment Results....118
9.3 Methods....118
  9.3.1 Image Classification....118
  9.3.2 Sampling Design....119
  9.3.3 Historical Aerial Photography....120
    9.3.3.1 Image Collection, Preparation, and Site Selection....121
    9.3.3.2 Photograph Interpretation and Assessment....121
  9.3.4 Digital Orthophoto Quadrangles....121
    9.3.4.1 Interpreter Calibration....121
    9.3.4.2 Sample Point Selection....122
  9.3.5 Airborne Videography....122
    9.3.5.1 Video and GIS Data Preparation....122
    9.3.5.2 Video Sample Point Selection....122
    9.3.5.3 Random Frame Selection and Evaluation....123
9.4 Results....123
  9.4.1 Aerial Photography Method....123
  9.4.2 Digital Orthophoto Quadrangle Method....124
  9.4.3 Airborne Videography Method....124
9.5 Discussion....125
  9.5.1 Map Accuracies....125
  9.5.2 Class Confusion....125
  9.5.3 Future Research....128

9.6 Conclusions.....128
9.7 Summary.....128
Acknowledgments.....129
References.....129

9.1 INTRODUCTION

There is intense interest among federal agencies, states, and the public to evaluate environmental conditions on community, watershed, regional, and national scales. Advances in computer technology, geographic information systems (GIS), and the use of remotely sensed image data have provided the first opportunity to assess ecological resource conditions on a number of scales and to determine cross-scale relationships between landscape composition and pattern, fundamental ecological processes, and ecological goods and services. Providing quantifiable information on the thematic and spatial accuracy of land-cover (LC) data derived from remotely sensed sources is a fundamental step in achieving goals related to performing large spatial assessments using space-based technologies.

Remotely sensed imagery obtained from Earth-observing satellites now spans three decades, making possible the mapping of LC across large regions by the classification of satellite images. However, the accuracy of these derived maps must be known as a condition of the classification. Theoretically, the best reference data against which to evaluate classifications are those collected on the ground at or near the time of satellite overpass. However, such data are rarely available for retrospective multitemporal studies, thus mandating the use of alternative data sources. Accordingly, the U.S. Environmental Protection Agency (EPA) has established a priority research area for the development and implementation of methods to document the accuracy of classified LC and land characteristics databases (Jones et al., 2000).

To meet the ever-growing need to generate reliable LC products from current and historical satellite remote sensing data, the accuracy of derived products must be assessed using methods that are both effective and efficient. Therefore, our objective was to demonstrate the viability of utilizing new high-resolution digital orthophotography along with other airborne data as an effective substitute when historical ground-sampled data were not available. The achievement of consistent accuracy assessment results using these diverse sources of reference data would indicate that these techniques could be more widely applied in retrospective LC studies.

In this study, classification accuracies for four separate LC maps of the San Pedro River watershed in southeastern Arizona and northeastern Sonora, Mexico (Figure 9.1) were evaluated using historical aerial photography, digital orthophoto quadrangles, and high-resolution airborne video. Landsat Multispectral Scanner (MSS) data (60-m pixels) were classified for the years 1973, 1986, and 1992. Lastly, 1997 Landsat Thematic Mapper (TM) data (30-m pixels) were resampled to 60 m to match the MSS resolution and classified. All data were analyzed at the Instituto del Medio Ambiente y el Desarrollo Sustentable del Estado de Sonora (IMADES) in Hermosillo, Mexico. Map accuracy was assessed by Lockheed-Martin (Las Vegas, Nevada) for 1973 and 1986 and at the University of Arizona (Tucson, Arizona) for 1992 and 1997. This study incorporated previous accuracy assessment methods developed for the San Pedro watershed by Skirvin et al. (2000) and Maingi et al. (2002).

9.2 BACKGROUND

9.2.1 Upper San Pedro Watershed Study Area

The study location comprised the upper watershed of the San Pedro River, which originates in Sonora, Mexico, and flows north into southeastern Arizona. Covering approximately 7600 km^2 (5800 km^2 in Arizona and 1800 km^2 in Sonora, Mexico), this area represents the transition between

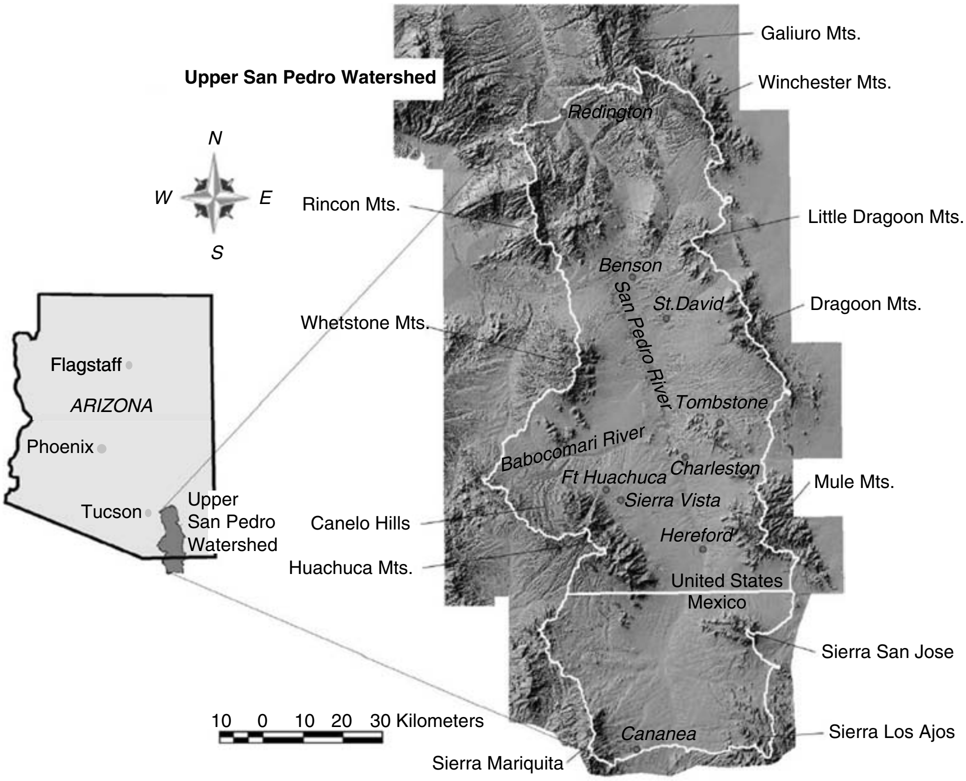

Figure 9.1 Location of the upper San Pedro River watershed study area with shaded relief map.

the Sonoran and Chihuahuan deserts, and topography, climate, and vegetation vary substantially across the watershed. Elevation ranges from 900 to 2900 m and annual rainfall ranges from 300 to 750 mm. Biome types include desertscrub, grasslands, oak woodland-savannah, mesquite woodland, riparian forest, and conifer forest, with limited areas of irrigated agriculture. Urban areas, including several small towns and the rapidly growing U.S. city of Sierra Vista, are fringed by low-density development that also occurs far from population centers. Numerous geospatial data sets covering the upper San Pedro watershed can be viewed and downloaded at the U.S. Environmental Protection Agency San Pedro Web site (USEPA, 2000).

9.2.2 Reference Data Sources for Accuracy Assessment

Aerial photography has long served in the creation of LC maps, both as a mapping base and more recently as a source of higher-resolution reference data for comparison with maps produced by classification of satellite imagery. Coverage for the conterminous U.S. at a scale of 1:40,000 is available through the National Aerial Photography Program (NAPP) and is scheduled for update on a 10-year, repeating cycle. Digital orthophoto quarter quadrangles (DOQQs) are produced from the 1:40,000-scale NAPP or equivalent high-altitude aerial photography that has been orthorectified using digital elevation models (DEMs) and ground control points of known location. A DOQQ image pixel represents 1 m^2 on the ground, permitting detection of landscape features as small as approximately 2 m in diameter. However, the image analyst may need site visits and/or supplementary higher-resolution images to visually calibrate for DOQQ-based LC interpretation.

Marsh et al. (1994) described the utility of airborne video data as a cost-effective means to acquire significant numbers of reference data samples for classification accuracy assessment. In that study, very similar classification accuracies were derived from airborne video reference data and from aerial color 35-mm reference photography acquired under the same conditions. With the addition of Global Positioning System (GPS) coordinate data encoded directly onto the videotape for georeferencing, sample points can be rapidly located for interpretation during playback.

9.2.3 Reporting Accuracy Assessment Results

The current standard for reporting results of classification accuracy assessment focuses on the error or confusion matrix, which summarizes the comparison of map class labels with reference data labels. Some easily computed summary statistics for the error matrix include overall map accuracy, proportion correct by classes (user and producer accuracy), and errors of omission and commission. Additional summary statistics usually include a Kappa (Khat) coefficient that adjusts the overall proportion correct for the possibility of chance agreement (Congalton et al., 1983; Rosenfield and Fitzpatrick-Lins, 1986; Congalton and Green, 1999). Although Kappa is widely used, some authors have criticized its characterization of actual map accuracy (Foody, 1992). Ma and Redmond (1995) proposed some alternatives to the Kappa coefficient, including a Tau statistic that is more readily computed and easier to interpret than Kappa. Stehman (1997) reviewed a variety of summary statistics and concluded that overall map accuracy and user and producer accuracies have direct probabilistic interpretations for a given map, whereas other summary statistics must be used with caution. The error matrix itself is recognized as the most important accuracy assessment result when accompanied by descriptions of classification protocols, accuracy assessment design, source of reference data, and confidence in reference sample labels (Stehman and Czaplewski, 1998; Congalton and Green, 1999; Foody, 2002).

9.3 METHODS

Four LC maps for the upper San Pedro River Watershed (Plate 9.1) were generated using 1973, 1986, and 1992 North American Landscape Characterization (NALC) project MSS data (Lunetta et al., 1993) and the 1997 TM data. All images were coregistered and georeferenced to a 60- × 60-m Universal Transverse Mercator (UTM) ground coordinate grid with a nominal geometric precision of 1 to 1.5 pixels (60 to 90 m).

9.3.1 Image Classification

The same LC classes ($n = 10$) were used to develop all four maps (Table 9.1). Vegetation cover classes represented very broad biome-level categories of biological organization, similar to the ecological formation levels as described in the classification system for biotic communities of North America (Brown et al., 1979). The classes included forest, oak woodland, mesquite woodland, grassland, desertscrub, riparian, agriculture, urban, water, and barren and were selected after direct consultation with the major land managers and stakeholder groups within the San Pedro watershed in Arizona and Mexico (Kepner et al., 2000). Most of the watershed was covered by grassland, desert scrub, and mesquite and oak woodland (Table 9.2).

The classification process for each data set began with an unsupervised classification using the green, red, and near-infrared spectral bands to produce a map with 60 spectrally distinct classes. The choice of 60 classes was based on previous experience with NALC data that usually gave a satisfactory trade-off between the total number of classes and the number of mixed classes. In this context, it proved helpful to define a set of 21 intermediate classes, which were easier to relate to the spectral information. For example, the barren class contained bare rock, chalk deposits, mines, tailing ponds, etc., that had unique spectral signatures. Each class was then displayed over the false-color image and assigned to one of the LC categories or to a mixed class.

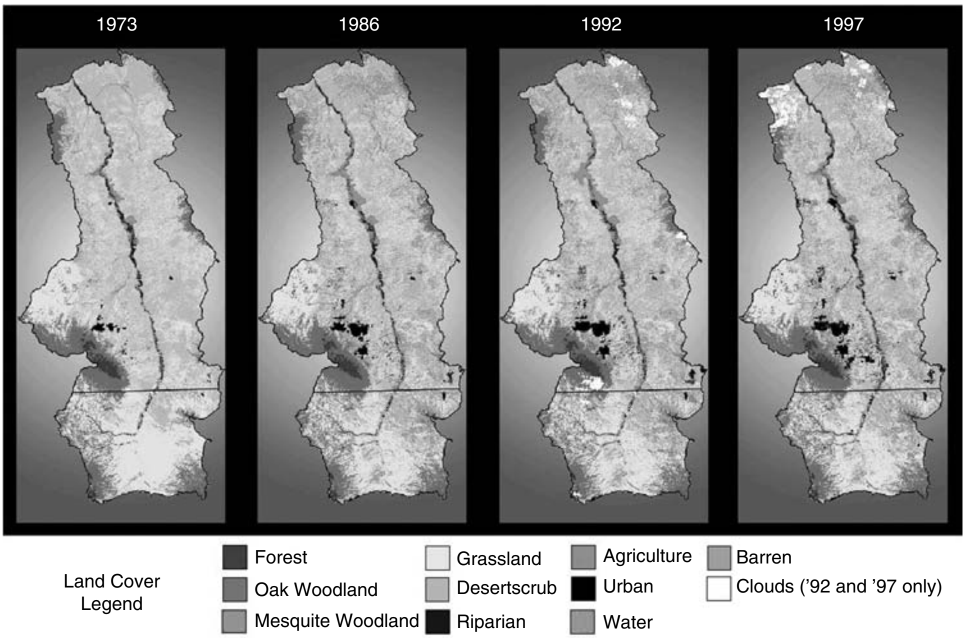

Plate 9.1 (See color insert following page 114.) 1973, 1986, 1992 and 1997 land-cover maps of the upper San Pedro River watershed with key to classes.

Interactive manipulation of spectral signatures for each class permitted many of the mixed classes to be resolved. The remaining mixed classes were separated into different categories using a variety of ancillary information sources, such as topographic maps produced by the Mexican National Institute of Statistics, Geography and Information (INEGI) (1:50,000 scale) and the U.S. Geological Survey (1:24,000 scale). The ancillary information used depended on the image being analyzed; for example, classification of the 1992 image relied heavily on field visits to establish ground control. Five 3-day site visits were conducted from September 1997 to June 1998 to enable analysts to collect specific LC data with the aid of GPS equipment.

9.3.2 Sampling Design

Because available reference data only partially covered the study area, pixels within each map were not equally likely to be selected for sampling; thus, a trade-off between practical constraints and statistical rigor was necessary (Congalton and Green, 1999). Sample points were selected using a stratified random sampling design, stratified by LC area for each of the four accuracy assessments. Reference data covering the Mexican portion of the study area were not available. The number of sample points was calculated using the following equation based on binomial probability theory (Fitzpatrick-Lins, 1981):

$$N = \frac{Z^2 pq}{E^2}$$

where N = number of samples, p = expected or calculated accuracy (%), $q = 100 - p$, E = allowable error, and Z = standard normal deviate for the 95% two-tail confidence level = 1.96.

For the lowest expected map accuracy of 60% with an allowable error of 5%, 369 sample points were required. Under area-stratified sampling, rare classes of small total area (i.e., water and barren) would not be sampled sufficiently to detect classification errors, so the minimum sample size was

Table 9.1 Land-Cover Class Descriptions for the Upper San Pedro Watershed

Forest	Vegetative communities comprised principally of trees potentially over 10 m in height and frequently characterized by closed or multilayered canopies. Species in this category are evergreen (with the exception of aspen), largely coniferous (e.g., ponderosa pine, pinyon pine), and restricted to the upper elevations of mountains that arise off the desert floor.
Oak Woodland	Vegetative communities dominated by evergreen trees (*Quercus* spp.) with a mean height usually between 6 and 15 m. Tree canopy is usually open or interrupted and singularly layered. This cover type often grades into forests at its upper boundary and into semiarid grassland below.
Mesquite Woodland	Vegetative communities dominated by leguminous trees whose crowns cover 15% or more of the ground, often resulting in dense thickets. Historically maintained maximum development on alluvium of old dissected flood plains; now present without proximity to major watercourses. Winter deciduous and generally found at elevations below 1200 m.
Grassland	Vegetative communities dominated by perennial and annual grasses with occasional herbaceous species present. Generally grass height is under 1 m and they occur at elevations between 1100 and 1700 m, sometimes as high as 1900 m. This is a landscape largely dominated by perennial bunch grasses separated by intervening bare ground or low-growing sod grasses and annual grasses with a less-interrupted canopy. Semiarid grasslands are mostly positioned in elevation between evergreen woodland above and desertscrub below.
Desertscrub	Vegetative communities comprised of short shrubs with sparse foliage and small cacti that occur between 700 and 1500 m in elevation. Within the San Pedro river basin this community is often dominated by one of at least three species (i.e., creosotebush, tarbush, and whitethorn acacia). Significant areas of barren ground devoid of perennial vegetation often separate individual plants. Many desertscrub species are drought deciduous.
Riparian	Vegetative communities adjacent to perennial and intermittent stream reaches. Trees can potentially exceed an overstory height of 10 m and are frequently characterized by closed or multilayered canopies depending on regeneration. Species within the San Pedro basin are largely dominated by two species: cottonwood and Goodding willow. Riparian species are largely winter deciduous.
Agriculture	Crops actively cultivated and irrigated. In the San Pedro River basin these are primarily found along the upper terraces of the riparian corridor and are dominated by hay and alfalfa. They are minimally represented in overall extent (less than 3%) within the basin and are irrigated by ground and pivot-sprinkler systems.
Urban (Low and High Density)	This is a land-use dominated by small ejidos (farming villages or communes), retirement homes, or residential neighborhoods (Sierra Vista). Heavy industry is represented by a single open-pit copper mining district near the headwaters of the San Pedro River near Cananea, Sonora (Mexico).
Water	Sparse free-standing water is available in the watershed. This category would be mostly represented by perennial reaches of the San Pedro and Babocomari rivers with some attached pools or repressos (earthen reservoirs), tailings ponds near Cananea, ponds near recreational sites such as parks and golf courses, and sewage treatment ponds east of the city of Sierra Vista, Arizona.
Barren	A cover class represented by large rock outcropping or active and abandoned mines (including tailings) that are largely absent of above-ground vegetation.

set to 20 where available (van Genderen and Lock, 1977). Work by Congalton (1991) and Congalton and Green (1999) suggests that sample sizes derived from multinomial theory are appropriate for comparing class accuracies, with a minimum sample size of 50 per class; however, this goal was not attainable for rare classes in this study.

After evaluation of selected sample points in each reference data set, an error matrix was constructed, comparing map class labels to reference data labels for each LC classification. Overall map accuracy and class-specific user and producer accuracies were calculated for each class. A Khat (Cohen's Kappa) and Tau (Ma and Redmond, 1995) were computed for the four error matrices, followed by a significant difference test (Z-statistic) based on Khat values (Congalton and Green, 1999).

9.3.3 Historical Aerial Photography

Reference data for the 1973 and 1986 LC maps were developed using aerial photography stereo pairs covering the Arizona portion of the study area (1:40,000 scale). A team, including a photo interpreter, an image processing specialist, a GIS specialist, and a statistician, conducted accuracy assessments. A preliminary study was conducted, using data collected during a field trip to the study area, to evaluate the effectiveness and accuracy of using aerial photographs to discriminate grassland, desertscrub, and mesquite woodland classes. These classes were particularly difficult to distinguish on the aerial photographs.

9.3.3.1 Image Collection, Preparation, and Site Selection

Landsat MSS data registration and other data integrity issues were reviewed for the 1973 and 1986 maps. These efforts included checking projection parameters and visual alignment using GIS data layers (i.e., roads, streams, digital raster graphics, and digital elevation models). Random sample points were generated using DOQQs acquired in 1992 (1:25,000 scale), and individual sample points were located on the aerial photographs using the DOQQs for accurate placement. A 180- × 180-m interpretation grid was generated and overlaid onto the LC maps.

Two mutually exclusive sets of sample points were generated for both 1973 and 1986 maps. The second set of sample points served as a pool of substitute points when no aerial photographs were available for a sample point in the first set. Whenever possible, pixels selected as sample sites represented the center of a 3 × 3 pixel window representing a homogeneous cover type. For rare classes (e.g., water), pixel sample points were chosen with at least six pixels in the window belonging to the same class. A total of 813 reference samples were used to assess the 1973 (n = 429) and 1986 (n = 384) maps. Multiple dates of aerial photographs were used in assessment: June 1971 and April 1972 (1973 map) and June 1983, June 1984, and September 1984 (1986 map).

9.3.3.2 Photograph Interpretation and Assessment

Photointerpreter training included using a subset of the generated sample points identified during visits to the San Pedro watershed locations as interpretation keys. To avoid bias, photointerpreters did not know what classifications had been assigned to sample points on the digital LC maps. To locate the randomly chosen sample sites on the aerial photographs, the site locations were first displayed on the DOQQ. Interpreters could then visually transfer the location of each site to the appropriate photograph by matching identical spatial data such as roads, vegetation patterns, rock outcrops, or other suitable features visible on the DOQQ and on the photograph. Each transferred sample point was examined on stereoscopic photographs and identified using the definitions shown in Table 9.1. LC categories for each sample point were recorded on a spreadsheet. A comment column on the spreadsheet allowed the interpreter to enter any notes about the certainty or ambiguity of the classification. The senior photointerpreter checked the accuracy of 10% of the sample point locations and 15% of the spreadsheet entries to ensure completeness and consistency. All LC class interpretations noted by a photointerpreter as "difficult" were classified by consensus opinion of all the interpreters.

9.3.4 Digital Orthophoto Quadrangles

Approximately 60 panchromatic DOQQs acquired in 1992 for the U.S. portion of the study area were available as reference data to evaluate the 1992 results. To obtain a precise geographic matching between the DOQQs and the satellite-derived map, the 1992 source MSS image data were geometrically registered to an orthorectified 1997 TM scene, and the resulting transformation parameters were applied to the 1992 thematic map.

9.3.4.1 Interpreter Calibration

To effectively visualize conditions represented by the LC class descriptions (Table 9.1), University of Arizona and IMADES team members participated in a field visit to numerous sites in the San Pedro watershed study area, including areas that were intermediate between classes. The analyst performing the 1992 assessment also reviewed high-resolution color airborne video data for comparison with the appearance of LC classes in the DOQQs. The video data were acquired over the watershed in 1995 and vegetation in selected frames at 1:200 scale was identified to species or species groups (Drake, 2000). Image "chips" were extracted from the DOQQs as an aid to LC class recognition in the reference data (Maingi et al., 2002).

9.3.4.2 Sample Point Selection

Generation of sample points from LC maps relied on a window majority rule. A window kernel of 3 × 3 pixels was moved across each cover class and resulted in selection of a sample point if a majority of six of the nine pixels belonged to the same class. This ensured that points were extracted from areas of relatively homogenous LC. A 180- × 180-m DOQQ sample size was used to match the 3 × 3 pixel map window and a map class was assigned and recorded for the DOQQ sample. A total of 457 points were sampled to assess the 1992 map.

9.3.5 Airborne Videography

Accuracy assessment of the 1997 LC map was performed using airborne color video data encoded with GPS time and latitude and longitude coordinates. The video data were acquired on May 2 through May 5, 1997, and were therefore nearly coincident with the June Landsat TM scene. There were 11 h of continuously recorded videography of the San Pedro Watershed for the area north of the U.S.–Mexico border, acquired at a flying height of 600 m above ground level. The nadir-looking video camera used a motorized 15× zoom lens that was computer controlled to cycle every 12 sec during acquisition, with a full-zoom view held for 3 sec. The swath width at wide angle was about 750 m and was approximately 50 m at full zoom. At full zoom, the ground pixel size was about 7.0 cm and the frame was approximately 1:200-scale when displayed on a 13-inch monitor. Although the nominal accuracy of the encoded GPS coordinates was only 100 m, ground sampling revealed that average positional accuracy was closer to 40 m (McClaran et al., 1999; Drake, 2000). The video footage was acquired by flying north–south transects spaced 5 km apart and the total flight coverage encompassed a distance of nearly 2000 km.

9.3.5.1 Video and GIS Data Preparation

The encoded GPS time and geographic coordinate data were extracted from the video into a spreadsheet for each flight line. Coordinate data from the spreadsheets were used to create GIS point coverages of frames from each flight line. Individual frames of the video data were identified during viewing by a time display showing hours, minutes, and seconds, in addition to a counter that numbered the 30 frames recorded per second. The time display information was included as an attribute to the GIS point coverages, which were inspected for erroneous coordinate or time data indicated by points that fell off the flight lines or were out of time sequence; such points were deleted.

9.3.5.2 Video Sample Point Selection

To minimize the likelihood of video sample points falling on boundaries between cover classes, selection of random sample points along the video flight lines was restricted to relatively homogeneous areas within classes. This was accomplished by applying a 3 × 3 diversity or variety filter to the 1997 map, which replaced the center pixel in a moving window by the number of different data file values (cover classes) present in the window. Pixels assigned the value of one therefore

Table 9.2 Upper San Pedro Watershed Land-Cover Classes: Absolute and Relative Areas; Representative Values from 1997 Land-Cover Classification

Land-Cover Class	Area (ha)	Proportion of Total Area (%)
Grassland	263,475	36
Desertscrub	229,571	31
Woodland Mesquite	101,559	14
Woodland Oak	90,540	12
Urban	16,562	2
Agriculture	14,530	2
Riparian	9,217	1
Forest	7,193	1
Barren	6,814	1
Water	417	<0.1
Total	739,878	100

represented centers of 180- × 180-m homogeneous areas on the map. Background, clouds, and cloud-shadowed pixels were excluded to prevent the selection of pixels that fell at the edge of the map, within openings in clouds, or in cloud-shadowed areas where the adjacent cover classes were not known.

Video flight line coverages were overlaid on the map of homogeneous cover, and a subset of frames falling on homogeneous areas (n = 4,567) was drawn from all study area frames (n = 18,104). The map class under each subset frame was added as an attribute to the "candidate frames" GIS point coverage for stratification purposes.

9.3.5.3 Random Frame Selection and Evaluation

Video sample points were drawn randomly from the homogeneous subset, stratified by map class area, and were distributed throughout the Arizona portion of the study area. The water class was excluded from analysis for lack of adequate reference data (n = 6) and was not presented in the final error matrix. A surplus of approximately 15% over the calculated minimum number of frames needed for each cover class was selected. The videography interpreter was provided with spreadsheet records containing the videotape library identifier, latitude, and longitude for each sample frame, along with GPS time for frame location on the tape. A cover class was assigned to each sample point and recorded in the spreadsheet.

Although the accuracy of video frame interpretation was not assessed in this study, it is expected to be very high. Drake (1996) reported that LC identification of similar airborne videography at the more detailed biotic community level averaged 80% accuracy after only 3 h of interpreter training. The interpreter for this study had substantial prior experience in both video frame interpretation and ground sampling for videography accuracy assessment in this region.

9.4 RESULTS

9.4.1 Aerial Photography Method

Results of accuracy assessment are presented in Table 9.3 (1973) and Table 9.4 (1986). Overall map accuracies were similar at 70% for 1973 and 68% for 1986. Khat and Tau statistics were also similar at 0.62 and 0.59 (Khat), and 0.66 and 0.65 (Tau) for 1973 and 1986, respectively. The user's and producer's accuracies were similar to overall accuracy for all except the mesquite woodland

Table 9.3 Error Matrix Comparing Aerial Photo Interpretation and 1973 Digital Land-Cover Classification, with Producer's and User's Accuracy by Class

1973 Land-Cover Class	Reference (Aerial Photo Interpretation Class)										
	1	2	3	4	5	6	7	8	9	10	Grand total
1	19	1	0	0	0	0	0	0	0	0	20
2	1	33	0	3	0	0	0	0	0	0	37
3	0	1	16	1	0	2	0	0	0	0	20
4	0	0	13	92	21	0	0	0	1	1	128
5	0	0	14	11	96	0	0	0	0	1	122
6	0	0	3	0	2	15	0	0	0	0	20
7	0	0	3	0	7	1	10	0	0	1	22
8	0	0	0	2	5	0	0	13	0	0	20
9	0	0	4	3	6	0	1	0	3	3	20
10	0	0	0	2	15	0	0	0	1	2	20
Grand Total	20	35	53	114	152	18	11	13	5	8	429

Land-Cover Class	1973 Map Total	Photointerpreter Total	Number Correct	Producer's Accuracy (%)	User's Accuracy (%)
1. Forest	20	20	19	95	95
2. Woodland Oak	37	35	33	94	89
3. Woodland Mesquite	20	53	16	30	80
4. Grassland	128	114	92	81	72
5. Desertscrub	122	152	96	63	79
6. Riparian Forest	20	18	15	83	75
7. Agriculture	22	11	10	91	45
8. Urban	20	13	13	100	65
9. Water	20	5	3	60	15
10. Barren	20	8	2	25	10
Total	429	429	299		

Note: Overall accuracy = 70%; Tau = 0.66; Cohen's Kappa (Khat) = 0.62; standard error = 0.027.

and barren classes, which showed substantially less than average accuracies in both years. The water class in 1973 had very low accuracies of 25% (producer's) and 10% (user's) and could not be assessed for 1986.

9.4.2 Digital Orthophoto Quadrangle Method

Accuracy assessment results are summarized in Table 9.5. Overall accuracy was about 75%, with Khat of 0.70 and Tau of 0.72. The producer's accuracy was 100% for four classes (forest, urban, water, and barren), indicating that all pixels examined in the DOQQs for these classes were correctly labeled in the 1992 map. The user's accuracy was also high for forest and water classes but was substantially less for urban and barren classes at 44 and 55%, respectively. Accuracies of mesquite woodland and grassland classes were lower than those for other classes.

9.4.3 Airborne Videography Method

Overall 1997 map accuracy was 72%, with Khat of 65% and Tau of 68% (Table 9.6). A detailed examination of results by cover class shows substantial variability in classification accuracy, with producer's accuracies ranging from 54 to 100% and user's accuracies from 13 to 100%. For most classes the two measures were roughly comparable and fell within the range of 60 to 90%. Exceptions were the mesquite woodland class with accuracies around 50% and agriculture and barren classes with relatively high producer's accuracies (71 to 100%) but lower user's accuracies (13 to 21%).

Table 9.4 Error Matrix Comparing Aerial Photo Interpretation and 1986 Land-Cover Classification, with Producer's and User's Accuracy by Class

1986 Land-Cover Classes	Reference (Aerial Photo Interpretation Class)									
	1	2	3	4	5	6	7	8	10	Grand Total
1	19	1	0	0	0	0	0	0	0	20
2	3	35	0	1	0	0	0	0	0	39
3	0	0	17	3	19	0	1	0	2	42
4	0	0	12	77	12	0	1	0	2	104
5	0	0	8	13	74	0	0	0	0	95
6	0	0	0	1	1	19	2	0	0	23
7	0	0	1	4	3	2	9	1	0	20
8	0	0	0	5	3	0	0	13	0	21
10	0	0	3	10	7	0	0	0	0	20
Grand Total	22	36	41	114	119	21	13	14	4	384

Land-Cover Class	1986 Map Total	Photointerpreter Total	Number Correct	Producer's Accuracy (%)	User's Accuracy (%)
1. Forest	20	22	19	86	95
2. Woodland Oak	39	36	35	97	90
3. Woodland Mesquite	42	41	17	42	41
4. Grassland	104	114	77	68	74
5. Desertscrub	95	119	74	62	78
6. Riparian Forest	23	21	19	91	83
7. Agriculture	20	13	9	69	45
8. Urban	21	14	13	93	62
10. Barren	20	4	0	0	0
Total	384	384	263		

Note: Overall accuracy = 68%; Tau = 0.65; Cohen's Kappa (Khat) = 0.61; standard error = 0.029.

9.5 DISCUSSION

9.5.1 Map Accuracies

Statistics describing map accuracy were very similar among the four dates tested regardless of differences in assessment methods and reference data. Overall map accuracies ranged from 67 to 75% and Tau values from 0.65 to 0.72. There were no statistically significant differences among Khat values (0.61 to 0.70) for all possible date comparisons.

One aspect of sampling that differed among the assessments was the application of homogeneity standards to the context of map sample points. Selection was made from the center of uniform 3 × 3 pixel windows for the1973 and 1986 assessments, with an exception for rare cover classes requiring only a majority of five or more pixels to match the center pixel. All sample points were selected from uniform 3 × 3 windows in the 1997 assessment. In contrast, for the 1992 assessment, a map class label was assigned as the majority of six or more pixels within a 3 × 3 window centered on the sample point. Although a positive bias may have been introduced by sampling only in homogeneous areas (Hammond and Verbyla, 1996), this effect was not apparent in results presented here.

9.5.2 Class Confusion

For all dates evaluated the producer's and user's accuracies tended to be similar to the overall classification accuracies and ranged between 61 and 100%. Generally low classification accuracies were expected in a spatially heterogeneous setting such as the San Pedro watershed, where cover types were distributed in a patchy fashion across the landscape due to climatic and edaphic effects

Table 9.5 Results of DOQQ-Based Accuracy Assessment of 1992 Land-Cover Classification: Error Matrix and Producer's and User's Accuracy by Class

1992 Land-Cover Classes	Reference (Digital Orthophoto Quads)										
	1	2	3	4	5	6	7	8	9	10	Grand Total
1	22	2	0	0	0	0	0	0	0	0	24
2	0	44	0	3	1	0	0	0	0	0	48
3	0	2	40	9	10	1	0	0	0	0	62
4	0	6	12	68	17	0	0	0	0	0	103
5	0	1	8	11	89	0	0	0	0	0	109
6	0	0	0	0	0	20	3	0	0	0	23
7	0	0	1	0	0	4	18	0	0	0	23
8	0	0	2	1	10	0	1	11	0	0	25
9	0	0	1	0	0	0	0	0	19	0	20
10	0	0	0	7	2	0	0	0	0	11	20
Grand Total	22	55	64	99	129	25	22	11	19	11	457

Land-Cover Class	1992 Map Total	DOQQ Total	Number Correct	Producer's Accuracy (%)	User's Accuracy (%)
1. Forest	24	22	22	100	92
2. Woodland Oak	48	55	44	80	92
3. Woodland Mesquite	62	64	40	63	65
4. Grassland	103	99	68	69	66
5. Desertscrub	109	129	89	69	82
6. Riparian Forest	23	25	20	80	87
7. Agriculture	23	22	18	82	78
8. Urban	25	11	11	100	44
9. Water	20	19	19	100	95
10. Barren	20	11	11	100	55
Total	457	457	342		

Note: Overall accuracy = 75%; Tau = 0.72; Cohen's Kappa (Khat) = 0.70; standard error = 0.025.

and land-use practices. Classes mapped with lower than average accuracy included the small-area agriculture, urban, water, and barren classes and the widespread mesquite woodland class. Factors likely to have contributed to class confusions included: (1) LC changes between the dates of image and reference data (especially for the 1973 and 1986 maps), (2) high spatial variability within classes (including areas dominated by soil background reflectance), (3) variable interpretations of class definitions by independent assessment teams, and (4) errors in reference data interpretation. Geometric misregistration did not appear to be a factor in the results presented here.

The agriculture class had higher producer than user accuracies for all dates and was most frequently confused with riparian, desertscrub, and mesquite woodland classes. The spatial distribution of agricultural areas in the watershed essentially outlined the riparian corridors, contributing to mixed pixel spectral response and classification confusion. There may have been difficulty in distinguishing fallow and abandoned agricultural fields from adjacent desertscrub and mesquite woodland, since the spectral response of these cover types was generally dominated by soil background.

The urban class included low-density settlement on both sides of the border. Low-density development was difficult to distinguish from surrounding cover types even at the DOQQ scale, suggesting the possibility of error in both maps and reference data. The accelerating pace of development in the watershed, particularly in Arizona, may have contributed to cover changes occurring between the dates of imagery and reference data.

The water class had the smallest area and was likely to have changed between the dates of images and reference data, due to the ephemeral nature of most surface water in this semiarid environment. For example, the 1973 NALC scene was acquired after a high-rainfall, El Niño–Southern Oscillation (ENSO) event during the winter of 1972–73 and portrayed wetter conditions than

Table 9.6 Results of Video-Based Accuracy Assessment of the 1997 Land-Cover Classification: Error Matrix and User's and Producer's Accuracy by Class

1997 Land-Cover Classes	Reference (Video Frame Data)									
	1	2	3	4	5	6	7	8	10	Grand Total
1	20	4	0	0	0	0	0	0	0	24
2	2	50	0	3	0	0	0	0	0	55
3	0	1	27	13	12	2	0	1	0	56
4	0	8	16	113	21	0	0	1	0	159
5	0	4	4	12	115	0	0	2	0	137
6	0	0	0	0	0	21	2	1	0	24
7	0	0	1	0	15	2	5	1	0	24
8	0	0	0	0	0	0	0	24	0	24
10	0	0	2	0	19	0	0	0	3	24
Grand Total	22	67	50	141	182	25	7	30	3	527

Land-Cover Class	1997 Map Total	Video Total	Number Correct	Producer's Accuracy (%)	User's Accuracy (%)
1. Forest	24	22	20	91	83
2. Woodland Oak	55	67	50	75	91
3. Woodland Mesquite	56	50	27	54	48
4. Grassland	159	141	113	80	71
5. Desertscrub	137	182	115	63	84
6. Riparian	24	25	21	84	88
7. Agriculture	24	7	5	71	21
8. Urban	24	30	24	80	100
9. Water	N/A	N/A	N/A	N/A	N/A
10. Barren	24	3	3	100	13
Total	527	527	378		

Note: Overall accuracy = 72%; Tau = 0.68; Cohen's Kappa (Khat) = 0.65; standard error = 0.024.

reference aerial photography acquired in 1971 and 1972 (Easterling et al., 1996; NOAA, 2001). The water class was not evaluated in 1986 and 1997 assessments due to insufficient representation in reference data.

The barren class was mapped with poor accuracy overall, including 0% correct in 1986. This class was most often confused with mesquite woodland, grassland, and desertscrub. These classes generally have sparse vegetation cover, with many image pixels dominated by soil or rock spectral responses, and were difficult to distinguish from truly barren areas at the MSS 60-m pixel size. A total of 38% of samples interpreted as barren on reference aerial photography from 1971 and 1972 were mapped as water in 1973; this was probably due to the interannual variations in precipitation mentioned above.

The mesquite woodland class may be interpreted as an indicator of landscape change in the San Pedro Watershed (Kepner et al., 2000, 2002). Conversion of many grassland areas to shrub dominance during the last 120 years is well documented for this region (Bahre, 1991, 1995; Wilson et al., 2001), and these change detection results were of potential interest to many researchers. However, both user and producer accuracies of all four dates were generally low for mesquite woodland (30 and 80%, respectively, for 1973 and 40 to 65% for other years). Class confusions included all but the forest class, with especially large errors in the grassland and desertscrub classes. This result may substantially reflect both the spatially and temporally transitional nature of the class and differences in interpretation among the groups performing image classification and accuracy assessment. Additionally, it was likely that neither the spectral nor the spatial resolution of MSS imagery was adequate to distinguish the mesquite woodland class in a heterogeneous semiarid environment, where most pixels are mixtures of green and woody vegetation, standing litter, and soils of varying brightness (Asner et al., 2000).

9.5.3 Future Research

Assessment of future LC classifications for the upper San Pedro area should incorporate some measure of the reference data variability, perhaps also allowing a secondary class label (Zhu et al., 2000; Yang et al., 2001). This may help to clarify the results for some cover classes. For example, the low accuracies and class confusions associated with the mesquite woodland class may have been due, in large part, to its gradational nature. If the interpreter would have been able to quantify the confidence associated with reference point interpretations, there would not have been a need to select sample points from homogeneous map areas, thus reducing the possibility of a positive accuracy bias (Foody, 2002). Another useful tool for future San Pedro LC work is the map of all sample points used in the accuracy assessment. Each point was attributed with geographic coordinates and both map and reference data labels (Skirvin et al., 2000). These data could be applied to generate a geographic representation of the continuous spatial distribution of LC errors (Kyriakidis and Dungan, 2001) to highlight especially difficult areas that should be field checked or otherwise handled in the future.

9.6 CONCLUSIONS

The results discussed in this chapter indicate that historical aerial photography, DOQQ data, and high-resolution airborne video data can be used successfully to perform classification accuracy assessment on LC maps derived from historical satellite data. Archived aerial photographs may be the only reference data available for retrospective analysis before 1992. However, their resolution (1:40,000 scale for NAPP data) often makes this task difficult. Successful use of DOQQ data requires precise geometric registration of the LC map to allow the overlay of orthorectified DOQQs. The use of georeferenced high-resolution airborne videography as a proxy for actual ground sampling in accuracy assessment provided the best method for current reference data development in the San Pedro watershed. The advantages include: (1) cost-effective collection of a statistically meaningful number of sample points, (2) effective control of coordinate locational error, and (3) variable-scale videography that permits the identification of specific plant species or communities of interest. Additionally, the videography provides a clear depiction of cultural features and land-use activities. The main limitation of this method is that data are collected along predetermined flight paths, thus constraining the sampling frame design.

9.7 SUMMARY

Because the rapidly growing archives of satellite remote sensing imagery now span decades, there is increasing interest in the study of long-term regional LC change across multiple image dates. However, temporally coincident ground-sampled data may not be available to perform an independent accuracy assessment of the image-derived LC map products. This study explored the feasibility of utilizing historical aerial photography, DOQQs, and high-resolution airborne color video data to assess the accuracy of satellite-derived LC maps for the upper San Pedro River watershed in southeastern Arizona and northeastern Sonora, Mexico. Satellite image data included Landsat Multi-Spectral Scanner (MSS) and Landsat Thematic Mapper (TM) data acquired over an approximately 25-year period. Four LC classifications were performed using three dates of MSS imagery (1973, 1986, and 1992) and one TM image (1997). The TM imagery was aggraded from 30 to 60 m to match the coarser MSS pixel size.

A stratified random sampling design was incorporated with samples apportioned by LC area, using a minimum sample size of $n = 20$ for rare classes. Results indicated similar map accuracies were obtained using the three alternative methods. Aerial photography provided reference data to assess the 1973 and 1986 LC maps with overall classification accuracies of 70% (1973) and 67% (1986). Assignments of class labels to sample points on 1992 reference DOQQs were verified by comparison with higher-resolution airborne video data, with overall 1992 map classification accuracy of 75%. Accuracy assessment of the 1997 products used contemporaneous airborne color video data and resulted in an overall map accuracy of 72%. There was no evidence of positive bias in accuracy resulting from use of homogeneous vs. heterogeneous pixel contexts in sampling the LC maps.

The use of historical aerial photography, high-resolution DOQQs, and airborne videography as a proxy for actual ground sampling for satellite image classification accuracy has merit. Selection of a reference data set for this study depended on the date of image acquisition. For example, prior to 1992, historical aerial photographs were the only data available. DOQQs covered the period since initiation of the high-resolution NAPP in 1992, and high-resolution airborne videography provided a cost-effective means of acquiring many reference sample points near the time of image acquisition. Problems that were difficult to avoid included inadequate sampling of rare classes and reconciling cover changes between acquisition dates of aerial photography or DOQQs and satellite image data. Other issues, including the need for consistent geometric rectification and criteria for mutually exclusive and reproducible LC class descriptions, need special attention when satellite image classification and subsequent LC map accuracy assessment are performed by different teams.

ACKNOWLEDGMENTS

The U.S. Environmental Protection Agency, Office of Research and Development provided funding for this work. The authors wish to thank participants from U.S. EPA, Lockheed Martin Environmental Services, Instituto del Medio Ambiente y el Desarrollo Sustentable del Estado de Sonora (IMADES), and the Arizona Remote Sensing Center at the University of Arizona for their assistance.

REFERENCES

Asner, G.P., C.A. Wessman, C.A. Bateson, and J.L. Privette, Impact of tissue, canopy, and landscape factors on the hyperspectral reflectance variability of arid ecosystems, *Remote Sens. Environ.*, 74, 69–84, 2000.

Bahre, C.J., *A Legacy of Change*, The University of Arizona Press, Tucson, 1991.

Bahre, C.J., Human impacts on the grasslands of southeastern Arizona, in *The Desert Grassland*, McClaran, M.P. and T.R. Van Devender, Eds., The University of Arizona Press, Tucson, 1995.

Brown, D.E., C.H. Lowe, and C.P. Pase, A digitized classification system for the biotic communities of North America, with community (series) and association examples for the Southwest, *J. Arizona-Nevada Acad. Sci.*, 14 (Suppl. 1), 1–16, 1979.

Congalton, R., A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sens. Environ.,* 37, 35–46, 1991.

Congalton, R.G. and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, CRC Press, Boca Raton, FL, 1999.

Congalton, R.G., R.G. Oderwald, and R.A. Mead, Assessing Landsat classification accuracy using discrete multivariate statistical techniques, *Photogram. Eng. Remote Sens.,* 49, 1671–1678, 1983.

Drake, S.E., Climate-Correlative Modeling of Phytogeography at the Watershed Scale, Ph.D. dissertation, University of Arizona, Tucson, 2000.

Drake, S.E., Visual interpretation of vegetation classes from airborne videography: an evaluation of observer proficiency with minimal training, *Photogram. Eng. Remote Sens.*, 62, 969–978, 1996.

Easterling, D.R., T.R. Karl, E.H. Mason, P.Y. Hughes, D.P. Bowman, R.C. Daniels, and T.A. Boden, United States Historical Climatology Network (U.S. HCN) Monthly Temperature and Precipitation Data, Revision 3, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN, 1996.

Fitzpatrick-Lins, K., Comparison of sampling procedures and data analysis for a land-use and land-cover map. *Photogram. Eng. Remote Sens.*, 47, 343–351, 1981.

Foody, G.M., On the compensation for chance agreement in image classification accuracy assessment, *Photogram. Eng. Remote Sens.*, 58, 1459–1460, 1992.

Foody, G.M., Status of land cover accuracy assessment, *Remote Sens. Environ.*, 80, 185–201, 2002.

Hammond, T.O. and D.L. Verbyla, Optimistic bias in classification accuracy assessment, *Int. J. Remote Sens.*, 17, 1261–1266, 1996.

Jones, K.B., L.R. Williams, A.M. Pitchford, E.T. Slonecker, J.D. Wickham, R.V. O'Neill, D. Garofalo, and W.G. Kepner, A National Assessment of Landscape Change and Impacts to Aquatic Resources: A 10-Year Strategic Plan for the Landscape Sciences Program, EPA/600/R-00/001, U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC, 2000.

Kepner, W.G., C.J. Watts, and C.M. Edmonds, Remote Sensing and Geographic Information Systems for Decision Analysis in Public Resource Administration: A Case Study of 25 Years of Landscape Change in a Southwestern Watershed, EPA/600/R-02/039, U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC, 2002.

Kepner, W.G., C.J. Watts, C.M. Edmonds, J.K. Maingi, S.E. Marsh, and G. Luna, A landscape approach for detecting and evaluating change in a semi-arid environment. *Environ. Monit. Assess.*, 64, 179–195, 2000.

Kyriakidis, P.C. and J.L. Dungan, A geostatistical approach for mapping thematic classification accuracy and evaluating the impact of inaccurate spatial data on ecological model predictions, *Environ. Ecol. Stat.*, 8, 311–330, 2001.

Lunetta, R.L., J.G. Lyon, J.A. Sturdevant, J.L. Dwyer, C.D. Elvidge, L.K. Fenstermaker, D. Yuan, S.R. Hoffer, and R. Werrackoon, North American Landscape Characterization: Research Plan, EPA/600/R-93/135, U.S. Environmental Protection Agency, Las Vegas, NV, 1993.

Ma, Z. and R.L. Redmond, Tau coefficients for accuracy assessment of classification of remote sensing data, *Photogram. Eng. Remote Sens.*, 61, 435–439, 1995.

Maingi, J.K., S.E. Marsh, W.G. Kepner, and C.M. Edmonds, An Accuracy Assessment of 1992 Landsat-MSS Derived Land Cover for the Upper San Pedro Watershed (U.S./Mexico), EPA/600/R-02/040, U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC, 2002.

Marsh, S.E., J.L. Walsh, and C. Sobrevila, Evaluation of airborne video data for land-cover classification accuracy assessment in an isolated Brazilian forest, *Remote Sens. Environ.*, 48, 61–69, 1994.

McClaran, M., S.E. Marsh, D. Meko, S.M. Skirvin, and S.E. Drake. Evaluation of the Effects of Global Climate Change on the San Pedro Watershed: Final Report, Cooperative Agreement No. A950-A1-0012 between the University of Arizona and the U.S. Geological Survey, Biological Resource Division, 1999.

NOAA (National Oceanic and Atmospheric Administration), Climate Prediction Center: ENSO Impacts on the U.S.: Previous Events, Web page [accessed 22 October 2002], available at http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.html.

Rosenfield, G.H. and K. Fitzpatrick-Lins, A coefficient of agreement as a measure of thematic classification accuracy, *Photogram. Eng. Remote Sens.*, 52, 223–227, 1986.

Skirvin, S.M., S.E. Drake, J.K. Maingi, S.E. Marsh, and W.G. Kepner, An Accuracy Assessment of 1997 Landsat Thematic Mapper Derived Land Cover for the Upper San Pedro Watershed (U.S./Mexico), EPA/600/R-00/097, U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC, 2000.

Stehman, S.V., Selecting and interpreting measures of thematic classification accuracy, *Remote Sens. Environ.*, 62, 77–89, 1997.

Stehman, S.V. and R.L. Czaplewski, Design and analysis for thematic map accuracy assessment: fundamental principles, *Remote Sens. Environ.*, 64, 331–344, 1998.

USEPA (U.S. Environmental Protection Agency), Upper San Pedro River, Web page [accessed 17 October 2002]. Available at http://www.epa.gov/nerlesd1/land-sci/html2/sanpedro_home.html.

van Genderen, J.L. and B.F. Lock, Testing land use map accuracy, *Photogram. Eng. Remote Sens.*, 43, 1135–1137, 1977.

Wilson, T.B., R.H. Webb, and T.L. Thompson, Mechanisms of Range Expansion and Removal of Mesquite in Desert Grasslands of the Southwestern United States, General Technical Report RMRS-GTR-81, U.S. Forest Service, Rocky Mountain Research Station, 2001.

Yang, L., S.V. Stehman, J.H. Smith, and J.D. Wickham, Thematic accuracy of MRLC land cover for the eastern United States, *Remote Sens. Environ.*, 76, 418–422, 2001.

Zhu, Z., L. Yang, S.V. Stehman, and R.L. Czaplewski, Accuracy assessment for the U.S. Geological Survey regional land-cover mapping program: New York and New Jersey region, *Photogram. Eng. Remote Sens.*, 66, 1425–1435, 2000.

Remote Sensing and GIS Accuracy Assessment

Edited by
Ross S. Lunetta
John G. Lyon

- - -

CRC PRESS
Boca Raton London New York Washington, D.C

This is a work of the United States Government under the provisions of Title 17, Section 105 of the U.S. Code and, therefore, U.S. copyright protection is not available. Under U.S. law, no U.S. copyright may be assigned. U.S. Government works are in the public domain and may be used by members of the U.S. public without copyright restrictions.

Library of Congress Cataloging-in-Publication Data

Remote sensing and GIS accuracy assessment / edited by Ross S. Lunetta, John G. Lyon.
p. cm.
Includes bibliographical references (p.).
ISBN 1-56670-443-X
1. Remote sensing—Congresses. 2. Geographic information systems—Congresses. 3. Spatial analysis (Statistics)—Congresses. I. Lunetta, Ross S. II. Lyon, J. G. (John G.)
G70.39.R45 2004
621.36'78—dc22 2004045728

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage or retrieval system, without prior permission in writing from the publisher.

The consent of CRC Press LLC does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific permission must be obtained in writing from CRC Press LLC for such copying.

Direct all inquiries to CRC Press LLC, 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431.

Trademark Notice:
Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

Visit the CRC Press Web site at www.crcpress.com
International Standard Book Number 1-56670-443-X
Library of Congress Card Number 2004045728
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper