LANDSAT VEGETATION MAPPING OF THE SOUTHWEST AND CENTRAL IDAHO ECOGROUPS

FINAL REPORT
Contract #: 53-0261-6-25

Submitted to:

U.S. Department of Agriculture
Forest Service, Boise National Forest
1249 South Vinnell Way
Boise, ID
83709

By:

Roland L. Redmond
Troy P. Tady
Foster B. Fisher
Michele Thornton
J. Chris Winne

Wildlife Spatial Analysis Lab
Montana Cooperative Wildlife Research Unit
The University of Montana
Missoula, MT
59812

14 November 1997

ABSTRACT

We constructed a digital map of existing vegetation and land cover across nearly 8 million hectares in central Idaho (~19.8 million acres), based on the classification of 6 Landsat Thematic Mapper (TM) scenes. Information was stored in a raster database comprised of a series of ARC/INFO grids (one per TM scene). This database was built through a two-stage classification involving both unsupervised and supervised procedures. First, for each TM scene, an unsupervised classification of pixels was conducted. This pixel classification, based on Euclidean distance calculations, was designed to maintain patterns observed in a color composite of TM channels 4, 5, and 3. The resulting spectral classes were then regrouped and merged to a 0.4 ha minimum map unit (MMU; > 4 pixels) in areas where riparian vegetation was likely to occur, and to a 2 ha MMU (> 22 pixels) in upland settings. Next, a raster database was constructed in ARC/INFO: base regions (or raster polygons) were delineated, and attributes for each region were collected (including majority aspect values and mean values for TM channels 1-7, elevation, and slope). Adjacent TM scenes were then 'virtually' edge-matched so that the files could be seamlessly merged after classification. Meanwhile, ground-truth data were assembled by the Forest Service and sent to the University, where they underwent a series of logical and positional tests to verify their accuracy and their utility for supervised classification purposes. In all, information from 19,798 unique plots was compiled in a reference database and used for classification or accuracy assessment. Plots in each TM scene that passed all tests were separated into two data sets: 20% of the plots by cover type were set aside for use in accuracy analysis, and 80% were used as training data in a supervised classification which assigned cover type labels using the NEAREST MEMBER OF GROUP classifier. This process was repeated for size class, but for some scenes, the NEAREST MEAN classifier was used to assign labels. Training and test data sets were also built for canopy cover; however, labels were assigned using decision rules based on an examination of modified NDVI values for ground-truth plots. Decision rules were also used for manual modifications specific to each scene. In addition, some cover types were labeled through visual interpretation of imagery, such as urban areas, agricultural lands, and clouds. In all, 41 cover types were mapped for the project area; these included urban, agriculture (irrigated and dryland), three grassland types, five shrub types, aspen, 16 coniferous forest types, two classes of burns, five riparian types, plus water, rock, snow, barren, basalt, and clouds. Separate assessments of accuracy were conducted using fuzzy sets for land cover type, size class, and for canopy cover attributes.

ACKNOWLEDGEMENTS

This project involved a close collaboration between the U.S.D.A. Forest Service and The University of Montana. As such, many different individuals played active and vital roles. We begin with those from the Forest Service. Bill Rush (Boise National Forest) not only had the vision and determination to get the project underway, but he also provided the leadership and guidance to see it through to completion. Two different contracting officers, Ron Zimmerman, and Stan Bird, handled all contract matters promptly and courteously. The forest representatives for the project deserve special mention for their coordinating efforts on each forest: Susan Boudreau (Payette National Forest), Don Oman (Sawtooth National Forest), and Lynn Bennett and Cliff Keene (Salmon and Challis National Forest). A number of Forest Service personnel also deserve special mention for coordinating and carrying out project responsibilities on individual forests: Cameo Flood, Joe Frost, Diane McConnaughey, Lynn Morelan, Norm Ness, Linda Spillers, and Diana Wall (Boise National Forest); Ted Demetriades, Dick Foster, and Chance O'Brien (Payette National Forest); Rob Daley, Paul Demeule, Howard Hudak, Gary Kieser, Mike O'Farrel, Seth Phalen, and John Shelly (Sawtooth National Forest); Melissa Abbott, Doug Basford, David Booth, Carol Boyd, Sharon Bradley, Joe Butsick, Russ Camper, Jennie Cummings, Dave Fallis, Mike Foster, John Fowler, Ben Garechana, Tim Gionet, Cindy Haggas, Brad Harris, Ed Hayes, Bill Hickey, Breck Hudson, Vince Johnson, Richard Madril, Dave Morris, Cheri Ohl, Dave Reeder, Ken Rodgers, Dianne Schuldt, Kathy Seaberg, Don Smith, Steve Spencer, Gene Sundberg, Larry Thornton, Janet Valle, Doug Weaver, and Dick Wenger (Salmon and Challis National Forest).

Additional help also came from the Salmon Resource Area Bureau of Land Management office: Bob Berasi, Bill Diage, Bruce Easton, Glenn Elzinga, Julie Owen, and Russel Reibe.

Last but not least, we thank those who helped and supported us at The University of Montana: Ken Hubbard, Pat Bristol, and Sheila Hoffland in the Office of Research Administration, Sandi Russo in Business Services. Our friends, colleagues, and co-workers in the Wildlife Spatial Analysis Lab all contributed in many important ways to the successful completion of this project; Jim Schumacher deserves mention for creating report figures, as do Poody McLaughlin and Kathy Jurist for their efforts in contract administration and the editing of this report.

LANDSAT VEGETATION MAPPING OF THE SOUTHWEST AND CENTRAL IDAHO ECOGROUPS

— TABLE OF CONTENTS —

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

This report describes the assembly and contents of several digital databases produced to map existing vegetation and land cover in a standardized and consistent manner across nearly 8 million hectares (19.77 million acres) in central Idaho. This work was an extension of an ambitious, complex, and successful collaboration between the U.S. Forest Service and The University of Montana. As originally envisioned, these databases are intended to support landscape assessments of the Southwest and Central Idaho Ecogroup areas. They are ideally suited for analyses at the regional, subregional, and landscape levels, as well as for support of many resource management disciplines, including timber, grazing, wildlife and fisheries habitat, and recreation.

The report is organized as follows. General descriptions of the project area, methods, and results are presented, followed by a discussion of the strengths and limitations of the methods and resulting databases. Complete descriptions of the vegetation classification scheme and attributes are provided in Appendices A and B, respectively. More detailed results and comments for each individual TM scene are found in Appendix C.

Digital GIS data files for each of the six classified central Idaho TM scenes were delivered prior to this written report. In addition, the following digital data files were delivered:

1. An Arc Macro Language (AML) program (named CLIPPER) that allows users to copy a portion of the GIS database (including all attributes) for any area specified by a boundary file.

2. Files containing all ground-truth information used to classify each TM scene (i.e., all training data), most of which were supplied by the Forest Service, although in a different format. These data may be useful for any future revisions to the supervised classifications or to their accuracy assessments.

3. Files of the Ecological Limit rules used for TM scenes P40R29, P41R28, P41R29, and P41R30 (note: P40R30 and P42R29 did not use Ecological Limit rules).

PROJECT AREA

The project area covers approximately 8 million hectares (19.77 million acres) in central Idaho (Figure 1). It is bounded on the south and east by the Snake River Plain, and by the Snake River to the west. The northern boundary roughly follows the Montana/Idaho border through the Beaverhead Mountains from Bannack Pass to Nez Perce Pass, and then west across Idaho, parallel to and north of the Salmon River. Major river systems draining the area include the Boise, Payette, Weiser, Salmon, Lemhi, Pahsimeroi, Wood, Big Lost, and Little Lost Rivers. Elevations range from nearly 400 m (1312 ft) above sea level at the bottom of Hells Canyon to 3860 m (12,662 ft) at the summit of Borah Peak in the Lost River Range. The area is sparsely populated and predominantly forested. Much of the land is administered by the U.S.D.A. Forest Service through the Boise, Payette, Sawtooth, and Salmon/Challis National Forests. The U.S. Department of Interior also administers lands around the perimeter of the area through the Bureau of Land Management and National Park Service (Craters of the Moon National Monument). Other important land management authorities include the U.S. Department of Energy (Idaho National Engineering Lab) and the state of Idaho. Land management practices across the area vary from being relatively simple and straightforward in some remote areas of rock, snow, and ice, to being complex and sometimes controversial in and around places like Cascade, Ketchum, and the Sawtooth National Recreation Area.

METHODS:
CREATING A DATABASE OF EXISTING VEGETATION AND LAND COVER

To map existing vegetation and land cover across the project area, we employed a two stage classification process, unique to this lab, which integrates remote sensing and GIS technologies. In the first stage, land cover patterns are derived from a color composite of TM channels 3, 4, and 5 using an unsupervised classification algorithm (Ma et al. ms1.); adjacent pixels of the same spectral class are grouped into contiguous areas (Ford et al. ms.), and these spatial units are brought into a GIS as raster polygons, termed *regions* in ARC/INFO. The second stage involves a supervised classification process run within ARC/INFO to label all regions according to existing vegetation and land cover type (Ma et al. ms2.). The first stage is analogous to a process of manually digitizing polygon boundaries based on spectral patterns; literally, ***what you see*** on-screen ***is what you get*** in the output file. Similarly, the second stage, supervised classification, is analogous to a process of manually labeling polygons but uses digital attribute values for each region to determine the proper labels. In manual labeling techniques, non-digital attributes are typically assimilated from analysts' knowledge, aerial photo interpretation, and/or field validation (Scott *et al.*, 1993).

Euclidean distance (Equation 1) is central to the entire mapping process. This measure is used to assess similarities between input pixels and spectral classes in the unsupervised classification, as well as to assess similarities between raster polygons and cover types in the

$$D_{xy} = \sum_{i=1}^{M} (X_i - Y_i)^2 \qquad (1)$$

supervised classification. For the unsupervised classification, X represents a known spectral group and Y an unknown pixel; for the supervised classification, X is a known cover type in the training data set, and Y is an unknown raster polygon to be classified.

Naturally, this overview of the classification process is greatly simplified. In practice, numerous interdependent steps are required to produce the desired land-cover databases (Figure 2). Several digital databases were compiled as inputs to the classification process, and checkpoints were implemented at various stages to ensure accurate and consistent products.

Analyses were conducted on IBM RS/6000 workstations running AIX (version 4.1). Primary commercial software packages included: ARC/INFO (versions 7.04 and 7.11), ERDAS (version 7.5), and IMAGINE (version 8.1). In addition, for many major processing steps, we

constructed customized software written in FORTRAN and C (such as VIMAP, copyright 1994-96, Zhenkui Ma), or scripts written in Arc Macro Language (AML).

MAPPING INPUTS

Landsat TM imagery, digital elevation models (DEMs), digital hydrography, and ground-truth plots were the primary data layers incorporated in the mapping process. Data acquisition and database construction are outlined below. Details are provided in related appendices, as well as in digital metadata documents.

Landsat TM Imagery

Six TM scenes were used for this project. All imagery was collected during the growing season, mid-July to early August, and from 1991-1995. Two 1995 scenes (P41R29 and P41R30) were purchased from the EOSAT Corporation with terrain-correction done using proprietary techniques and ground control points. The Hughes/STX Corporation terrain-corrected three of the images (P40/R29, P41R28, and P42/R29) using proprietary techniques and ground control points. One scene (P40/R30) came from the Multiresolution Landcover Consortium (MRLC) archive and was terrain-corrected by the EROS Data Center (Sioux Falls, SD). The fourth scene (P34/R27) was purchased from EOSAT Corporation, which also did the terrain-correction using proprietary techniques and ground control points. All scenes were projected into an Albers Equal-Area Conic projection, NAD27 datum. Final pixel size was 30 m^2 for all six scenes.

Digital Elevation Data

Elevation, slope, and aspect information were derived from digital elevation data. U.S. Geological Survey (USGS) 7.5’ DEMs were used.

Digital Hydrography

USGS 1:100,000 digital line graphs (DLGs) were acquired for the full extent of all the TM imagery. When compared with recently published USGS topographic maps, inconsistencies were observed, such as missing features, miscoded features, and superfluous features in the digital version. We learned through consultation with USGS that their hardcopy maps (1:100,000 scale) were more reliable and accurate than the digital versions; consequently, we edited the DLGs to reflect information on the hardcopy maps. No physical changes were made to the original information contained in the DLGs; however, new attributes were added to flag miscoded, superfluous, and newly added features so that users could quickly modify the DLGs to match the hardcopy maps if necessary. Individual DLGs were then appended to create separate coverages for each TM scene, although data were archived and delivered in tiles corresponding to the USGS 1:100,000 quadrangles. Further details about this editing may be found in Redmond et al. (1996; Appendix C).

Ground-truth Data

Ground reference data were acquired from the Boise, Payette, Sawtooth, Salmon, and Challis National Forests. Additional ground reference data was supplied by the Salmon Resouce Area office of the BLM, the Boise Cascade Corporation, and from the USFS Region One classification project. ASCII files containing plot information were converted to ARC/INFO point coverages and then sorted and stored in separate coverages for each TM scene. To maximize the training data available for use in supervised classification, plots that fell within multiple scenes were maintained in multiple coverages. Additional training data were collected from USFS personnel during reviews of preliminary classifications at the University of Montana and from 7.5' quad maps printed after the initial classification of each scene using aerial photos or forest stand maps.

PROCESSING STEPS

Unsupervised Classification of Pixels

The two-pass, unsupervised classification procedure was designed by Zhenkui Ma to replicate the patterns observed in a color composite of TM imagery (channels 4, 5, and 3 assigned to red, green and blue). Again, what you see on-screen is what you get in the classified output file. If the colors of a classified image appear similar to the color composite from which it is drawn, important patterns presumably have been retained through the classification process. Spectral classes are defined based on Euclidean distance; the algorithm essentially is searching for the shortest distance between points in multivariate space, or in this case, the distance between RGB values for different pixels. Color similarity thus determines spectral class.

In the first pass of the unsupervised classification, a color palette file, which maps spectral values for the three TM channels to RGB values, was created by randomly sampling pixels to represent patterns evident in the color composite and thus to define spectral classes. In the second pass, Euclidean distances were calculated between pixels in the input image and spectral classes in the color palette; input pixels were assigned to the nearest spectral class as measured by Euclidean distance. Additional details may be found in Ma et al. (ms1).

Merging to Minimum Mapping Units

For this project, a specially designed merge procedure was developed to better delineate vegetation in riparian areas. The 30 m^2 pixel resolution was merged either to a one acre (0.4 hectare; greater than 4 pixels) minimum mapping unit (MMU) in zones where riparian vegetation was more likely to occur, or to a five acre (2.0 hectare; greater than 22 pixels) MMU outside this zone.

Groups of pixels smaller than these designated MMUs were merged with their most similar neighbor in a rule- and object-based process (Ford et al. ms). A similarity matrix controlled the incorporation of small patches into larger neighbors. This matrix was built based on the TM channel values for the input spectral groups. The merging process first identified regions smaller than the MMU, then listed neighboring regions and examined similarities between small regions and their neighbors. Finally, small regions were merged with larger neighbors having the most similar spectral values.

Zone Identification

Zones within which riparian vegetation was likely to occur were delineated for each TM scene using 7.5 minute DEMs (when available), 1:100,000 scale hydrography DLGs, and the spectral classes derived from the unsupervised classification. Delineation involved three steps:

1. Spectral classes generally associated with water or riparian areas were identified. Potential water spectral classes were subdivided into a ‘good’ and a ‘potential’ water type. ‘Good’ water spectral classes were almost always water, whereas ‘potential’ water often included other areas, such as cloud shadow or forested, northerly aspects.

2. Scene-wide data sets for hydrography (1:100K DLGs) and for topography (7.5 min DEMs patched with some 1 degree data) were prepared. Hydrologic features were buffered based on water body type (see Table 1) to define an initial zone within which riparian vegetation was likely to occur. Some types (with a buffer distance of 0) were excluded. This initial buffer area was further limited to a 5 meter rise from the stream or lake surface elevation to define the potential zone.

3. The final zone used for the merge procedure was defined as the area both within the potential riparian area and composed of potential riparian or water classes defined by the unsupervised classification.

Table 1. Buffer distance (meters) for hydrologic attributes used in merge.

Type of Water Body	Attribute Value	Buffer Distance
Glacier or Permanent Ice Field	103	0
Man Made Shoreline	201	0
Closure Line	202	0
Dam or Weir	406	0
Ditch or Canal	414	0
Aqueduct or Pipeline	415	0
Siphon	418	0
Tailings Pond	110	0
Intermittent Lake or Stream	610	150
Lakes	200	150
Left/Right Bank	605/606	450
All Other	—	300

Merging Steps

A seven-step merge process was used for the central Idaho scenes.

1. Riparian spectral classes within the zone, excluding water, were merged to a 5 pixel (1 acre) MMU within the riparian zone. This step provided the one-acre MMU for the riparian zone; however, smaller polygons remained along the edge of the zone or if they represented water.

2. Riparian and water classes outside the zone were merged to a 1 acre MMU. This merge helped to initially retain riparian areas outside the defined riparian zones.

3. All polygons inside of the zone and with a spectral class not likely to represent a

riparian cover type were merged to a 5 pixel MMU. This merge removed small island
pixels or polygons next to or within water polygons.

4. All polygons outside of the zone and not labeled as a 'good' water class were merged to a 23 pixel (5 ac) MMU. This step achieved the 5 ac MMU outside of the zone, while protecting the 'good' water.

5. All other pixels throughout the scene were merged to a 5 pixel (1 ac) MMU, excluding classes of potential water. This eliminated single pixels and small polygons along the edges of zones.

6. All potential water pixels within the scene were merged to a 12 pixel (2.5 acre) MMU. This step reduced the total number of polygons by grouping water pixels into larger polygons.

7. A final merge to 1 acre MMU was applied to all remaining pixels except those with a spectral class indicating 'good' water.

Once the merging process was finished, the shapes and sizes of all raster polygons (analogous to polygons but in raster format; termed ***regions*** in ARC/INFO) were established, and these were not changed in subsequent steps. These polygons serve as the base units in the raster database. At this point, all steps in the digital unsupervised classification were complete. To assign labels to these base regions through the supervised classification process, collection of ground-truth data was necessary.

Collection of Training Data for Supervised Classifications

Compilation of a training dataset is a critical step in classifying existing vegetation, because the resultant map can only be as good as the training data used to develop it. A two stage classification scheme was choosen for the central Idaho classification project. The initial classification of each scene used existing data supplied by the Boise, Payette, Sawtooth, Salmon, and Challis National Forests. In addition, new field data was collected by the forests using GPS units for plot locations. Emphasis was placed on only using existing data from large homogenous stands or areas. Use of existing data from highly hetergenous areas (where two or more different cover types occured) was discouraged due to the accuracy of the existing data plot locations. The following criteria was established to determine suitable existing data for use in the initial classification:

1. Existing data should not be over 10 years old, nor should the site have been disturbed (natural or management induced) since the data was collected.
2. The existing data represents an area 20 acres or larger (stand versus single plot data) for upland cover types and 1-5 acres for riparian cover types.
3. Location of the training data point must come from the center of the existing data stand or area.
4. For aerial or orthophoto interpretative training data the plot location should fall within a large homongenous stand of the cover type.

The new field data collected used two criteria. First, the plot should be near the center of a 5 acre or larger stand, and second, the stand be a homogeneous representation of the cover type. In addition to the existing and new field data supplied by the region four forests, field data collected with GPS units for the USFS Region One classification project (collected in 1994 and 1995) was used in the initial classification.

Personnel from the Boise, Payette, Sawtooth, Salmon, and Challis National Forests came to The University of Montana to review the initial classification and to help analysts input additional training data sites using timber stand maps and aerial photos. Again, emphasis was placed on choosing large areas where only one cover type occurred. A final source of training data for the second classification came from 7.5 minute quadrangles plotted from the initial classification for each scene. The 7.5 minute quadrangles (see table 2) were sent to each forest to be reviewed for correct and incorrect classification sites. The incorrect classification sites were marked with the correct cover type, tree size, and canopy cover if applicable.

Table 2. Names of 7.5 minute USGS quadrangles plotted for ground-truthing each TM scene.

P40/R29	P40/R30	P41/R28	P41/R29
Big Windy Peak	Arco Hills	Gibbonsville	Alturas Lake
Bohannon Spring	Bellevue	Ulysses	Amber Lakes
Double Spring	Copper Basin Knob	Ulysses Mountain	Blowfly Creek
Gold Stone Mountain	Hyndman Peak	Tincup Hill	Boiling Springs
Horse Basin	Mahony Butte		Brundage Mountain
Morrison Lake	Miller Peak		Bull Trout Point
Powderhorn Gulch	Sun Valley		Council Mountain
Warren Mountain			Custer

P41/R29	P41/R30	P42/R29
Easley Hot Springs	Anderson Ranch Dam	Crooked River Point
Gold Fork Rock	Deer Mountain	Pollock Mountain
Gooseberry Creek	Dollarhide Mountain	Rocky Comfort Point
Hazard Lake	Dunnigan Creek	Sturgell Peak
King Hill Lake	Idaho City	Weasel Gulch
Lake Mountain	Sprout Mountain	
Landmark	Sydney Butte	
Leesburg	Trinity Mountain	
Opal Lake		
Paddy Flat		
Pats Creek		
Pioneerville		
Scott Creek		
Stanley		
Swanholm Peak		
Twin Peak		
Twin Sisters		
Warren		

Ground-truth Data Processing Pipeline

After the data was obtained from the Forest Service in digital format or submitted on 7.5 minute quadrangles, it was subjected to a series of positional checks to ensure accuracy. In TM scenes P40R29, P41R28, P41R29, and P41R30, all training data points were run through geographic and ecological checks (supplied by the Boise, Sawtooth, Salmon, and Challis National Forests). The Region One data was checked for positional accuracy, and the geographic and ecological checks were used where appropriate. A total of 19,798 unique plots were compiled in the ground-truth database; of these, 7,905 fell within adjacent TM scenes. ARC/INFO point coverages with numerous attributes were then created and manipulated to obtain training datasets for each TM scene. Because of the overlap between scenes, individual plots may have been stored in multiple training set coverages; a total of 27,703 plots (Table 3) were available. These scene-based plot coverages were then passed to the image analysts for use in the supervised classification process (described below).

Table 3. Plots in the Central Idaho ground-truth database, summarized by TM scene.

TM Scene	R1 Data	Exist/Field	UM Review	7.5’ Quad	Totals
P40R29	1679	432	513	1229	3853
P40R30	0	2126	0	3452	5578
P41R28	614	676	334	1813	3437
P41R29	146	2469	622	4369	7606
P41R30	0	940	0	2948	3888
P42R29	128	438	300	2475	3341
Totals	2567	7091	1769	16286	27703

Setting up the Raster Database

The supervised classification, where regions were assigned labels based on ground-truth data, required a raster database containing multiple TM and ancillary attributes. Logically, then, classifications could not take place until such databases (or scene grids) had been constructed for each TM scene. Once a classified and merged image was created through the unsupervised classification process, the resultant file was converted from ERDAS GIS to ARC/INFO GRID format, thus maintaining its raster file structure. Each scene grid contained roughly 600,000 regions. Next, for each scene grid, a value attribute table (VAT) was built to contain statistics by region for spectral and biophysical (TM and DEM) data. In addition to the mean values for TM channels 1-7, mean elevation and slope values for each region were calculated and stored within the attribute table. Because mean values were unlikely to offer representative measures for aspect (e.g., when averaged, northeast and northwest slopes would be recorded as south), we classified aspect into eight groups and stored majority values for each region. We also calculated a version of modified NDVI (normalized difference vegetation index) adapted from Nemani et al. (1993):

$$MNDVI = (TM4 - TM3) / (TM4 + TM3 + 1) * (256 / (TM5 + 1)) * 100$$

Spatial attributes (hectares, perimeter length, and xy coordinates) were also calculated and recorded. The latter two were derived by converting the raster file to vector format; these attributes were automatically created in the conversion, and then their values were transferred back to the raster file. See Appendix I for a complete description of database attributes.

Training Data Analysis

As mentioned above, the supervised classifications could not occur until raster databases were constructed. Even more importantly, the supervised classifications would not be possible without the input of high quality data to 'train' the computer in which labels should be assigned to each region. Thus, the analysts next examined the scene-based plot coverages created during the ground-truth processing pipeline stage. Plots expected to cause problems were identified. If problems could not be resolved, the associated plots were eliminated from the training set. Examples include multiple plots with different vegetation types in a single polygon and plots with low or unknown locational accuracy. Each analyst examined the data for the same set of problems, but in a slightly different order. Generally, plots were set aside if they had cover type codes that would be manually rather than digitally labeled; these included urban, agricultural, and water cover types, but the specific set of excluded types varied by TM scene.

Training data analysis did not truly end until supervised classifications were complete. To assess the quality of training data at various stages of the entire process, we used a bootstrap method. This approach involved removing each training plot from the data set in turn and using the remaining plots to classify that single plot. A matrix then was generated from the classified output files to evaluate potential confusion among training plots. Diagonal elements in the 'confusion' matrix represent the number of plots properly classified; misclassified plots (omission and commission errors) were further evaluated, and in some cases, dropped altogether from subsequent analyses. Note that even if plots were correctly classified by others, they still may have been deleted; clusters of problem plots were fairly common. Confusion matrices for cover type based on the final training set are provided for each TM scene in Appendix C.

Supervised Classification of Regions

Once the raster database and training data sets were in place for a given TM scene, supervised classifications were conducted to assign cover type and size class labels to each region. This proved to be an iterative process. Multiple classifications may have been conducted for a single attribute, with intermediate modifications to training data, until satisfactory results were obtained. The algorithm used to classify regions is a supervised, nonparametric classification called the NEAREST MEMBER of GROUP (NMG). The mathematical description of NMG is:

> An unknown region Y belongs to group i if
>
> $ED(Y,X_i) < ED(Y,X_j)$ for $j <> i$
>
> where ED is Euclidean distance, X_i and X_j are supervised training data, and X_i is any one of the known training regions for group i.

With NMG, Euclidean distances are calculated between each unknown (i.e., unsampled) region and every training region in the data set; thus, all training regions are treated independently, and have an equal chance of affecting the assignment of labels. Each region is assigned a label corresponding to the group that contains the training region closest to the unknown region in terms of Euclidean distance.

The first step in each supervised classification was to overlay training plots with regions in the raster database and to extract the necessary attributes from each region for use in the classification. For each training plot, an attribute was added to identify the exact region in which it fell. Attribute tables were then related for the training plot and raster files, and the attribute values desired for each classification were exported into a training data file (ASCII format), sorted by group (cover type or size class). To classify cover types (Fig. 3), we used mean values for TM channels 1-7 and elevation. Because TM values range from 0-255 (well below typical values for elevation (m)), elevation was rescaled by dividing the raw value by 25 so that it would not be accorded extra weight in the classification. In a separate classification of forest size class, mean values of TM channels 1-7 and M_NDVI were incorporated; presumably, elevation should not influence size class as it could cover type. In addition to creating ASCII files for training data, similar files were created by exporting a matching set of attributes for every region in the

raster database. Using VIMAP software, the file of training data was compared with the file of regions to be classified, and every region was classified using the NMG algorithm. The NMG algorithm used Euclidean distances derived as follows:

$$(TM1_{train} - TM1_{unknown})^2 + (TM2_{train} - TM2_{unknown})^2 + \ldots (TM7_{train} - TM7_{unknown})^2 + (ELE_{train} - ELE_{unknown})^2$$

Distances between attribute values were squared to avoid mixing positive and negative values, and to magnify the amplitude of distances, thus helping to distinguish differences among groups. Attribute values thus played the primary role in determining which labels should be assigned to each region; these were the values between which Euclidean distances were calculated.

Three cover type labels were assigned to each region, in decreasing order of likelihood (or rather, in increasing Euclidean distance). Labels were maintained without any modifications in the raster database as COV_CODE_1, COV_CODE2, and COV_CODE3 (Appendix B). The actual Euclidean distance values (rescaled by dividing ED by 1000) also were recorded. Smaller values indicated a higher likelihood of correct classification: even if a cover type proved to be incorrect for a given polygon, a very small ED value would indicate that, based on the training data available, the assigned label offered the best possible fit. These ED values were used in evaluating classification results by looking at the relative differences between all three values for individual regions and at the various combinations of values across regions. ED values were also instrumental in making some modifications to cover type labels. Only the smallest Euclidean distance (COV_PROB_1) was maintained in the raster database. After the cover type classification was complete, a new attribute called COVERTYPE was added and populated with values from COV_CODE_1; this attribute was later manipulated through manual modifications. If COVERTYPE equalled COV_CODE_1, no modifications were made.

Manual Modifications

Agriculture

Agricultural lands were classified into two types: irrigated and dryland. Fields in active growth at the time the satellite imagery was acquired were assumed to be irrigated, whereas dryland agriculture included fallow and/or sparsely vegetated fields. All agricultural lands were identified and labeled manually according to a 6-step process described below; these in turn were based on the following four observations and assumptions about agricultural lands:

1. They tended to be associated with particular spectral classes.
2. They tended to occur in larger patches (> 25 pixels) than other types.

3. These patches tended to be more homogeneous in term of their spectral composition than other types.

4. They tended to be spatially clumped across entire TM scenes.

The first step involved determining the relationship between the two agricultural types and the 120-130 spectral classes derived from the unsupervised classification. Each spectral class was visually inspected and assigned to one of the following seven groups based on its perceived association with agricultural lands:

0) Never agriculture
1) Occasionally dry agriculture
2) Sometimes dry agriculture
3) Usually dry agriculture
4) Occasionally irrigated agriculture
5) Sometimes irrigated agriculture
6) Usually irrigated agriculture

The resulting file was used in conjunction with the unsupervised classification (the u-grid), the merged image (the m-grid), and the full database file (the z-grid) to produce a new output image (the outperc grid). An AML assigned output class values to every raster polygon based on the following three attributes: agricultural class value (0-6; see above list), the size of the polygon (number of pixels), and a homogeneity class value based on the spectral similarity of pixels in the u- versus m-grids in 10 percentile groups (see Table 4). In other words, if less than 10% of the pixels in the u-grid had the same spectral class value in the m-grid, then the polygon was assigned to the lowest homogeneity class (1). All raster polygons then were assigned to one of two size classes: small (<= 25 pixels) or large (> 25 pixels). Thus, the 6 possible agricultural classes, each of which could occur in 10 different homogeneity classes and two different sizes, yield a possible list of 120 agricultural type codes, plus another 10 codes for those that are never agriculture. Thus, the AML assigned one of these 130 output code values to each raster polygon in the grid. Referring again to Table 7, polygons with spectral class 0 were assigned to output classes 1-10, depending on their homogeneity percentile ranking. Polygons with spectral classes 1-6 were assigned to output values of 11-70 if they were less than or equal to 25 pixels in size, or 71-130 if they were larger than 25 pixels; a polygon with an output value of 115 should have a spectral class that indicated sometimes irrigated agriculture, a size greater than 25 pixels, and a 50-60% spectral correspondence between pixels in the u-grid and the m-grid.

In the third step, an analyst evaluated each output class for its association with agricultural land use and determined final breakpoints for irrigated versus dryland agriculture in each of the seven groupings. For example, the analyst might determine that polygons with spectral code 6 (indicating "usually irrigated agriculture") actually represent irrigated agriculture when the output class value is between 65 and 70 for small polygons or between 122 and 130 for large polygons.

The next two steps required that an analyst delineate areas within each TM scene where the output rules should be applied, and then apply them there. In other words, the classification of agricultural types was restricted to certain specific areas in each TM scene. In the sixth step, the final classification of the two agricultural cover types (2010 and 2020) was merged into the z-grid. Although this method was more time consuming and somewhat less objective than others, it had the advantage of producing considerably more accurate results.

Table 4. Relationships among variables used to classify irrigated and dryland agriculture for each TM scene.

Spectral Code [a]	Agricultural Class	Occurrence freq.	Size [b]	Output Class Values [c]
0	Non-Agriculture	n/a	n/a	1-10
1	Dry Agriculture	Occasional	SM	11-20
2	"	Sometimes	SM	21-30
3	"	Usually	SM	31-40
4	Irrigated Agriculture	Occasional	SM	41-50
5	"	Sometimes	SM	51-60
6	"	Usually	SM	61-70
1	Dry Agriculture	Occasional	LG	71-80
2	"	Sometimes	LG	81-90
3	"	Usually	LG	91-100
4	Irrigated Agriculture	Occasional	LG	101-110
5	"	Sometimes	LG	111-120
6	"	Usually	LG	121-130

[a] Recoded from original 120-130 spectral classes from unsupervised classification (u-grid).

[b] Number of pixels in raster polygon; SM < 26, LG > 25.

[c] One value per homogeneity percentile group; i.e., output class 1 = non-agriculture with 10% correspondence between pixels in u- versus m-grids; 2 = non-agriculture with 11-20%, etc.

Other Manual Modifications

Once supervised classifications had been conducted, manual modifications proved to be simple from within the raster database structure. Three basic types of manual modifications were used: attribute recoding based on decision rules, geographic limits (defined by the Forest Service for some cover types), and calculation of new attribute values. All cover type modifications were applied to the COVERTYPE attribute only.

As an example of a rule-based modification, ELE, ASP, and SLP were used in an ecological limit rule to check the classification of Lodgepole Pine. If COV_CODE_1 = 4203 (Lodgepole Pine), ELE < 2000 (meters), ASP = 5 (south aspect), and SLP > 6 (degrees or 15 percent slope), then the COV_CODE_1 field was incorrect and the COVERTYPE field should be set to COV_CODE_2.

Edge-matching

We employed a simple solution for seamlessly edge-matching classified data from the six adjacent TM scenes once the raster databases had been constructed. This ‘cookie cutter’ method was designed to preserve the integrity of individual scene classifications and to minimize the perception of an ‘edge’ between adjacent overlapping scenes. It allows each image to be processed independently using all available spectral information and then edgematched to its neighbors based on natural boundaries observed from land cover patterns. Through this method, adjacent scenes are ‘virtually’ edge-matched. Rather than physically deleting regions, they are simply flagged to indicate whether or not they should be used. As a result, the original data can always be retrieved, and new edge-matching schemes can be devised and implemented at any time.

Edge-matching occurred only within overlapping areas for adjacent scenes. Before edge-matching could occur, each of the six overlapping scenes within the mapping area was evaluated and ranked in descending order of dominance from P41R29, P41R30, P41R28, P40R29, P42R29, and P40R30. Primarily, P41R29 and P41R30 were most dominant because they were 1995 TM scenes used to reflect the changes from the 1994 fires. Factors such as the distribution of cloud cover, the image acquisition date, and the classification accuracy levels for each scene were carefully weighed in determining the rest of the dominance ranking. Once dominance relationships were established, the following processing sequence was implemented:

1. Create a polygon coverage encompassing the outer boundary of the six TM scenes.
2. Generally define the portion of each TM scene that should be kept (i.e., not lost in the edge-matching process) by drawing a polygon through the areas of overlap with the adjacent scene(s) and excluding portions with less desireable characteristics like cloud cover.
3. Perform overlay and masking operation in ARC/INFO to identify all polygons at least partially within the portion of the scene to be kept; add attribute, KEEP, to the database and populate with value = 1 for all these polygons.
4. Calculate a KEEP grid for each scene database; assign the value of the scene id to each polygon to be kept (i.e., with KEEP=1 in scene grid); all other polygons have no data.
5. In order of dominance, merge the KEEP grids into one single grid.

Accuracy Assessment

Because the land cover classification scheme employed for Central Idaho is complex (more than 40 cover types; see Appendix A), and because many land cover types overlap to varying degrees, map accuracy was evaluated using fuzzy sets (Gopal and Woodcock, 1994). A fuzzy matrix, derived from two-way tabulation of cover types, was constructed to evaluate the acceptability of various misclassification possibilities (Table 5). Acceptability was ranked through scores assigned to each cell in this matrix. For example, confusion between Douglas-fir (*Pseudotsuga menziesii*) and Douglas-fir/lodgepole pine (*Pinus contorta*) was determined to be less troublesome than was confusion between Douglas-fir and upland grasslands. Acceptability was rated on a scale from 1 to 5, as outlined by Gopal and Woodcock (1994): 1) absolutely wrong; 2) understandable, but wrong; 3) acceptable; 4) good; or 5) perfect match. When evaluating cover types according to this scheme, the following logic was applied:

• If cover type codes matched exactly, a score of 5 was assigned.

• If the codes did not match exactly, but the types shared a dominant species in the cover type name, a score of 4 was given (i.e., 4225 Douglas-fir - grand fir received score 4 if it was classified either 4212 Douglas-fir or 4207 grand fir).

• If the cover type was a commonly-occurring species in a mixed type, it received a score of 3 (i.e., 4203 lodgepole pine is a common component of 4220 mixed subalpine forest, and thus the confusion between these types is scored 3).

• If the cover type fell within the correct lifeform but was not similar to the species in the label, it was assigned a score of 2 (i.e., 4206 ponderosa pine and 4208 subalpine fir).

• If the lifeform was mismatched, a score of 1 was given.

By rating acceptability in this manner, accuracy assessments could be conducted at both the acceptable and ideal levels, thus offering more information than traditional approaches.

Separate assessments of accuracy were conducted for cover type, size class, and canopy cover. Plots from the test data sets were overlaid with the raster database; plot attributes were compared with classification results, and scores of 1-5 were drawn from the fuzzy matrix. Producer and user accuracies were assessed both individually (for scores 5, 4, 3, 2, and 1 in ‘match matrices’) and cumulatively (for scores 5+4+3+2+1 in ‘accumulate matrices’). For standard definitions of accuracy and general guidelines on interpreting error matrices, see Appendix J; Lachowski et al. (1995) also offers information on accuracy assessment and error matrices.

Final accuracy figures were weighted for each cover type using the proportion of the scene mapped as that cover type (as a surrogate for its population size), and results were scaled to a hypothetical population of 1000 units (to make numbers easier to interpret). Thus, the number of these units correctly classified for each cover type was estimated by multiplying the percentage of the scene mapped as that cover type, times the hypothetical population size of 1000, times the cumulative percentage of test plots correctly classified (using scores 5+4+3). These estimates were summed for each cover type, and the results were divided by 1000 to yield an overall area-weighted producer’s accuracy:

$$\sum \text{(\% cover type in scene)} * 1000 * \text{(\% correct plots)}\ 1000$$
$$= \sum \text{(\% cover type in scene)} * \text{(\% correct plots)}$$

Note that accuracy was not assessed for all the classes that were mapped. Specifically, classes like urban, agriculture, and water were omitted from the accuracy assessment, in part because the available test data were not representative of common occurrences of these classes. Nevertheless, because urban, agriculture, and water classes can be readily identified through visual interpretation, their actual accuracy should exceed 80%.

Quality Control

As a last step, the raster databases for each TM scene were examined for consistency and completeness, using an AML to standardize final products and to make them more convenient to use. To reduce file size, large background polygons (described below) were deleted from each TM scene grid, and the value attribute table was reordered to reflect this change. Because TM scenes are diagonal, and raster file systems deal with rectangles, each background polygon filled in the smallest possible square around the legitimate data values for that scene. The TM scenes were originally imported into ERDAS; this software package uses 0 to define NODATA, whereas ARC/INFO treats 0 as a legitimate value distinct from NODATA. Thus, when files were transferred to ARC/INFO, large meaningless background polygons were created.

In addition to deleting background polygons, attribute names were made more intuitive, attribute order was logically reorganized, and extraneous attributes were deleted; the attributes in the resulting database are described in Appendix B. Key attributes were queried to ensure that their values matched a range of acceptable possibilities.

RESULTS & DISCUSSION

We created an ARC/INFO raster database of existing vegetation and land cover for the Boise, Payette, Salmon/Challis, and Sawtooth National Forests and surrounding lands, mapping 4,420,268 hectares (10,922,482 acres) administered by the U.S. Forest Service and 7,962,766 hectares (19,675,994 acres) in all. In so doing, we compiled digital databases of topography and hydrography and classified portions of 6 Landsat TM scenes. Specific results for each TM scene are presented in Appendix C as frequency tables that summarize the total hectares mapped for each cover type. Results for the entire project area are highlighted below and are followed by a discussion of accuracy assessments for the individual TM scenes.

Mapping Trends in the Project Area

Cover Type

As mapped, the single most common cover type in the project area is 3350, or sagebrush (19.20%; Table 6). However, taken as a group, forest types dominate the region, covering 50.22% of the project area, 67.12% of the Southwest Idaho Ecogroup, and 61.77% of the Central Idaho Ecogroup (Table 7). For results specific to each National Forest, see Tables 8 and 9. Seventeen coniferous or broadleaf forest types are found in the project area, the most common of which are Douglas-fir (4212), mixed subalpine forest (4270), and lodgepole pine (4203). Fires of moderate or high intensity are mapped across 65,084 hectares, or 0.82% of the project area. Only 1.78% of the project area is occupied by urban areas (1100) or agricultural lands (2010 and 2020), pointing to the area's remoteness. Grassland types (3100 series) cover 12.02% of the project area; nearly a third of that area is mapped as altered grassland (3110). Mesic shrubs (3210) occupy about 4% of the area, and for the most part are interspersed with forest stands. With the exception of sagebrush, xeric shrubs are relatively uncommon, covering less than 3% of the area. Riparian areas are also fairly uncommon. Five riparian types are mapped within the project area; the most common is shrub riparian (6310, 0.66%), but all types are limited in area, and no one type is truly dominant. In all, riparian areas cover 1.91% of the project area, 1.36% of the Southwest Idaho Ecogroup, and 1.07% of the Central Idaho Ecogroup. Almost 7% of the project area is occupied by rock and barren types (7300, 7301, and 7800). Because of snow or cloud cover, 0.35% of the project area could not be mapped.

Size Class/Canopy Closure

For coniferous forests (4200 series), an additional attribute combining size class and canopy closure level was also assigned to each stand. For size class, 8% of all conifer stands are classified as sapling tree (1.0-4.9" dbh), 48% as small tree (5.0-12.0" dbh), 31% as medium tree (12.1-20" dbh), and 12% as large tree (>20.0" dbh). For canopy closure, 32% of all conifer stands are classified as low (10-40%), 48% as moderate (40-70%), and 19% as high (>70%). The most commonly mapped combination is small tree-moderate closure (9.83% of the project area, Table 10). This combination, assigned to more than 20% of all conifer stands, is more common in the Central Idaho (16.40%) than in the Southwest Idaho Ecogroup (11.70%, Table 11). In fact, the Central Idaho Ecogroup has a higher proportion of smaller trees than does the Southwest Idaho Ecogroup. Similarly, 10.99% of the Southwest Idaho Ecogroup and only 3.68% of the Central Idaho Ecogroup is mapped as large tree stands. For specific breakdowns by National Forest, see Tables 12 and 13.

Accuracy Assessment

Cover type classification accuracies are reported by TM scene in Table 14, including both Level 5 (exactly right) and Level 3 (acceptable) producer's accuracies. Accuracy levels were highest for P40/R30 (74% exactly right and 89% acceptable; Table 14) and lowest for P41/R29 (51% exactly right vs. 79% acceptable; Table 14). Level 5 accuracies for the other four TM scenes were all nearly 57%, and Level 3 accuracies exceeded 82% for three of the four. Generally, the improvement between Level 5 and Level 3 accuracies was attributable to the forest cover types (4000 codes; Table 14); except for cases with very few test plots, there was little difference between Level 5 and Level 3 accuracies for the grass (3100), shrub (3200), riparian (6000), or rocky/barren (7000) cover types. Nevertheless, these results should be viewed with caution; in particular, attention should be focused on the number of test and training plots available for each cover type. In many cases, these numbers were quite low and may lead to accuracy levels that are not true representations of accuracy across the entire TM scene.

In general, and when ample training data were available, the grassland types (3100) tended to classify quite well, especially the mesic montane parkland/subalpine meadow type (3180) and the altered herbaceous type (3110). Accuracies for the shrub (3200-3350), riparian (6000), and mixed barren (7800) types were variable, although for big sagebrush (3350), they

were quite high, ranging between 67% and 94%. Not surprisingly, exposed rock (7300) also classified well (6395%) in all six scenes.

Data Limitations

In general, many factors can influence the accuracy of classifications derived from Landsat TM data (see Congalton, 1991; Congalton and Green, 1993; Lachowski et al., 1995). These range from limitations associated with input data, including TM imagery or ground-truth data, to errors introduced in the classification process itself. We discuss these below.

TM Data

Given both the spatial and spectral resolutions of TM data, not all vegetation patterns can be delineated or classified accurately. For instance, we typically find variation in the spectral composition of 30 meter pixels representing the same cover type. If variation within a cover type is greater than the variation among different cover types, then these cover types necessarily will be confused spectrally with others. In the case of ponderosa pine, its spectral signature varied quite widely among five of the TM scenes, depending on tree size and density, as well as on the presence and amount of understory grass or shrubs. This contributed to spectral overlap between the ponderosa pine cover type and certain non-forest types, and undoubtedly led to lower classification accuracies (e.g., P41/R29 vs P40/R29; Table 14), not just for cover types, but also for lifeform and size class.

Time of year and atmospheric conditions affect the quality of TM data and any resulting classifications. Information about existing vegetation and land cover is best obtained through TM data acquired at certain times of the year. For existing vegetation, acquisition times close to the peak of the growing season are generally best, although to distinguish particular vegetation cover types, such as aspen (4101), spring or fall images might be best. Sun angle and atmospheric conditions also can adversely affect the quality of TM imagery. For example, in September, the sun is lower in the sky and casts more shadows in steep terrain than it does earlier in the season. Similarly, smoke or haze can interfere with spectral reflectance patterns and thereby limit the variation available in TM data. In the western U.S., hazy atmospheric conditions commonly occur in late August and September as a result of wildfires. For all these reasons, mid-summer (late July through mid-August) should be the ideal time to acquire TM data for this project area; we were fortunate to be able to obtain high quality images from this time period and to have them all terrain-corrected.

Reflectance data from the TM thermal channel (band 6) were missing from two scenes: P40/R30 and P41R28, covering portions of Sawtooth and Salmon/Challis National Forests. These scenes were obtained from the Multi-Resolution Land Characterization Consortium archive at the USGS EROS Data Center. Because our supervised classification method used mean values for all seven TM channels, the lack of thermal information for this one scene hindered our ability to distinguish among general cover types, like water, rock, grass, shrub, and forest, that we know absorb, radiate, and reflect heat differently. For the other four TM scenes, we observed general patterns of association between thermal values and cover classes, such as low values for water, moderate values for forest, and relatively high values for grass, shrub, and rock. Similarly, cover types occurring on south and west facing slopes tended to have higher thermal values than did cover types on north and east facing slopes. Thus, we feel that the absence of thermal data for the two TM scenes reduced the power of our classification method and may have reduced the resulting map accuracies as well, although this may only be evident for the exposed rock class in Table 14.

Map Unit Definition

Our unsupervised classification and merging process produced regions as small as 0.4 ha MMU for the six TM scenes. First, spectral classes corresponding to water were preserved at 30 m resolution in the merging process. This did not guarantee that they would — or even should — be labeled water in the supervised classification. Such instances are especially common on steep, north facing slopes that appear shadowed on the TM imagery. Although there are numerous regions of this sort, they occupy little area overall. Second, along edges, regions smaller than 2 ha may have been created and maintained. The edge-matching process in particular reshaped and/or subdivided many regions on subordinate scene edges. In addition to creating 'remnant' polygons smaller than the MMU, this process also created some non-contiguous regions wherever subordinate regions were subdivided by overlapping dominant ones. Any attempt to rectify these would have changed the region attributes; because these are the attributes we used in the classification process, we opted to leave the database intact to be fixed by individual users if desired.

Ground-truth Data

Much of the reference data used in this project came from existing sources within the Forest Service. In other words, very few new data plots were collected in the field, and we made no effort to devise an unbiased sampling methodology. Nonetheless, given the relative abundance of reference data available for each TM scene, as well as the results of the accuracy assessments, we do not believe that the classification results were overly biased in favor of cover types that were managed by the Forest Service, rather than in favor of those types occurring either in inaccessible areas, or on privately owned lands.

We were very selective in our use of training data for each TM classification, but it was not possible to eliminate all possible sources of error. For example, the accuracy of ground-truth plots, both in terms of their location and content, will directly affect classification accuracies. Again, we tried to identify plots with potential errors early in the processing pipeline. If these errors could not be corrected, then the individual plots were eliminated. Undoubtedly, however, some errors were missed. If these ended up in the training data set, they could still be detected by further quality control measures. If they were not detected and used for training, however, their effects would be distributed throughout the classification and result in a lower map accuracy.

It was inevitable that more plots would be available for some scenes than for others (e.g., P41/R28 vs. P41/R29; Table 14). To maximize the use of all available data, we tried to use data from plots that were located on multiple scenes. However, because adjacent scenes were sometimes acquired in different years, and possibly under different environmental or atmospheric conditions, this may not have been appropriate. When these plots were selected for training two or more adjacent image classifications, they were removed if they did not meet certain filtering criteria.

When the riparian training data was compared to the modeled riparian zones, some limitations were noted in the size of the riparian zones. In all six TM scenes, some of the modeled riparian zones were too narrow, thus missing areas where riparian training data occurred. Since only polygons with a majority of their area inside the modeled riparian zone could be labeled as riparian covertypes, the overall riparian classification could be underrepresented. One limitation in riparian classification is the 30 meter pixel size, which in some cases is larger than the riparian area.

All TM scenes were acquired between 1991 and 1995, whereas the ground-reference data could have been acquired either before or after the TM imagery. For stable, slow changing cover types, these temporal differences should not cause classification problems, unless land use suddenly changed. Again, our training data analysis was designed to catch gross inconsistencies between field and satellite data, but we certainly did not eliminate all errors.

Related to the topics of plot location and patch uniformity, some pre-existing field data were collected for purposes and at scales that do not match the 30 m spatial resolution of TM data. Even though single pixels might represent relatively pure cover types on the ground, it is important to remember that we were classifying patches 0.4 ha and larger. Thus, when single pixels are combined into larger mapping units, both training and test data must match this scale. Once again, we tried to minimize this potential conflict in the classification process by careful analysis of the training data and by removal of questionable plots.

CONCLUSIONS

In spite of the limitations discussed above, we maintain that our two-stage procedure for mapping large areas was a suitable and effective approach for a project of this magnitude. By providing an integrated GIS database, users can query and process information contained in many layers in a cost-efficient manner and at a controlled level of accuracy. Furthermore, all stated accuracies represent conservative estimates, and users should find the GIS databases to be more accurate and useful than these figures might otherwise indicate. Where inaccuracies are real, the nature of the database makes modifications relatively simple and straightforward. Although the database covers a large geographic area, the full scene units have been maintained for archival, online storage and for eventual updating with new TM imagery. Being relatively small (~100 megabytes each), these full scene databases can be archived on CD-ROM, retrieved quickly, and processed efficiently. Thus, many problems arising from conventional classifications of Landsat images across large areas have been resolved through methodology developed for this project.

REFERENCES

Anderson, J. R., E. E. Hardy, and J. T. Roach. 1975. A land use and land cover classification system for use with remote sensor data. USGS Professional Paper 964. U. S. Government Printing Office. iii+ 28pp.

Baret, F., G. Guyot, A. Begue, P. Maurel, and A. Podaire. 1988. Complementarity of middle infrared with visible and near-infrared reflectance for monitoring wheat canopies. *Remote Sensing of Environment*, 26: 213-225.

Brown, J. F., T. R. Loveland, J. W. Merchant, B. C. Reed, and D. O. Ohlen. 1993. Using multisource data in global land cover characterization: concepts, requirements, and methods. *Photogrammetric Engineering and Remote Sensing,* 59(6): 977-987.

Butera, C. 1986. A correlation and regression analysis of percent canopy closure and TM spectral response for selected forested sites in San Juan National Forest, Colorado. *I.E.E.E. Transactions on Geosciences and Remote Sensing*, GE-24, 122-129.

Congalton, R. G. 1988. A comparison of sampling schemes used in generating error matrices for accessing the accuracy of map generated from remote sensing data. *Photogrammetric Engineering and Remote Sensing* , 54(5): 593-600.

Congalton, R.G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37: 35-46.

Congalton, R.G., and K. Green. 1993. A practical look at the sources of confusion in error matrix generation. *Photogrammetric Engineering and Remote Sensing*, 59(5): 641-644.

Crist, E. P., and R. C. Cicone. 1983. Investigation of Thematic Mapper data dimensionality and features using field spectrometer data. *Proceedings of the Seventeenth International Symposium on Remote Sensing of Environment*, Vol. III, pp. 1313-1322.

Davis, F.W., P.A. Stine, and D.M. Stoms. 1994. Distribution and conservation status of coastal sage scrub in southwestern California. *Journal of Vegetation Science,* 5(5):743-756.

Environmental Systems Resource Institute. 1991. Cell-based modeling with GRID. Version 6.0, Printed in the United States of America.

ERDAS. 1991. Field guide, Second Edition, Version 7.5, Printed in the United States of America, 394 p.

Ford, R., Z. Ma, S. Barsness, and R. Redmond. Ms. Rule-based aggregation of classified imagery. Submitted to Computational Science & Engineering, Feb. 1997.

Gopal, S. and C. Woodcock. 1994. Theory and methods for accuracy assessment of thematic maps using fuzzy sets. Photogrammetric Engineering and Remote Sensing, 60(2): 181-188.

Guo, J. 1993. An object-oriented model with efficient algorithms for identifying and merging raster polygons. M.S. Thesis, Department of Computer Science, The University of Montana, Missoula, Mont.

Heckbert, P., 1982. Color image quantization for frame buffer display. *Computer Graphics,* 16(3): 297-307.

Heuvelink, G. B. M., P. A. Burroughs, and A. Stein. 1989. Propagation of errors in spatial modelling with GIS. *International Journal of Geographical Information Systems*, 3(4): 303-322.

Hitchcock, C.L. and A. Cronquist. 1973. Flora of the Pacific Northwest. Univ. Washington Press, Seattle. 730pp.

Hixson, M., D. Scholz, N. Fuhs, and T. Akiyama. 1980. Evaluation of several schemes for classification of remotely sensed data. *Photogrammetric Engineering and Remote Sensing,* 46(12):1547-1553.

Justice, C., and J. Townshend. 1982. A comparison of supervised and unsupervised classification procedures on LANDSAT MSS data for an area of complex surface conditions in Basilicata, southern Italy. *Remote Sensing of Environment,* 12:407420.

Keane, R. E., M. E. Jensen, and W. J. Hann. 1990. ECODATA and ECOPAC: analytical tools for integrated resource management. *The Compiler,* 8(3):24-29, 31, 3337.

Lachowski, H., P. Maus, M. Golden, J. Johnson, V. Landrum, J. Powell, V. Varner, T. Wirth, J. Gonzales, S. Bain. 1995. Guidelines for the use of digital imagery for vegetation mapping. USDA: Forest Service, Gen. Tech. Rep. EM-7140-25, Washington, D.C.

Lillesand, T. M. and R. W. Kiefer. 1987. Remote Sensing and Image Interpretation, 2nd ed., John Wiley & Sons, New York, New York, 721 p.

Ma, Z., 1995. Using a rule-based merging algorithm to eliminate 'salt-pepper' and small regions of classified image. *Proceedings of the Ninth Annual Symposium on Geographic Information Systems,* Vol. II, pp. 834-837.

Ma, Z. and C. E. Olson. 1989. Spectral matching in supervised classification of terrain features. *IGARSS '89*, Vol. II, pp. 638-640.

Ma, Z. and C. E. Olson. 1989. A measurement of spectral overlap among cover types. *Photogrammetric Engineering and Remote Sensing*, 55(10): 1441-1444.

Ma, Z. and C. E. Olson. 1990. Accuracy evaluation of computed classified thematic map with a systematic cluster sampling scheme. *IGARSS '90*, Vol. III, pp. 23792382.

Ma, Z. and R. L. Redmond. 1996. Building attribute tables for raster GIS files with Arc/Info. Pp. 198-201 *in*: Raster Imagery in Geographic Information Systems, S. Morain and S. Lopez Baros, Eds. OnWord Press, Santa Fe, NM.

APPENDIX A:

A Vegetation & Land Cover Classification System for Central Idaho

This vegetation and land cover classification system for central Idaho was developed for the ground-truthing component of the Forest Service Upper Columbia Basin Assessment and modified for the central Idaho National Forests. The system is based on a classification developed for the Montana and Wyoming Gap Analysis projects to aid in the preliminary classification of unsupervised Landsat TM imagery at a mid-scale level of 1:100,000.

Each cover type in the hierarchical classification system is given a four-digit code. The first digit of the class codes are based on the USGS standard (Anderson 1986) for assigning satellite imagery land cover types (1=urban, 4=forests, etc.). Cover types were mapped to one of three levels in the key: General Group (e.g., 3100, Upland Grasslands), Parent Group (e.g., 3110, Altered Herbaceous Grasslands), and Sub-code Group (e.g., 3111, Non-native Grasses).

The classification categories and the rules for making cover type assignments are intended to be exclusive. Criteria are provided to assist field crews in assessing vegetation and land type cover using estimates of canopy cover for life forms and/or specific species types. Canopy cover is defined as an expression of the percentage of the total natural spread of foliage of an individual species, including interstitial spaces of branches. This classification system defines classes based on the distinction between absolute and relative canopy cover. The following criteria are used in sequence to determine the land cover type:

Land cover type breaks:

1. Developed or Non-Developed Lands? (Urban-Agricultural or Non-Urban-Ag.)

2. Riparian or Upland Site?

3. Life form type?

A. Is the site Forest dominated?	Forest Cover (FC) ≥ 10%
B. Is the site Shrub dominated?	FC < 10%, Shrub Cover (SC) ≥15%
C. Is the site Grass dominated?	FC < 10%, SC < 15%, Herbaceous Cover (HC) ≥ 15%
D. Is the site Barren or Rock?	FC < 10%, SC < 10% and HC < 10%
E. Is the site Alpine?	Vegetation above tree line

4. Species/Group type.

	General	Parent	Sub-code
DEVELOPED LANDS			
URBAN—AGRICULTURAL LANDS 1000-2999			
Manually Classified			
Urban	1100		
Agricultural	2000		
Agriculture-Dry		2010	
Agriculture-Irrigated		2020	
NON-DEVELOPED LANDS			
GRASSLANDS 3100-3199			
Forest Cover <10%, Shrub Cover <15%, and Herbaceous Cover $15%			
Upland Grasslands	3100		
Altered Herbaceous		3110	
Non-Native grass			3111
Noxious Weeds			3121
Mesic Montane Parklands & Subalpine Meadows		3180	
SHRUBLANDS 3200-3499			
Forest Cover <10%, Shrub Cover $15%			
Mesic Shrubs	3200		
Mesic Shrubs (single or mixed species)		3210	
Xeric Shrubs	3300		
Single Species Xeric Shrubs (one species > 66% of Shrub Cover)			
Mtn Mahogany		3301	
Bitterbrush		3304	
Rabbitbrush		3312	
Shadscale		3318	
Big Sagebrush Steppe		3350	
Mountain Big Sagebrush			3351
Wyoming Big Sagebrush Steppe			3352
Basin Big Sagebrush			3353
Black Sagebrush Steppe			3354
Low Sagebrush Steppe			3355
Tri-tip Sagebrush			3356

	General	Parent	Sub-code
FORESTLANDS 4000-4999			
Forest Cover $10%			
A. Is it Broadleaf or Conifer Dominated or Mixed Broadleaf/Conifer?			
(Broadleaf > 66% Forest Cover; Conifer > 66% Forest Cover)			
1. Is it a Single Species Stand? (one species > 66% Forest Cover)			
2. Is it a Two Species Stand? (Sum of two species $ 80% Forest Cover)			
3. Is it a Mixed Species Stand?			
Broadleaf Forest 4100-4199			
Broadleaf Forest	4100		
Single Broadleaf Species			
Aspen		4101	
Conifer Forest 4200-4299			
Conifer Forest	4200		
Single Conifer Species Stands			
Engelmann Spruce		4201	
Lodgepole Pine		4203	
Whitebark Pine		4204	
Limber Pine		4205	
Ponderosa Pine		4206	
Grand Fir		4207	
Subalpine Fir		4208	
Douglas-fir		4212	
Rocky Mtn Juniper		4214	
Utah Juniper		4216	
Two-conifer Species Stands			
Douglas-fir/Lodgepole Pine		4223	
Douglas-fir/Grand Fir		4225	
Douglas-fir/Ponderosa Pine		4230	
Douglas-fir/Limber Pine		4231	
Douglas-fir/Engelmann Spruce		4232	
Mixed Conifer Species Stands			
Mixed Whitebark Pine Forest		4260	
WBP $10%			
Mixed Subalpine Forest		4270	
WBP 1-9% or SF $10% or ES $10%			
Mixed Subalpine Forest with Douglas-fir			4271
Mixed Subalpine Forest without Douglas-fir			4272

	General	Parent	Sub-code
Mixed Mesic Forest		4280	
RC or GF or WL $10% & DF or PP			
Mixed Mesic Forest with Ponderosa Pine			4281
Mixed Mesic Forest without Ponderosa Pine			4281
Mixed Xeric		4290	
RMJ, UJ, PF, DF, PP			
Other Forest Types 4300-4500			
Mixed Broadleaf and Conifer Forest		4300	
Standing Burnt or Dead Forest		4400	
Moderate Intensity Fire			4402
High Intensity Fire			4403
Timber Harvest Units		4500	

WATER 5000-5999

	General	Parent	Sub-code
Water	5000		

RIPARIAN 6000-6999

A. Is it Forest Dominated Riparian? FC > 10%

B. Is it Shrub Dominated Riparian? FC < 10%, SC $ 15%

C. Is it Graminoid Dominated Riparian? FC < 10%, SC < 15%, HC > 15%

	General	Parent	Sub-code
Tree Dominated Riparian	6100		
Conifer Dominated Riparian		6110	
Broadleaf Dominated Riparian		6120	
Mixed Tree Riparian		6130	
Shrub Dominated Riparian	6300		
Shrub Dominated Riparian		6310	
Xeric Shrub Dominated Riparian			6311
Mesic Shrub Dominated Riparian			6312
Willow Dominated Riparian			6313
Herbaceous Dominated Riparian	6200		
Graminoid & Forb Dominated		6210	
Xeric Herbaceous Riparian			6215
Moist Herbaceous Riparian			6216

	General	Parent	Sub-code
BARREN LAND 7000-7999			
Tree Cover, Shrub Cover, and Herbaceous Cover <10%			
Barren Land	7000		
Rock-Dominated Sites		7300	
Basalt Flows			7301
ALPINE AREAS 8000-8199			
Areas above Tree Line			
Alpine Areas	8000		
Alpine Meadows		8100	
SNOW AND CLOUDS 9000-9999			
Snowfields or Ice	9100		
Clouds	9800		
Cloud Shadow	9900		

Tree Size Class		**Code**	**Canopy Cover**	**Code**
Sapling	(1.0 - 4.9” DBH)	1	1-9%	1
Small	(5.0 - 12.0” DBH)	2	10-40%	2
Medium	(12.1 - 20.0” DBH)	3	41-70%	3
Large	(>20.0” DBH)	4	71-100%	4

Description of Land Cover Types Used in the Final Classification

1. 1100 - Urban:
Developed cities or towns. Manually classified.

2. 2010 - Agricultural Dry-Lands:
Agricultural lands using dry-land farming techniques. Manually classified.

3. 2020 - Agricultural Irrigated-Lands:
Agricultural lands using irrigation farming techniques. Manually classified.

4. 3100 - Upland Grasslands:
Regions with < 10% tree cover, < 15% shrub cover, and grass cover ≥ 15%. Dominated by native grass species of the interior basin or palousse praire. Some polygons may have pockets of trees or shrubs present. Non-native grass species maybe present up to 30% cover.

5. 3110 - Altered Grasslands:
Regions with < 10% tree cover, < 15% shrub cover, and grass cover ≥ 15%. Dominated by non-native grass species or noxious weeds. Some polygons may have pockets of trees or shrubs present.

6. 3180 - Mesic Montane Parklands and Subalpine Meadows:
Regions with < 10% tree cover, < 15% shrub cover, and grass cover ≥ 15%. Dominated by upper elevation mesic grass or forb species. Timber harvest units or recent-older fire areas can be classified as this type if mesic grass or forb species dominate the site.

7. 3210 - Mesic Shrublands:
Regions with < 10% tree cover and ≥ 15% shrub cover. Dominated by mesic shrubs and found above the great basin/interior basin valleys. Timber harvest units or recent-older fire areas can be classified as this type if mesic shrubs dominate the site.

8. 3301 - Mountain Mahogany:
Regions with < 10% tree cover and ≥ 15% shrub cover. Dominated by Mountain Mahogany shrubs, typically found with exposed rocks. Some polygons may have pockets of Douglas-fir present within them.

9. 3304 - Bitterbrush:
Regions with < 10% tree cover and ≥ 15% shrub cover. Dominated by Bitterbrush and grasses, or exposed soils. Sagebrush and/or rabbitbrush may also be present in the area with dominance shifting between Bitterbrush and sagebrush.

10. 3318 - Shadscale:
Regions with < 10% tree cover and ≥ 15% shrub cover. Dominated by Shadscale and grasses, or exposed soils. Sagebrush and/or rabbitbrush may also be present in the area.

11. 3350 - Sagebrush:
Regions with < 10% tree cover and ≥ 15% shrub cover. Dominated by one or more of the following sagebrush types: Mountain Big Sagebrush, Wyoming Big Sagebrush, Basin Big Sagebrush, Black Sagebrush, Low Sagebrush, or Tri-tip Sagebrush. Grasses may have equal or higher coverage than the sagebrush. Timber harvest units occassionally are mapped as this type.

12. 4101 - Aspen:
Regions with ≥ 10% tree cover. Dominated by aspens, occuring as homogeneous stands or in aspen/grass pockets. Sites may have exposed rock present. Typically limited to stands less than 20 acres.

13. 4200 - Conifer Forest Type:
Regions with ≥ 10% tree cover. Polygons that did not pass the ecolimits tests for cov_code_1, cov_code_2, and cov_code_3 and all three cov_code types were conifer. The default code of 4200 was assigned to the polygon.

14. 4203 - Lodgepole Pine:
Regions with ≥ 10% tree cover. Dominated by Lodgepole Pine and found in mid-high elevation areas. May have other conifer species present typically Douglas-fir, Engelmann Spruce, or Subalpine fir.

15. 4204 - Whitebark Pine:
Regions with ≥ 10% tree cover. Dominated by or having > 10% cover of Whitebark Pine in the stand. Typically occurs only at high elevation, near treeline, on southern or westerly aspects.

16. 4205 - Limber Pine:
Regions with ≥ 10% tree cover. Dominated by Limber Pine, found only in the south-eastern courner of the study area. Typically has low to moderate cover and may occur in patches.

17. 4206 - Ponderosa Pine:
Regions with ≥ 10% tree cover. Dominated by Ponderosa Pine, typically found in lower forest elevations.

18. 4207 - Grand Fir:
Regions with ≥ 10% tree cover. Dominated by Grand Fir, only found in western Idaho, and is associated with mixed mesic forest types.

19. 4208 - Subalpine Fir:
Regions with ≥ 10% tree cover. Dominated by Subalpine Fir, typically found in higher elevations. May occur in small pure patches with mixed subalpine fir stands adjoining them.

20. 4212 - Douglas-fir:
Regions with ≥ 10% tree cover. Dominated by Douglas-fir within the stand. Commonly found in low to mid elevation conifer stands, may have patches of Lodgepole Pine or Ponderosa Pine within them.

21. 4223 - Douglas-fir/Lodgepole Pine:
Regions with ≥ 10% tree cover. Dominated by both Douglas-fir and Lodgepole Pine together within the stand. Polygons may have patches of pure Douglas-fir or Lodgepole Pine within them.

22. 4225 - Douglas-fir/Grand Fir:
Regions with ≥ 10% tree cover. Dominated by both Douglas-fir and Grand Fir together within the stand. Polygons may have patches of pure Douglas-fir or Grand Fir within them.

23. 4230 - Douglas-fir/Ponderosa Pine:
Regions with ≥ 10% tree cover. Dominated by both Douglas-fir and Ponderosa Pine together within the stand. Polygons may have patches of pure Douglas-fir or Ponderosa Pine within them.

24. 4260 - Mixed Whitebark Pine Forest:
Regions with ≥ 10% tree cover. High elevation conifer stands with whitebark pine present. Other conifer species present are: Subalpine Fir, Engelmann Spruce, and Lodgepole Pine.

25. 4270 - Mixed Subalpine Fir Forest:
Regions with ≥ 10% tree cover. Upper elevation conifer stands with a combination of the following speices: Subalpine Fir, Engelmann Spruce, Lodgepole pine, or Douglas-fir.

26. 4280 - Mixed Mesic Forest:
Regions with ≥ 10% tree cover. Low to mid-elevation conifer stands with a combination of the following species: Douglas-fir, Grand Fir, Larch, Engelmann Spruce, Ponderosa Pine, Lodgepole Pine, or Subalpine Fir.

27. 4290 - Mixed Xeric Forest:
Regions with ≥ 10% tree cover. Lower elevation dry conifer sites with a combination of the following species: Ponderosa Pine, Douglas-fir, Limber Pine, or Lodgepole pine.

28. 4300 - Mixed Broadleaf and Conifer Forest:
Regions with ≥ 10% tree cover. Comination of Aspen and Conifer trees, usually Douglas-fir. Typically found in stands less than 20 acres.

29. 4400 - Standing Burn Timber:
Areas identified as having recent fires based on fire layers from the National Forests. Two sub-codes are 4402 (moderate intensity burn) and 4403 (high intensity burn). Manually classified using the fire layers from the national forests.

30. 5000 - Water:
Regions where water is dominant feature of polygon.

31. 6110 - Conifer Dominated Riparian:
Riparian regions with ≥ 10% tree cover. Dominated by conifer trees within the stand. Patches of broadleaf trees or shrubs may be present within the polygon.

32. 6120 - Broadleaf Dominated Riparian:
Riparian regions with ≥ 10% tree cover. Dominated by broadleaf trees within the stand. Patches of conifer trees or shrubs may be present within the polygon.

33. 6130 - Conifer and Broadleaf Dominated Riparian:
Riparian regions with ≥ 10% tree cover. Dominated by both conifer and broadleaf trees together within the stand. This class is often confused with conifer dominated stands.

34. 6210 - Herbaceous Dominated Riparian:
Riparian regions with < 10% tree cover, < 15% shrub cover, and ≥ 15% herbaceous cover. Typically dominated by carex or sedge communities with or without standing water. Cattail marshes are also included in this cover type.

35. 6310 - Shrub Dominated Riparian:
Riparian regions with < 10% tree cover and ≥ 15% shrub cover. Two distinct types of shrub dominated riparian are willow dominated areas with standing water, and non-willow dominated areas without standing water.

36. 7300 - Exposed Rock:
Regions dominated by exposed rock, talus or scree. Patches of trees, shrubs, or grass may occur within the rock areas. 7301 Basalt is a subcode of the 7300 (exposed rock) code.

37. 7800 - Barren Areas:
Regions that are barren or are dominated by barren areas with patches of trees, shrubs, or grass within them.

38. 9100 - Snow:
Regions that had snow as a dominant cover when the imagery was taken.

39. 9800 - Clouds:
Regions that had a cloud covering the area in the image.

40. 9900 - Cloud shadow:
Regions that had a cloud shadow covering the area in the image.

APPENDIX B: Land Cover Attributes for the Central Idaho Classification Project

The following items are contained in the raster databases (ARC/INFO GRIDs) of existing vegetation and land cover created for each of the 6 Landsat TM scenes. Each of the 6 land cover grids is roughly 80 megabytes in size and has an average of 520,000 *regions*, or raster polygons (range ~ 470,000 - 620,000). Land cover grids are named according to the following convention: p*r*z (e.g., p41r28z). Actual item definitions, units, and valid ranges are included in Tables B-1 (for land cover) at the end of this appendix.

Note that after edge-matching, some attribute values (especially mean and majority values) may be rendered inaccurate for subordinate regions in one scene that are partially obliterated or subdivided by dominant regions from the adjacent scene. Regions that are subdivided will be noncontiguous; without a second region-grouping to reassign values, these regions cannot be identified in the output (GRID) database. They can be assigned unique id values, however, using the vector polygon option in the revised clipper AML. By repeating the operations used to derive those values (such as ZONALSTATS), attributes may be updated as needed. Original values were maintained because these are the values on which the classifications of cover type, size class, and canopy cover were based.

VALUE: Unique region identification number created by ARC/INFO; value should never be changed because it plays a critical part in maintaining raster topology, or the link between spatial and attribute data.

COUNT: Number of 30 m^2 pixels (or 15 m^2 for P30R27) in the region; value calculated by ARC/INFO and should not be changed.

LIFEFORM: Dominant lifeform of the region; assigned by an AML that queries the COVERTYPE value. If null, the region falls outside of the study area boundary.

N = non-vegetation
T = tree
S = shrub
G = grass/forb
B = barren
R = rock
W = water
A = agriculture
U = urban

COVERTYPE: Land cover type assigned to the region (Appendix A), e.g. 4206 = Ponderosa Pine Forest, 3201 = Mesic Upland Shrub, 7300 = Rock, etc. If this value does not equal the value in COV_CODE_1, then it has been manually modified. If zero, the region falls outside of the study area boundary.

TREESIZE: Tree size class assigned to the region (Appendix A). e.g. 1 = sapling, 2 = small tree, 3 = medium tree, 4 = large tree, 0 = non conifer covertype. Tree size class was only classified for regions with a conifer covertype code (4200-4290). If zero, the region falls outside the study area boundary.

CANOPYCOVER: The amount of Shrub or Tree Canopy Cover for the region (Appendix A). e.g. 2 = 10-40%, 3 = 41-70%, 4 = 71-100%, 0 = non shrub or tree covertype. If zero, the region falls outside the study area boundary.

BURN_INTENSITY: The fire burn intensity for the region (only regions that fell inside of fire boundaries provided by the Boise & Payette NF's). Codes are: 1 = low intensity, 2 = moderate intensity, 3 = high intensity, 0 = region outside of fire boundaries. Note: covertype was manually changed to 4402 for moderate intensity regions and 4403 for high intensity regions.

ELE: Mean elevation (in meters) of region; value calculated in ARC/INFO (ZONALSTATS) using 7.5 minute DEM data when available; otherwise, 3 arc second data resampled to 30 m^2 by Hughes/STX.

SLP: Mean slope value of region (in degrees); value calculated in ARC/INFO (ZONALSTATS) using 7.5 minute DEM data when available; otherwise, 3 arc second data resampled to 30 m^2 by Hughes/STX.

ASP: Majority aspect value of region; value calculated in ARC/INFO (ZONALSTATS) using 7.5 minute DEM data when available; otherwise, 3 arc second data resampled to 30 m^2 by Hughes/STX. Manually recoded to the following nine classes:

0 = flat
1 = north
2 = northeast
3 = east
4 = southeast
5 = south
6 = southwest
7 = west
8 = northwest

SPECTRAL_CLASS: Alternate name LINK. Spectral class code from unsupervised classification; value input directly into ARC/INFO and should not be changed.

COV_CODE_1: Most likely cover type assigned to region by supervised classification (Nearest Member of Group classifier; VIMAP).

COV_PROB_1: Euclidean distance that led to the cover type assigned by the supervised classification (COV_CODE_1); this is the smallest Euclidean distance that was calculated between attributes for this region and those for any region in the training data set. Assigned through supervised classification based on Nearest Member of Group classifier. Value has been rescaled by dividing by 1000:

$$((TM1_{train} - TM1_{unk})^2 + (TM2_{train} - TM2_{unk})^2 + ...(TM7_{train} - TM7_{unk})^2 + (ELE_{train} - ELE_{unk})^2) / 1000$$

COV_CODE_2: Second most likely cover type assigned to region by supervised classification (Nearest Member of Group classifier; VIMAP).

COV_CODE_3: Third most likely cover type assigned to region by supervised classification (Nearest Member of Group classifier; VIMAP).

TREE_SIZE_1: Most likely tree size assigned to region by supervised classification (Nearest Member of Group classifier; VIMAP).

TREE_PROB_1: Euclidean distance that led to the tree size assigned by the supervised classification (TREE_SIZE_1); this is the smallest Euclidean distance that was calculated between attributes for this region and those for any region in the training data set. Assigned through supervised classification based on Nearest Member of Group classifier. Value has been rescaled by dividing by 1000:

$$((TM1_{train} - TM1_{unk})^2 + (TM2_{train} - TM2_{unk})^2 + ...(TM7_{train} - TM7_{unk})^2 + (ELE_{train} - ELE_{unk})^2) / 1000$$

TREE_SIZE_2: Second most likely tree size assigned to region by supervised classification (Nearest Member of Group classifier; VIMAP).

TREE_SIZE_3: Third most likely tree size assigned to region by supervised classification (Nearest Member of Group classifier; VIMAP).

MNDVI: Modified normalized difference vegetation index; calculated in ARC/INFO according to the following equation modified from Nemani et al. (1993):

$$(TM4 - TM3) / (TM4 + TM3 + 1) * (256 / (TM5 + 1)) * 100$$

TM1: Mean spectral value of region (TM channel 1); value calculated through overlay of region boundaries and 30 m pixels for TM channel 1 (ZONALSTATS).

TM2: Mean spectral value of region (TM channel 2); calculated as for TM1.

TM3: Mean spectral value of region (TM channel 3); calculated as for TM1.

TM4: Mean spectral value of region (TM channel 4); calculated as for TM1.

TM5: Mean spectral value of region (TM channel 5); calculated as for TM1.

TM6: Mean spectral value of region (TM channel 6); calculated as for TM1. If null for all regions, data for TM channel 6 were not available for this scene.

TM7: Mean spectral value of region (TM channel 7); calculated as for TM1.

PERIMETER: Perimeter of region (meters); value automatically calculated by ARC/INFO when a temporary conversion from raster to vector format was performed.

HECTARES: Area for the region in hectares (COUNT * 0.09).

X-COORD: X-coordinate of region center (in meters defined by Albers Equal Area projection); corresponds to polygon label point automatically assigned by ARC/INFO when a temporary conversion from raster to vector format was performed.

Y-COORD: Y-coordinate of region center (in meters defined by Albers Equal Area projection); corresponds to polygon label point automatically assigned by ARC/INFO when a temporary conversion from raster to vector format was performed.

SCENEPOLY_ID: Unique region identifier for entire project area. Generated by multiplying the scene number (see below) by 1 million and adding the VALUE item. Required for clipping, and should not be modified.

19 P42R29

23	P41R30
24	P40R30
45	P40R29
46	P41R28
47	P41R29

RIPARIAN: Indicates whether the region meets the model for riparian zones or not. The model for riparian zones is defined by 1:100,000 hydrology grid, a change in slope > 5 degrees, and specific spectral classes that represent riparian vegetation reflectance. The codes are: 0 = upland, 1 = potential riparian area.

VEG_LIMIT: Geographic or ecological limit zone the region is in. If zero, the scene had no geographic or ecological limit zones defined by the forests.

P40R29	1-7
P40R30	0
P41R28	1-4
P41R29	1-7
P41R30	1-2
P42R29	1-2

FIRE_LIMIT: Indicates whether the region is within a fire boundary area supplied by the Boise and Payette National Forests, 0 = outside fire boundary, 1 = inside fire boundary.

VEGCHECK: Defines how covertype field was determined. 1 = cov_code_1, 2 = cov_code_2, 3 = cov_code_3, 4 = default conifer code, 5 = manually modified or classified.

SIZECHECK: Defines how treesize field was determined. 1 = tree_size_1, 2 = tree_size_2, 3 = tree_size_3.

KEEP: Integer value for each region indicating whether or not the region will be kept when adjacent files are merged according to the currently defined edge-matching scheme for the project area. A value of 1 indicates that the region will be kept, at least in part (regions along edges may be subdivided when files are merged); 0 indicates that the region will be entirely dropped. As with COVERTYPE and similar attributes, the KEEP item has only been populated for regions that fall at least partly within the project area boundary.

Note that if a new grid is created based solely on the KEEP item (i.e., OUTPUT = P41R29Z.KEEP), there may appear to be holes (areas where KEEP = 0 entirely surrounded by KEEP = 1) well within the edge-matched boundary on edges where the scene is subordinate. These holes are an artifact of the manner in which KEEP was created; if even a part of a region will be kept in the edge-matching process, the entire region will be coded KEEP = 1.

Where large polygons from the dominant scene extend into the subordinate scene, some configurations of subordinate polygons will be such that a subordinate polygon (S1) closer to the edge spans some portion of a dominant polygon (D). S1 will be coded KEEP = 1; when displayed, S1 will essentially cut off any subordinate polygon farther from the edge (S2) where KEEP = 0 because it is entirely within D. If completely surrounded by regions like S1, S2 will appear to be a hole. When the two scenes are merged into one, S2 will be entirely overwritten by D, as will the extra portion of S1. Thus, KEEP is used by the clipper AML for edge-matching, based on a predetermined dominance order among TM scenes (see below). KEEP also provides some extra information along scene borders (e.g., the part of S1 that will later be overwritten) and, when mapped out, will not present an exact geographic match with the specific area to be kept per scene. Finally, KEEP is not intended for single-scene operations.

SCENE DOMINANCE ORDER: To be used when merging adjacent scenes together based on KEEP in conjunction with other attributes.

1	P41R29
2	P41R30
3	P41R28
4	P40R29
5	P42R29

Table B-1. Items as defined in the value attribute tables (VATs) for the land cover databases (ARC/INFO grid format) created for the 6 Landsat TM scenes covering the Boise, Payette, Sawtooth, Salmon and Challis National Forests.

ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC	Units	Valid Range
VALUE	4	10	B	-		-
COUNT	4	10	B	-		-
LIFEFORM	1	2	C	-		A,B,G,R,S,T,U,W,N
COVERTYPE	2	5	B	-		0,(≥1000, <10,000),
TREESIZE	2	4	B	-		0-4

CANOPYCOVER	2	4	B	-		0,2,3,4
BURN_INTENSITY	2	4	B	-		0-3
ELE	2	4	B	-	m	0 - 5000
SLP	4	4	F	0	degs	0 - 90.0
ASP	2	2	B	-		0 - 8
SPECTRAL_CLASS	2	4	B	-		(Alternate Name = LINK)
COV_CODE_1	2	5	B	-		(≥1000, <10,000)
COV_PROB_1	4	8	F	5		-
COV_CODE_2	2	5	B	-		(≥1000, <10,000)
COV_CODE_3	2	5	B	-		(≥1000, <10,000)
TREE_SIZE_1	2	2	B	-		0-4
TREE_PROB_1	4	8	F	5		-
TREE_SIZE_2	2	2	B	-		0-4
TREE_SIZE_3	2	2	B	-		0-4
MNDVI	2	6	B	-		-
TM1	2	4	B	-		0 - 255
TM2	2	4	B	-		0 - 255
TM3	2	4	B	-		0 - 255
TM4	2	4	B	-		0 - 255
TM5	2	4	B	-		0 - 255
TM6	2	4	B	-		0 - 255
TM7	2	4	B	-		0 - 255
PERIMETER	4	10	B	-	m	cellsize**2
HECTARES	4	12	F	2	ha	0-10000
X-COORD	4	10	F	1		-
Y-COORD	4	10	F	1		-
SCENEPOLY_ID	4	10	B	-		sceneid *1E6 + VALUE
RIPARIAN	2	4	B	-		0,1
VEG_LIMIT	2	4	B	-		based on scene
FIRE_LIMIT	2	4	B	-		0,1
VEGCHECK	2	4	B	-		0-5
SIZECHECK	2	4	B	-		0-3
KEEP	2	2	B	-		0,1

TM SCENE P40/R29

Image Analyst: Michele Thornton

Training Data Analysis

A total of 4,345 ground-truth plots were available for labeling P40/R29. The following classes were manually labeled in the scene: urban, agriculture, water, cloud, and cloud shadow. Plots with these VEG CODEs were not used in the digital classification as either training or test data. Overlap plots from two adjacent TM scenes were included as training data. A total of 208 plots, all representing Lodgepole Pine (4203), were added from TM scene P41/R29. TM scene P41/R28 provided an additional 623 training plots; these included 164 Ponderosa Pine (4206) plots, as well as 47 plots of Douglas-fir/Ponderosa Pine (4230). Ponderosa Pine and Lodgepole (4203) training plots were somewhat under-represented in this scene before the addition of the overlap plots.

Due to a lack of training plots, numerous of sub-codes were combined to a parent code or eliminated for the purposes of the supervised classification. These combinations were discussed with and approved by Lynn Bennett of the Salmon/Challis National Forest. The following codes were either dropped or combined:

3111 and 3121 were combined with code 3110;
3210 was dropped from the classification;
3304 was dropped from the classification;
3312 was dropped from the classification;
4204 was combined with 4260;
4208 was combined with 4270;
4214 was dropped from the classification;
4290 was dropped from the classification;
4400 was dropped from the classification;
4500 was dropped from the classification;
6112 and 6111 were combined with code 6110;
6215 and 6216 were combined with code 6210; and
6311 and 6312 were combined with code 6310.

Additional plots were removed by identifying duplicate plots that occurred within the same raster polygon. 'Good Duplicates' were multiple plots falling in the same polygon but assigned the same Cover Type code; for these, one plot per polygon was kept. 'Bad Duplicates' were multiple plots falling in the same polygon but assigned different cover type codes. A number of these were viewed by the image analyst to determine which code was most appropriate to keep. If an appropriate determination could not be ascertained, all 'Bad Duplicate' plots were dropped for that raster polygon.

Total Plots in Ground Truth File	3,853		
Plots from P40/R29		3,022	
Plots from P41/R28		623	
Plots from P41/R29		208	
Plots removed or Recoded			
'Good Duplicates'			412
'Bad Duplicates'			414
Eliminated Codes			499
Final Number of Training Plots	2,528		

Plots underwent further spectral examination after removal of duplicate and manually identified plots. Outliers were identified for each cover type by examining plots in relation to TM and ancillary data for their respective polygon (both visually and in relation to calculated standard deviations). As part of the bootstrap analysis, a considerable amount of time was spent visually examining plots that were classified as a different life-form. When these plots appeared to be incorrectly located, they were either eliminated or moved to a more appropriate polygon nearby.

Land Cover Types

Land cover types were labeled through a digital classification of seven TM channels (1,2,3,4,5,6 & 7) and elevation using a Nearest Member of Group classifier. The classification produced three intermediate cover type labels in descending order of likelihood: COV_CODE_1 was the most likely cover type based on the training data provided for the digital classification; COV_CODE_2 was the second most likely cover type, followed by COV_CODE_3. For cover types not subjected to additional modification (see below), COV_CODE_1 values were used and assigned directly to the COVER_CODE field in the database attribute table. Modifications were carried out using spectral class values (i.e., LINK) and/or intermediate cover code labels in conjunction with other attributes, including elevation, slope, or upland/riparian position.

Manual modifications were as follows:

1100 Urban	Manually identified by Willard Gustafson
2010 Non-irrigated Agriculture	Manually identified by Willard Gustafson
2020 Irrigated Agriculture	Manually identified by Willard Gustafson
5000 Water	Water was selected by the LINK values equal to 1, 3, 25, 68, and 71, and SLOPE less than or equal to 5.

The following upland codes were recoded to riparian codes if they occurred within the riparian buffer zone:

3100 (Upland Grassland)	—>	6210 (Graminoid/Forb Dominated Riparian)
3110 (Altered Herbaceous)	—>	6210 (Graminoid/Forb Dominated Riparian)
4203-4270 (Upland Conifer)	—>	6110 (Conifer Dominated Riparian)
4300 (Mixed Tree Upland)	—>	6130 (Mixed Tree Riparian)

Frequency Analysis

The table below summarizes the coverage extent for each cover type defined for P40/R29.

Cover Type	Class Name	Percent	Area (ha)
1100	Urban or Developed Lands	0.04	1,436
2010	Agriculture-Dry	0.66	22,293
2020	Agriculture-Irrigated	2.26	76,684
3100	Upland Grasslands	8.03	272,815
3110	Altered Herbaceous Grasslands	2.43	82,494
3180	Mesic Montane Parklands & Subalpine Meadows	3.18	107,907
3301	Mountain Mahogany	6.47	219,921
3318	Shadscale	2.69	91,478
3350	Big Sagebrush Steppe	30.40	1,031,375
4101	Aspen	0.49	16,546

4200	Conifer Forest	0.067	2,272
4203	Lodgepole Pine	6.03	204,841
4205	Limber Pine	0.76	25,956
4206	Ponderosa Pine	1.08	36,815
4212	Douglas-fir	10.60	360,564
4223	Douglas-fir/Lodgepole Pine	4.53	154,048
4230	Douglas-fir/Ponderosa Pine	5.32	18,066
4260	Mixed Whitebark Pine Forest	2.29	77,747
4270	Mixed Subalpine Forest	4.36	148,165
4300	Mixed Broadleaf and Conifer Forest	1.75	59,596
5000	Water	0.18	6,067
6110	Conifer Dominated Riparian	0.95	32,174
6120	Broadleaf Dominated Riparian	0.40	13,475
6130	Mixed Tree Riparian	0.25	8,348
6210	Graminoid & Forb Dominated	0.96	32,762
6310	Shrub-Dominated Riparian	0.91	31,042
7300	Rock-Dominated Sites	5.23	177,541
7800	Mixed Barren Sites	2.51	85,347
	TOTAL	100.00	3,397,788

Canopy Cover Classes

Canopy cover labels (CANOPY_CODE = low (2 = 10-40%), medium (3 = 41-70%), and high (4 = >70%)) were assigned to tree and shrub classes based on M_NDVI values. These values were plotted as frequency histograms for all forest types and for upland shrub types. The histograms were examined to determine break points based upon the distribution modes. Regions with M_NDVI values falling below the lower break point were assigned to a low canopy cover class; those greater than or equal to the lower breakpoint, but below the higher break point were assigned to the medium class; and those falling above the higher breakpoint were assigned to the high canopy cover class (see below).

Canopy Stratification Type	*Breakpoints*	
	Lower	Upper
3210 - 3356 Upland Shrubs	47	85
4101 - 4300 Tree	150	235

Size Classes

Four live size/structure classes were classified for conifer forest types:

1 = 1 - 4.9 inches DBH

2 = 5 - 12.0 inches DBH

3 = 12.1 - 20.0 inches DBH

4 = > 20.0 inches DBH

The seven TM spectral bands were used for the size/structure classification. The Nearest Member of Group classifier was used. Tree size classes were stratified based on the three canopy cover classes described above.

Accuracy Analysis

Accuracy was determined by randomly selecting 20% of the plots for each cover type classified, and holding these out of the Nearest Member of Group Classification routine. These training plots then were used in a fuzzy set analysis of accuracy (Gopal and Woodcock 1994). A fuzzy code was assigned to each cell of a matrix from two-way tabulation of cover types. Each cell on the diagonal was always assigned a score of 5 to represent an absolutely correct classification (perfect); score 4 indicates an accurate but not exact classification (good); score 3 indicates acceptable results (OK); score 2 indicates an unacceptable, but understandable misclassification; and score 1 indicates that the classification is absolutely wrong.

Urban, agriculture and water were not included in the assessment because sample data for these cover types often represented unusual conditions. These plots were thus not used for training and were not considered suitable for accuracy testing. The manually labeled cover types should have better than 80% accuracy.

Match and accumulate matrices for producer and user accuracies were constructed based on score distributions of each cover type. At the acceptable level (score 3) and perfect level (score 5), overall weighted producer's accuracies are as follows:

Classification for P40/R29	Number of Classes	Score 3 (Acceptable)	Score 5 (Perfect)
	28	82.14%	57.14%

Similar accuracy assessments were performed on canopy coverage and tree size class for conifer forest types. The accumulate matrix for producer for tree size class for conifer types resulted in 66.28% agreement at level 5 (perfect) and 86.80% agreement for level 3 (acceptable). The accumulate matrix for producer for canopy coverage for tree species resulted in 40.09% agreement at level 5 (perfect) and 84.05% agreement at level 3 (acceptable).

A life form accuracy assessment was also performed in which cover types were grouped into basic life forms: tree, shrub, grass, barren, and rock. The results, at level 5, are as follows:

Life Form	Tree	Shrub	Grass	Barren	Rock
Accuracy	96.03%	77.12%	56.34%	22.22%	63.16%

Discussion

P40/R29 went through three iterations of classification. An initial classification was performed in December of 1996. Eight different 7.5 minute quadrangles were selected by the Forest Service; maps of both cover type and size class/canopy closure class were plotted for each quadrangle and sent to the Forest Service for review. Polygons were identified as correct or, if incorrect, were labeled with the proper cover type code by Forest Service personnel. Quadrangles reviewed for this scene included:

Big Windy Peak
Bohannon Spring
Double Spring
Gold Stone Mountain
Horse Basin
Morrison Lake
Powderhorn Gulch
Warren Mountain

These additional training data were digitized into the plot coverage from the quad maps by Wildlife Spatial Analysis Lab staff. These plots can be identified in the point coverage as MR1 in the PLOTID item. An additional 887 plots were added to the training data by this means. A second classification, with these additional plots, was performed in May of 1997 and reviewed by Cliff Keene of the Salmon/Challis National Forest. Additional data and comments from Cliff were incorporated into the third and final classification (June 1997).

Certain cover types were limited geographically during the supervised classification. Limitations were defined by Salmon/Challis National Forest personnel. Geographic limits were incorporated by using polygons that covered large spatial areas within the scene (Figure C1). Geographic limits were provided for the following cover types: 4206 (Ponderosa Pine) and 4230 (Douglas-fir/Ponderosa Pine) were limited to Zone 1. 3318 (Shadscale) was excluded from Zone 1. 4205 (Limber Pine) was limited to Zones 3, 4, and 5. 3354 (Black Sagebrush) was excluded from Zones 1 and 2. 4101 (Aspen) was limited to polygons smaller than 20 acres within Zone 1. After the final classification was performed, cover type codes for conifer types were further evaluated using an Arc/Info AML that checked for ecological limits for specified cover types. The ecological limits were provided by the Salmon/Challis National Forest. This resulted in 20,038 cover type changes. Essentially, if COV_CODE_1 did not meet the specified ecological limit, COV_CODE_2 was selected. If COV_CODE_2 did not meet the ecological limit, COV_CODE_3 was selected. Finally, if COV_CODE_3 did not meet the limit, then a Cover type of 4200 (Conifer) was assigned. Overall, in the final accuracy assessment, the Ecolimit AML increased level 5 accuracy slightly from 56.75% correct to 57.14% correct.

Overall, considerable effort was spent to improve the spectral differentiation between forest cover types and to improve the locations of training data. Examination of plot data indicates that a certain amount of confusion remains among forest cover types, especially conifers. Upland xeric shrubs and upland grass types show similar confusion. Other noted confusion exists among Aspen (4101) which tended to over-classify into recently cut or deciduous shrub polygons. Mountain Mahogany (3301) tended to classify areas where it was present, but also tended to classify into areas that may have a strong Mountain Mahogany understory such as within Douglas-fir stands. Training plots for 3301 were often found near and/or associated with Douglas-fir (4212) plots, resulting in some confusion between the two. Given both the spatial and spectral limitation of TM data, it may not be possible to further reduce the remaining confusion.

Error Matrix for Upland and Riparian Cover Type Training Plots

Scene P40/R29

COVER TYPES	3100	3110	3180	3301	3318	3350	4101	4203	4205	4206	4212	4223	4230	4260	4270	4300	6110	6120	6130	6210	6310	7300	7800	SUM
3100	68	0	7	15	1	48	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1	4	147
3110	0	54	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	60
3180	4	0	41	3	0	7	1	0	0	0	0	0	0	0	0	2	0	0	2	3	0	3	4	70
3301	16	2	1	64	0	21	0	2	5	11	11	2	0	2	0	2	0	2	0	1	0	1	4	147
3318	0	0	0	0	7	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	15
3350	36	14	7	18	7	396	0	0	0	2	1	0	0	0	0	0	0	0	0	1	4	2	7	495
4101	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	1	0	2	4	2	5	0	0	18
4203	0	0	0	1	0	0	0	118	1	0	43	23	1	5	12	1	2	0	0	0	0	0	0	207
4205	0	0	0	4	0	0	0	0	22	0	13	1	0	6	5	1	0	0	0	0	0	0	0	52
4206	1	0	0	7	0	1	0	0	0	86	7	0	3	0	0	0	0	1	0	2	0	0	0	108
4212	0	0	0	5	0	0	0	53	16	9	298	27	5	4	18	4	2	1	0	0	0	0	0	442
4223	0	0	0	3	0	0	0	35	2	0	32	52	0	1	13	1	3	0	0	0	1	0	0	143
4230	0	0	0	0	0	0	0	0	0	6	9	0	7	0	0	1	0	1	0	0	0	0	0	24
4260	0	0	1	3	0	0	0	5	4	0	4	1	0	16	3	1	0	0	0	0	0	0	0	38
4270	0	0	0	1	0	0	0	18	4	0	15	17	0	4	53	0	0	0	0	0	0	0	0	112
4300	0	0	2	2	0	1	0	1	0	0	2	3	0	0	0	4	1	3	2	0	3	0	1	25
6110	0	0	0	0	0	0	0	1	0	0	7	3	0	0	0	2	0	1	1	1	0	0	0	16
6120	0	0	0	2	0	0	3	0	0	1	0	0	0	0	0	2	1	30	5	3	11	0	0	58
6130	0	0	0	0	0	0	3	0	0	1	0	0	0	0	0	1	1	4	4	1	2	0	0	17
6210	1	0	2	0	0	2	4	0	0	2	0	0	0	0	0	1	2	5	3	38	21	0	1	82
6310	0	0	1	0	0	0	9	0	0	0	0	3	0	0	0	3	0	11	7	22	52	0	0	108
7300	0	0	2	2	1	4	0	0	3	0	0	0	0	1	0	0	0	0	0	1	0	71	12	97
7800	7	0	5	7	0	8	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	6	11	47
SUM	133	70	69	137	16	501	24	233	58	118	444	132	16	39	104	27	12	61	28	78	99	85	44	2528

PERCENTAGE AGREEMENT IS 59.18
TAU WITH EQUAL PROBABILITY IS 5734
TOTAL DIAGONAL ELEMENTS ARE 1496
TOTAL POINTS OF DATA ARE 2528

TM P40/R29

Fuzzy Set Accuracy Assessment. Upland and Riparian Cover Types

Accumulate Matrix (producers)

GROUP	5	5(%)	4	4(%)	3	3(%)	2	2(%)	1	1(%)
3100	10	34.48	10	34.48	11	37.93	11	37.93	29	100.00
3110	11	91.67	11	91.67	11	91.67	11	91.67	12	100.00
3180	10	71.43	10	71.43	10	71.43	10	71.43	14	100.00
3301	12	41.38	12	41.38	20	68.97	20	68.97	29	100.00
3318	2	66.67	3	100.00	3	100.00	3	100.00	3	100.00
3350	77	77.78	79	79.80	86	86.87	86	86.87	99	100.00
4101	2	50.00	2	50.00	2	50.00	2	50.00	4	100.00
4203	18	43.90	33	80.49	40	97.56	40	97.56	41	100.00
4205	1	10.00	1	10.00	9	90.00	9	90.00	10	100.00
4206	20	90.91	20	90.91	22	100.00	22	100.00	22	100.00
4212	62	70.45	68	77.27	86	97.73	86	97.73	88	100.00
4223	7	24.14	22	75.86	29	100.00	29	100.00	29	100.00
4230	1	20.00	4	80.00	5	100.00	5	100.00	5	100.00
4260	3	37.50	3	37.50	7	87.50	7	87.50	8	100.00
4270	10	45.45	13	59.09	22	100.00	22	100.00	22	100.00
4300	1	20.00	1	20.00	3	60.00	3	60.00	5	100.00
6110	2	66.67	2	66.67	2	66.67	2	66.67	3	100.00
6120	10	66.67	10	66.67	10	66.67	10	66.67	15	100.00
6210	6	37.50	6	37.50	6	37.50	11	68.75	16	100.00
6310	9	40.91	9	40.91	9	40.91	15	68.18	22	100.00
7300	12	63.16	12	63.16	17	89.47	17	89.47	19	100.00
7800	2	22.22	2	22.22	4	44.44	5	55.56	9	100.00
SUM	288	57.14	333	66.07	414	82.14	426	84.52	504	100.00