

# FEATURES

## A Brief Overview of the Southwest Regional GAP Land Cover Mapping Effort

John Lowry

Remote Sensing/GIS Laboratory, College of Natural Resources,  
Utah State University, Logan, Utah

### Introduction and Project Area

The Southwest Regional Gap Analysis Project (SWReGAP) was initiated in 1999 as a multi-institutional cooperative effort to map and assess biodiversity for a five-state region (Arizona, Colorado, New Mexico, Nevada, and Utah). This area comprises approximately 150 million hectares (560,000 square miles), representing approximately one-fifth of the coterminous United States. A key task in this effort was to develop a seamless land cover map for the region. The five-state region was divided into 20 ecologically and spectrally similar mapping zones. Each mapping zone provided a functional working area for project management, data collection, and modeling. Each state was responsible for the mapping zones roughly corresponding to their state jurisdiction (Figure 1).

### Methods

#### Data Preparation

Landsat 7 enhanced thematic mapper plus (ETM+) images were selected from 1999–2001 for three seasons: spring, summer, and fall. Scenes were selected for optimal representation of seasonal phenology and minimal cloud cover. Landsat scenes

were standardized using a dark object subtraction method and mosaicked for each mapping area. Image transformations such as brightness, greenness, and wetness bands were created for each image mosaic. Digital elevation data, provided by the National Elevation Data Set (1999), were a subset for each mapping zone, as were subsequent digital elevation derivatives, such as aspect and landform. Each mapping zone had a 2 km overlap with the adjacent mapping area, providing an overall 4 km overlap region between modeling areas.

#### Training Sample Collection

Approximately 93,000 samples were available for the five-state region (Figure 2). The majority of samples were collected through field surveys conducted between 2001 and 2003. Field surveys involved recording ocular estimates of biotic characteristics (percent cover of dominant species for trees, shrubs, grasses, and forbs) and physical characteristics (elevation, slope, aspect, and landform). The location of each sample site was recorded with a global positioning system (GPS) reading and a polygon digitized using a laptop computer with thematic mapper (TM) imagery as a backdrop. In addition, two digital photographs were taken at each sample location. Sampling involved traversing all navigable roads in a mapping zone and opportunistically selecting samples based on appropriate size and composition (i.e., representative) of stands. Additional samples, obtained from other projects, imagery, or



Figure 1. Five-state region divided into mapping zones.

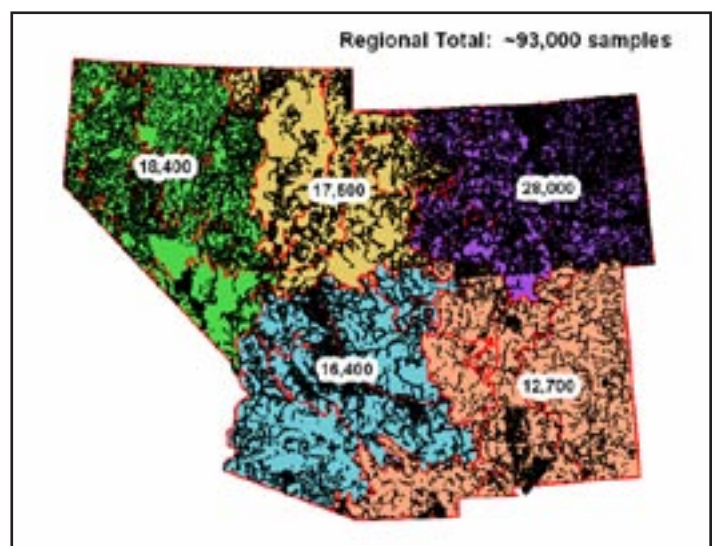


Figure 2. Approximately 93,000 training samples collected from various sources.

aerial photo interpretation, were also used, though these were in the minority.

Thematic Mapping Legend

The focus of the mapping effort was on natural and seminatural systems. The basic thematic mapping unit was the *ecological system* concept developed by NatureServe (Comer et al. 2003). Ecological systems represent recurring groups of biological communities that are found in similar physical environments and are influenced by similar dynamic ecological processes. They are intended to provide a thematic mapping unit mappable at a meso-scale level from remotely sensed imagery. Each sample site was assigned an appropriate land cover label in the database prior to the modeling process.

Land Cover Modeling

The majority of natural and seminatural land cover classes were modeled using a decision-tree (DT) classifier. DT classifiers are becoming a common approach used for land cover mapping (Lawrence and Wright 2001; Pal and Mather 2003; Brown de Colstoun et al. 2003). Advantages of DT include the ability to use both continuous and categorical predictor data sets with different measurement scales, good computational efficiency, and an intuitive hierarchical representation of discrimination rules. A major technical challenge in the past has been that of spatially

applying the decision-tree rules generated by the DT software within a geographic information system.

After experimenting with the development of several approaches, the project used See5/C5.0 (Rulequest Research 2004) for the DT classifier and ERDAS Imagine for spatially applying the DT-generated rules. The integration of these software systems was greatly facilitated by the use of a customized interface for ERDAS Imagine developed under contract by Earth Satellite Corporation for the U.S. Geological Survey Eros Data Center (Figure 3). Where the decision tree could not be used, other techniques, such as localized unsupervised clustering or screen digitizing, were used to map a minority of cover classes.

**Results**

Model Validation

DT models were validated by generating initial models using 80 percent of available samples, while withholding 20 percent of samples. Withheld samples were randomly selected and stratified by cover class. Withheld sample polygons were intersected through the land cover map to create an error matrix, presenting users, producers, and overall “accuracies.” The kappa statistic was also calculated for the error matrix. This validation process was performed on each of the 20 mapping

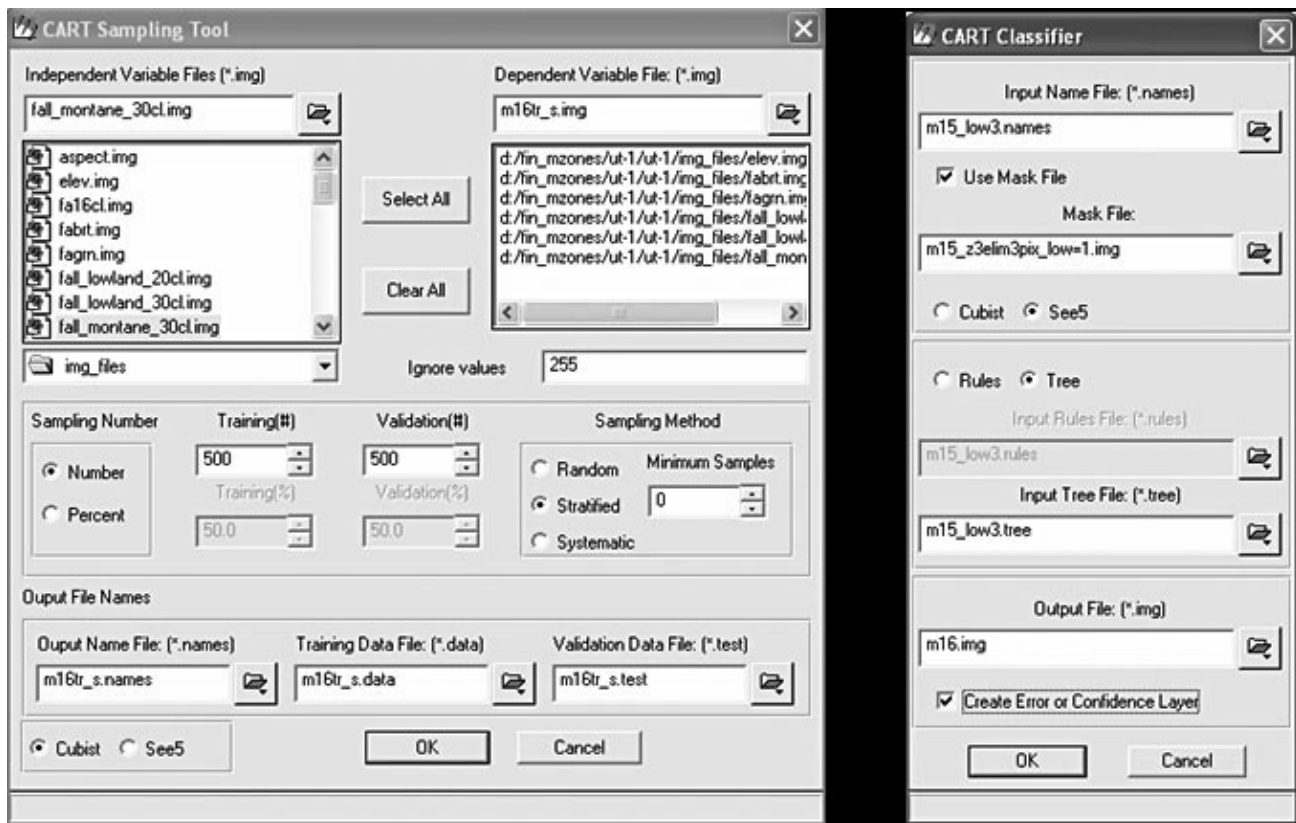


Figure 3. ERDAS Imagine custom interface for integrating Imagine with See5/C5.0.

areas for the five-state region. Overall accuracies (sum of diagonals) vary from mapping zone to mapping zone and will be presented in the final report.

#### Regional Mosaic and Data Set Delivery System

Using the 4 km overlap region between mapping zones, a “cutline” was used to edge-match adjacent mapping areas where land cover discontinuities resulted from the modeling process. The resulting five-state region mosaic was qualitatively reviewed by the five state teams and NatureServe. Following review, a limited number of errors were “flagged” for final editing. The “edits” that were determined to be relatively easy to correct with localized recoding, or a simple conditional model, were made to the regional map.

The SWReGAP land cover data set was completed in September 2004, and it is currently available to the public with “provisional” status from <<http://earth.gis.usu.edu/swgap/landcover.html>>. Because the data set encompasses such a large region, the web site allows users to download specific geographic segments of the region, such as individual states, counties, or ecoregions. Additionally, the web site offers an Internet map server from which users can interactively clip a specified rectangle in the region. The clipped data set is subsequently bundled with metadata and made available for download (Figure 4).

#### Literature Cited

Brown de Colstoun, E. C., M. H. Story, C. Thompson, K.

Commisso, T. G. Smith, and J. R. Irons. 2003. National Park vegetation mapping using multitemporal Landsat 7 data and a decision-tree classifier. *Remote Sensing of Environment* 85:316–27.

Comer, P., D. Faber-Langendoen, R. Evans, S. Gawler, C. Josse, G. Kittel, S. Menard, M. Pyne, M. Reid, K. Schulz, K. Snow, and J. Teague. 2003. *Ecological systems of the United States: A working classification of U.S. terrestrial systems*. Arlington, Va.: NatureServe.



Figure 4. Delivery system allows download by geographic subsets of the region.

Lawrence, R. L., and A. Wright. 2001. Rule-based classification systems using Classification and Regression Tree (CART) analysis. *Photogrammetric Engineering and Remote Sensing* 67:1137–42.

Pal, M., and P. M. Mather. 2003. An assessment of effectiveness of decision-tree methods for land cover classification. *Remote Sensing of Environment* 86:554–65.

Rulequest Research. 2004. See web site at <<http://www.rulequest.com/>> (last updated November 2004).