# Mapping Vegetation Communities Using Statistical Data Fusion in the Ozark National Scenic Riverways, Missouri, USA

**Robert A. Chastain Jr., Matthew A. Struckhoff, Hong S. He, and David R. Larsen**

## Abstract

*A vegetation community map was produced for the Ozark National Scenic Riverways consistent with the association level of the National Vegetation Classification System. Vegetation communities were differentiated using a large array of variables derived from remote sensing and topographic data, which were fused into independent mathematical functions using a discriminant analysis classification approach. Remote sensing data provided variables that discriminated vegetation communities based on differences in color, spectral reflectance, greenness, brightness, and texture. Topographic data facilitated differentiation of vegetation communities based on indirect gradients (e.g., landform position, slope, aspect), which relate to variations in resource and disturbance gradients. Variables derived from these data sources represent both actual and potential vegetation community patterns on the landscape. A hybrid combination of statistical and photointerpretation methods was used to obtain an overall accuracy of 63 percent for a map with 49 vegetation community and land-cover classes, and 78 percent for a 33-class map of the study area.*

## Introduction

Vegetation communities within Ozark National Scenic Riverways (ONSR) were mapped at both the Association level of the National Vegetation Classification System (NVCS; FGDC, 1996), and using an ecological systems approach (NatureServe, 2006) that aggregates similar and ecologically related associations. The mapping approach combined information gleaned from aerial photo interpretation, results from statistical classifications using remote sensing and topographic data, extensive and intensive field data, and an ecological classification system model to develop an automated vegetation classification system of the natural and altered vegetation community types present in the mapping area. This study was conducted in cooperation with the U.S. Geological Survey (USGS), National Park Service (NPS) Vegetation Mapping Program, a cooperative effort to classify, describe, and map vegetation communities in national park units across the United States (USGS–NPS, 2006). The USGS-NPS Vegetation Mapping Program specifies an overall and class-wise standard of 80 percent.

Many variables derived from remote sensing or ancillary data sources are typically hypothesized to serve as potentially useful inputs for a statistical classification model for vegetation community mapping, but often it is difficult to know *a priori* which variables to incorporate into the model. For example, multispectral remote sensing data captured during different seasons is frequently obtained to assist the discernment of different vegetation communities by capturing a range of phenological conditions on the ground. In addition, various data transforms can add to the information content of these multispectral image data. Frequently, high spatial resolution remote sensing data is obtained to exploit textural information in these data (Carleer and Wolff, 2005; Ivits *et al.*, 2005; Trietz and Howarth, 2000; Wulder *et al.*, 2004). However, spatial scales of remote sensing data that are relevant to the land-cover features of interest must be identified and utilized to achieve good classification results. In addition, information contained in digital elevation model (DEM) data is often used to model the indirect resource gradients related to topographic position. Choosing the most valuable input data to produce a parsimonious statistical classification model while taking advantage of the dimensionality of a large input variable set derived from combined data sources can prove to be a difficult balance. The value of the multi-dimensional input variable set must be retained while also minimizing degrees of freedom to produce a tractable classification model that performs optimally.

The primary goal addressed in this research is to develop a method to optimally exploit an information-rich array of field survey, remote sensing, and topographic datasets to produce a detailed vegetation community map. Multitemporal Landsat TM and ETM+ image data collected during leaf-on and leaf-off conditions, high-resolution softcopy color-infrared (CIR) photo image data, and 10-meter resolution DEM data were utilized in this research. Numerous

Robert A. Chastain Jr. is at the Fort Lewis Directorate of Public Works, Environmental and Natural Resources Division IMNW-LEW-PWE MS17, Box 339500, Bldg. 2012, Fort Lewis, WA 98433, and formerly with the University of Missouri-Columbia Department of Forestry, Columbia, MO 65211 (robert.chastain2@us.army.mil).

Matthew A. Struckhoff is with the U.S. Geological Survey, Columbia Environmental Research Center, 4200 New Haven Road, Columbia, MO 65201.

Hong S. He and David R. Larsen are with the University of Missouri-Columbia, Department of Forestry, 203 Anheuser Bush Natural Resources Building, Columbia, MO 65211.

Photogrammetric Engineering & Remote Sensing
Vol. 74, No. 2, February 2008, pp. 247–264.

0099-1112/08/7402–0000/$3.00/0

data transformations were performed to fully exploit the information content of the multi-scaled remote sensing as well as DEM data to produce input variables for a statistical classification model. For example, the spectral dimensionality of the multitemporal Landsat data was utilized by deriving various transformations, as was the textural information contained in the high-resolution CIR aerial photogrammetric image data, and a number of indices relating to direct and indirect gradients relevant to plant species niche differentiation were derived from the DEM data. A large set of statistical classification input variables was derived based on past experience regarding which variables have yielded successful results in vegetation community classifications using digital image data, with the intent that stepwise analysis would the indicate the most germane variables for inclusion into subsequent analysis.

Statistical classification approaches that are typically employed in remote sensing studies proved inadequate in a pilot investigation, so a novel approach was developed to use the large available dataset to its highest potential. The approach used in this research involved mining the available data with the intent of utilizing information relevant to (a) species-level niche differentiation, (b) past land-use history, (c) patterns of current land-use, and (d) differential responses to disturbance, as all of these factors influence patterns of current vegetation communities present in the ONSR mapping region. Discriminant analysis was used as a novel statistical approach to choose the most relevant classification inputs among nearly 100 variables derived from remote sensing and DEM data, then mathematically fuse the numerous input variables into independent canonical functions to map NVCS vegetation associations and NatureServe community types based on per-pixel probabilities of class membership.

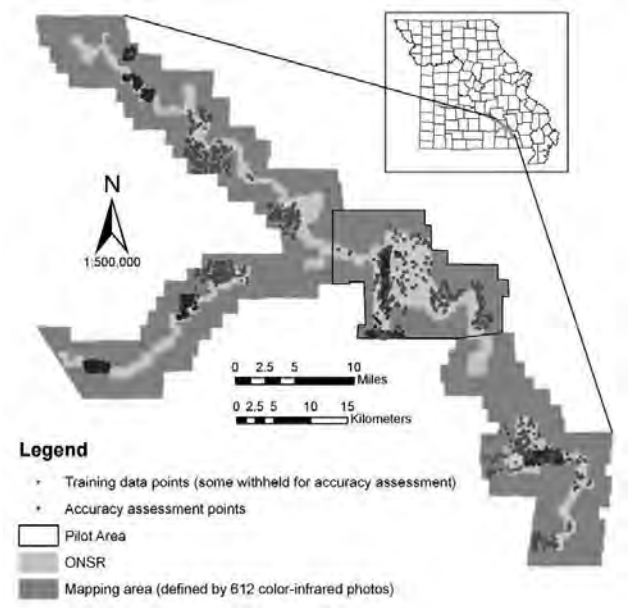


Figure 1. Study area map indicating mapping area and national park boundaries as well as the distribution of field sampling points used to train the statistical classification and validate the vegetation community map. The location of a pilot area within which various statistical classification approaches were tested is also shown, and the map extents shown in Figures 5 and 6 are located within the pilot area polygon.

### Study Area

The Ozark National Scenic Riverways (ONSR), which was established in 1964 from public and private lands with a variety of uses, encompasses approximately 33,257 ha along the Current and Jacks Fork Rivers in southeastern Missouri. It is located in the Current River Hills Subsection of the Ozark Highlands Section of the national hierarchical framework of ecological units (Avers *et al.*, 1994). The mapping region (UTM Zone 15 North, NAD83; 611300, 4074000 by 691400, 4149400 or 36.79N, 91.753W by 37.484N, 90.841W) is 141,854 ha (350,529 acres) and encompasses ONSR as well as areas immediately surrounding the parkland (Figure 1). The Ozark Highlands are perhaps the oldest continuously exposed landmass in North America. Because the region was not glaciated during any of the last four major continental glaciation events, plant life has persisted here for the last 100 million years, resulting in extreme biological diversity of plant communities and high levels of endemism (The Nature Conservancy, 2003). Because most of the area was logged between 1880 and 1920 (Cunningham and Hauser, 1992), the landscape is now dominated by second-growth forest. Pre-settlement vegetation was characterized by oak and pine woodlands and forests heavily influenced by aboriginal and natural fires (Guyette and Cutter, 1991), but fire regime changes in the second growth forests have resulted in a smaller pine component, and uniform, younger forests have replaced the woodland/forest mosaic.

### Classification Structure

The NVCS uses a seven level hierarchical scheme for the classification of natural and semi-natural terrestrial vegetation (Table 1). The first five levels describe the physiognomic characteristics, and the last two levels describe the floristics. The ONSR mapping area includes relatively intact

Table 1. Details Associated with the NVCS Classification Structure. Names of the Associations Identified and Mapped Within the ONSR Study Area can be Seen in Tables 6 and 7

| Level | Primary Basis For Classification | Example |
|---|---|---|
| Class | Growth form and structure of vegetation | Forest |
| Subclass | Growth form characteristics (e.g., leaf phenology) | Deciduous forest |
| Group | Leaf types, corresponding to climate | Cold-deciduous forest |
| Subgroup | Relative human impact (natural/semi-natural or cultural) | Natural/semi-natural |
| Formation | Additional physiognomic and environmental factors, including hydrology | Temporarily flooded cold-deciduous woodland |
| Alliance | Dominant/diagnostic species of uppermost or dominant stratum | *Pinus echinata* Forest Alliance |
| Association | Additional dominant/diagnostic species from any strata | *Pinus echinata/Vaccinium* (*arboreum, pallidum, stamineum*) Forest |

areas with minimal evidence of human disturbance and parcels of land that are either actively or were previously being used for grazing, logging, and/or row cropping. The result is a landscape mosaic containing both relatively natural plant communities as well as areas that have been modified through past and continuing human uses that are now in various stages of regrowth; some currently altered to such an extent that they are considered "cultural" or "early successional" types. Because the NVCS scheme focuses primarily on "natural" or "semi-natural" types, many of the most altered communities present in the ONSR mapping area are not adequately accounted for. In addition, certain vegetation communities are often spatially related to one another, though they may have little in common structurally. In order to address these issues, and to identify broader vegetation classes relevant to resource management, local "Community Types" were identified using the ecological systems approach developed by NatureServe (2006) to better categorize the some of the vegetation communities present in the mapping area. Ecological systems were identified in part due to recognized shortcomings within the NVCS when mapping communities. Consolidating like communities into broader categories (typically the NVCS alliance) in order to improve classification accuracy is an accepted process within the NVCS program. The vegetation community classification system used in this research is therefore an amalgam of the NVCS association level and NatureServe Community Types.

## Methods

A vegetation community map designed to meet the 80 percent accuracy standards required by the USGS-NPS Vegetation Mapping Program was produced using a hybrid method wherein the results of separate statistical classifications produced for different segments of the mapping area were combined using a spatial overlay process with elements digitized from current and historical high-resolution aerial photography. Digitized features corresponded for the most part to human-altered land-cover elements, and these were placed over the mapped predictions of NVCS vegetation associations obtained from statistical classifications. The logic behind the overlay order of derived mapping components was based on the manner in which anthropogenic land-cover modifications are superimposed over relatively natural vegetation communities.

The independent variables used as input data for the statistical classifications include CIR and multispectral remote sensing image data that depict the magnitude of reflectance of the vegetation communities in different electromagnetic wavelengths as well as the textural patterns of their brightness, and topographic data that represent indirect gradients spatially segregating vegetation communities (Parker, 1982; Austin and Smith, 1989; Franklin, 1995). These spatial classification inputs represent the spectral reflectivity of actual vegetation communities present on the landscape and characterize differences resulting from environmental niche specialization of component species in these communities, such as moisture and light environment differences. In this sense, these classification inputs represent both actual and potential vegetation community patterns in the ONSR mapping area.

### Data

*Classification Training Data*

A total of 1,069 field samples that served as training data for statistical classification were collected in 2003 and 2004 following NVCS guidelines for sampling large parks. Those guidelines suggest the use of a gradsect approach (Austin and Heyligers, 1989), which deliberately samples across environmental gradients known to influence vegetation. Within the study area, Nigh *et al.* (2000) had identified and modeled ecological land types (ELTs), which are land units characterized by distinct topography, parent materials, soil series, and potential vegetation associations. We applied a gradsect sampling system in which sampling area boundaries were drawn to incorporate the maximum ecological diversity, as indicated by the ELT model, and significant public land ownership. During the second season, a preliminary classification map was used to stratify sampling and identify sites with high provisional vegetation community diversity. Data collection at each point followed NVCS guidelines (The Nature Conservancy, 1994) and are described in the final technical report to the National Park Services (Chastain *et al.*, 2006). Geospatial training information collected at each sample location included identification of the NVCS association type within which the sample point was located, a ranking of its quality from 1 to 5 to describe the degree to which that community matched the description in the current NVCS, and an estimation of the extent in meters of the NVCS type in each cardinal direction from the sampling point.

Training data were extracted from 92 discriminating variables at a total of 3,237 locations (Tables 2 through 5; Figure 1). The 3,237 sample points were gleaned from 1,069 field observation points, with the additional 2,168 observations located at a distance of 50 meters in cardinal directions where field notations on the extent of the vegetation association at the primary sampling point verified spatial continuity. Data were extracted at the 3,237 sample points from raster files containing spatially continuous representations of 92 discriminating variables obtained from CIR aerial photo images, Landsat images, and topographic data to create a tabular input dataset for a discriminant analysis approach implemented using the Statistical Analysis Software (SAS; SAS Institute, Inc., 2004).

*Color Infrared Aerial Photography*

A set of 612 1:12,000 scale CIR aerial photographs were acquired over the ONSR mapping region in October 2002, at the onset of foliar senescence. They were scanned into digital form at a resolution of 0.25 m cells and orthorectified using the OrthoBASE functions contained in the Leica Photogrammetry Suite extension of the ERDAS Imagine® software, version 8.7. At least five ground control points (GCPs) were obtained for each photo frame by locating ground features that were discernable on both the October 2002 aerial photo images and corresponding National Agricultural Imagery Program (NAIP) 2-meter orthophotos serving as reference images. These GCPs were used to perform aerial triangulation between the camera and features on the images, which were then orthorectified using a 10-meter resolution DEM and a nearest neighbor resampling method to create an output raster image with 0.25 meter square pixels. Pan-ocular assessment of the positional accuracy of the orthorectified CIR photo images was performed by comparing them with NAIP orthophotos using a maximum tolerance of a 15 m spatial discrepancy between the two image sources. When larger discrepancies were found on any location on an image frame, it was re-rectified using additional GCPs. Mosaics of the individual orthorectified October 2002 photo image frames were created for use as a source of variables for statistical classification input and as a reference for photointerpretation.

Numerous data derivatives that facilitate the discrimination of vegetation associations based on differences in their color, level of greenness and brightness, and texture were obtained from the 2002 CIR photo image data to serve as

TABLE 2. DISCRIMINATING VARIABLES OBTAINED FROM THE 2002 CIR AERIAL PHOTO IMAGE DATA

| | |
|---|---|
| ir2m | Infrared wavelength grey values from 2-meter resolution image data |
| red2m[2] | Red wavelength grey values (2-meter resolution) |
| green2m | Green wavelength grey values (2-meter resolution) |
| illalb2m[2] | Illumination/albedo surface; 2-meter resolution (Sqrt($IR^2$ + $Red^2$ + $Green^2$)) |
| div112m[1,2,3] | Diversity of brightness values in the illumination/albedo surface (11 × 11 window) |
| sd112m[1,2] | Std deviation of brightness values in the illumination/albedo surface (11 × 11) |
| variance2m[1,2] | Variance of brightness values in the illumination/albedo surface (11 × 11) |
| propbright15 | Proportion of bright pixels in a 15 × 15 neighborhood (2-meter resolution) |
| propbright25[1] | Proportion of bright pixels in a 25 × 25 neighborhood (2-meter resolution) |
| sqvi2m[1] | Square root vegetation index; 2-meter resolution (Sqrt(IR/Red)) |
| ndvi2m[1,2] | Normalized difference vegetation index; 2-meter resolution (Red − IR/Red + IR) |
| pc12m | Principal component 1 (2-meter resolution) |
| pc22m[1,2] | Principal component 2 (2-meter resolution) |
| pc32m | Principal component 3 (2-meter resolution) |
| shadir2m | Infrared grey values with shadows masked out (2-meter) |
| shadred2m | Red grey values with shadows masked out (2-meter) |
| shadgreen2m | Green grey values with shadows masked out (2-meter) |
| msqvi2m | Square root vegetation index with shadows masked (2-meter resolution) |
| mndvi2m[1] | Normalized difference vegetation index with shadows masked (2-meter) |
| ir14m[2] | Infrared wavelength grey values (14-meter resolution) |
| red14m[2] | Red wavelength grey values (14-meter resolution) |
| green14m[2] | Green wavelength grey values (14-meter resolution) |
| illalb14m[1,2,3] | Illumination/albedo surface; 14-meter resolution (Sqrt($IR^2$ + $Red^2$ + $Green^2$)) |
| sqvi14m[1,3] | Square root vegetation index; 14-meter resolution (Sqrt(IR/Red)) |
| ndvi14m[1,3] | Normalized difference vegetation index; 14-meter resolution (Red − IR/Red + IR) |
| pc114m[2] | Principal component 1 (14-meter resolution) |
| pc214m[1,2,3] | Principal component 2 (14-meter resolution) |
| pc314m[2] | Principal component 3 (14-meter resolution) |
| shadir14m | Infrared grey values with shadows masked out (14-meter) |
| shadred14m | Red grey values with shadows masked out (14-meter) |
| shadgreen14m[2,3] | Green grey values with shadows masked out (14-meter) |
| msqvi14m[1,2,3] | Square root vegetation index with shadows masked (14-meter) |
| mndvi14m[1,2,3] | Normalized difference vegetation index with shadows masked (14-meter) |
| ir30m[1] | Infrared wavelength grey values (30-meter resolution) |
| red30m[1,2] | Red wavelength grey values (30-meter resolution) |
| green30m[1] | Green wavelength grey values (30-meter resolution) |
| illalb30m[1] | Illumination/albedo surface; 30-meter resolution (Sqrt($IR^2$ + $Red^2$ + $Green^2$)) |
| sqvi30m[1] | Square root vegetation index; 30-meter resolution (Sqrt(IR/Red)) |
| ndvi30m[1,2,3] | Normalized difference vegetation index; 30-meter resolution (Red − IR/Red + IR) |
| pc130m[1,2] | Principal component 1 (30-meter resolution) |
| pc230m[1] | Principal component 2 (30-meter resolution) |
| pc330m[1] | Principal component 3 (30-meter resolution) |
| shadir30m[1] | Infrared grey values with shadows masked out (30-meter) |
| shadred30m[1] | Red grey values with shadows masked out (30-meter) |
| shadgreen30m[1] | Green grey values with shadows masked out (30-meter) |
| msqvi30m[1,3] | Square root vegetation index with shadows masked (30-meter) |
| mndvi30m[1,3] | Normalized difference vegetation index with shadows masked (30-meter) |

1 = variable used in the discriminant analysis classification for the Hills and Breaks landscape strata
2 = variable used in the discriminant analysis classification for the Bottomland landscape strata
3 = variable used in the discriminant analysis classification for the Igneous Knobs landscape strata

statistical classification inputs (Table 2). Because the large file sizes of the 0.25 m resolution image mosaics precluded the extraction of derivative information, a correlation analysis was performed to determine a spatial resolution that would reduce the size of these image files to a manageable level while maintaining as much image detail as possible so that useful textural information would not be lost. Correlation analysis was performed in three test areas on illumination/albedo surfaces (calculated as the square root of the sum of the squares of the infrared, red, and green image bands from the CIR image data), because illumination/albedo accentuates total image texture in forested areas by combining illumination and overall reflectance variations in all of the image bands, highlighting boundaries of individual tree canopies and the shadows on their peripheries (Warner *et al.*, 1998). Correlograms produced from the illumination/albedo surfaces to examine scales of spatial dependence in the three test areas predict that a spatial resolution of 2-meters represented a good compromise between data size reduction and textural information retention, with a great deal (85 percent) of the spatial autocorrelation in the image remaining at that spatial scale (Figure 2).

Textural information extracted from the illumination/albedo surface derived from the 2-meter CIR image data included the standard deviation, variance, and diversity of brightness values within an 11 pixel × 11 pixel moving window, and the proportion of bright pixels calculated within 15 pixel × 15 pixel and 25 pixel × 25 pixel windows. Diversity of brightness values was calculated as the number of different grey-level values in the illumination/albedo surface summed within a moving 11 pixel × 11 pixel analysis window, and was computed using the functionality of the Leica Imagine® software. The proportion of bright pixels was calculated by first identifying a threshold that appeared from pan-occular comparison with the 2-meter CIR image data to mask out the shadowed areas, then calculating the proportion of remaining (*bright*) pixels within 11 pixel × 11 pixel and 25 pixel × 25 pixel windows.

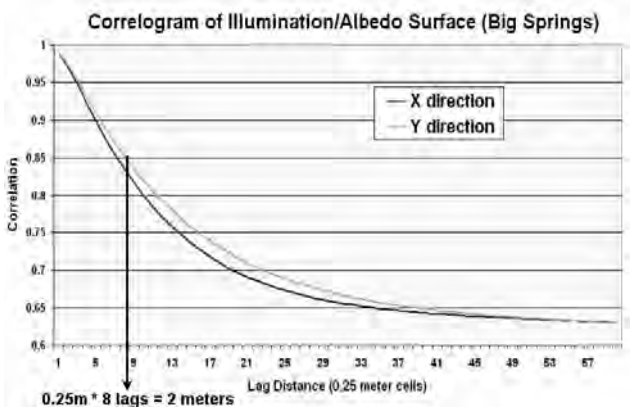


Figure 2. Correlogram of illumination/albedo surface indicating a high amount of spatial dependence at about 2-meters in both the *x* and *y* directions.

To further facilitate a statistical classification of the CIR image data, two separate data processing steps were taken in an attempt to defeat the "H-resolution problem" (after Strahler *et al.*, 1986), as small scale heterogeneity in these images resulting from tree canopy components such as sunlit leaves, shaded areas, and branches was apparent at both 0.25- and 2-meter resolution (Carleer and Wolff, 2005, Wulder *et al.*, 2004). First, the CIR photo image was degraded to 14- and 30-meter resolutions, and many of the same data transformations that were performed on the 2-meter CIR data were implemented on these degraded image data to produce classification model inputs. This was done because the H-resolution problem asserts that as the spatial resolution of a sensor increases, interclass variability also increases, thereby reducing statistical separability and classification accuracy in per-pixel classifiers (Hay *et al.*, 1996, Wang *et al.*, 2004). The 14- and 30-meter resolution scales for image degradation were chosen because they are consistent with the spatial resolutions of other remote sensing image data sources (e.g., SPOT, ASTER, Landsat), thus endowing this research with comparability to studies using these data sources. Second, shadowed areas (pixels) were masked out of the image data using the thresholding method described above, so that subsequent derivations of greenness indices were performed only on sunlit pixels in these masked data (Table 2).

*National Agriculture Imagery Program Photo Data*

Because the October 2002 aerial photos were obtained during a period of phenological browning-down in the ONSR mapping area, 2-meter CIR photo images from the 2003 National Agriculture Imagery Program (NAIP) were acquired to obtain information during peak growing season conditions (July). Various greenness indices were derived and reflectance values in the infrared, red, and green wavelengths were obtained, as well as medians of these derivatives calculated within a 15 pixel × 15 pixel moving window to serve as classification inputs (Table 3).

*Landsat TM/ETM+*

Both leaf-off Landsat ETM+ and growing season Landsat TM image data (path 24, row 34) were employed as multispectral data sources for this research, because (a) recent cloud-free Landsat image data were available for these two phenological conditions over the ONSR mapping area, and (b) the multispectral and multitemporal information contained in these data were expected to enhance the potential to distinguish vegetation community types compared to using CIR photo data alone. The winter (leaf-off) image was acquired on 13 March 2002, and was geographically referenced to UTM Zone 15 NAD83 coordinates using 45 GCPs that were identifiable both on the Landsat and the reference NAIP orthophoto images. The resulting global root mean square error (RMSE) was 0.2454 pixel units for the portion of the image that covered the ONSR mapping area, indicating that the spatial error between the map source and image data is expected to be less than one quarter of a 30 meter × 30 meter ETM+ pixel. The growing season TM image was obtained on 05 July 2000, and was geographically referenced to UTM Zone 15 NAD83 coordinates using 40 GCPs that were identifiable on both the Landsat and reference NAIP orthophoto images with an global RMSE of 0.3135 pixel units, indicating an expected spatial error of less than one-third of a TM pixel. The TM and ETM+ image data were corrected to at-sensor reflectance following Markham and Barker (1986) using parameters published by Chander and Markham (2003; TM) and in the Landsat-7 Science Data Users Handbook (Irish, 2000; ETM+).

An empirical topographic normalization technique was applied to the TM data to reduce the influence of differential solar illumination related to topography (Allen, 2000). An empirical model relating solar illumination angle to differential reflectance of forested pixels was developed band by band (cos(*i*)), computed pixel by pixel from a 10-meter digital elevation model and solar azimuth and elevation values for the Landsat overflights) to calculate illumination corrected reflectance using the following equation:

$$R_i = R_0 - \cos(i) * M - B + R(\text{hat})$$

where M and B are regression parameters (reported in Chastain *et al.*, 2006), $R_o$ is the original reflectance, and R(hat) is the mean reflectance (Meyer *et al.*, 1993).

TABLE 3. DISCRIMINATING VARIABLES OBTAINED FROM DATA DERIVATIVES TAKEN FROM THE SUMMER 2003 NAIP IMAGE DATA (2-METER RESOLUTION)

| | |
|---|---|
| naipir[3] | NAIP infrared reflectance |
| naipred | NAIP red reflectance |
| naipgreen | NAIP green reflectance |
| naipirmedian15[1,2] | Median of infrared reflectance in a 15 × 15 neighborhood |
| naipredmedian15[1,3] | Median of red reflectance in a 15 × 15 neighborhood |
| naipgreenmedian15[1,2,3] | Median of green reflectance in a 15 × 15 neighborhood |
| naipndvi | Normalized difference vegetation index |
| naipndvimedian15[1,2] | Median of NDVI in a 15 × 15 neighborhood |
| Naipsqvi | Square root vegetation index |
| naipsqvimedian15[1,2,3] | Median of square root vegetation index in a 15 × 15 neighborhood |

1 = variable used in the discriminant analysis classification for the Hills and Breaks landscape strata
2 = variable used in the discriminant analysis classification for the Bottomland landscape strata
3 = variable used in the discriminant analysis classification for the Igneous Knobs landscape strata

The six Landsat visible and infrared bands typically contain substantial redundancy, which was reduced using principal components analysis to three bands (PC1, PC2, and PC3), explaining more than 95 percent of the variance in both of the original images. In addition, a tasseled cap transformation was applied using coefficients from Huang *et al.* (2002) to derive brightness (soils), greenness (vegetation), and wetness (plant canopy and soils) indices (Crist and Kauth, 1986). The Normalized Difference Vegetation Index (NDVI) and the Normalized Difference Moisture Index (NDMI) were also derived from the original Landsat image data to produce input variables for a statistical classification model (Table 4).

*Topographic Data*

Differences in color and other spectral reflectance characteristics are often not sufficient to distinguish similar plant communities from each other, so ancillary data are frequently used as input in remote sensing classifications to reduce this confusion. Specifically, topographic data has proven valuable in distinguishing among different suites of plant species based on niche specialization related to topographic gradients (Frank, 1988; Ohmann and Gregory, 2002). Topographic indices and other derivatives developed as landscape-scale representations of gradients associated with landform shape and position were obtained from a 10-meter DEM of the study area (Table 5). These indices represent measurements of indirect gradients hypothesized to control species distribution (Austin and Smith, 1989; Franklin, 1995). From the DEM data, the indirect gradients of slope, Beers-transformed aspect (Beers *et al.*, 1966), elevation, slope position, and slope curvature were derived to act as measurable surrogates to the direct gradients of exposure, moisture availability, temperature, and growing season length. Also derived was the terrain relative moisture index (TRMI), which represents an additive combination of slope angle, slope position, slope aspect (Beers-transformed), and

TABLE 4. DISCRIMINATING VARIABLES OBTAINED FROM DATA DERIVATIVES TAKEN FROM THE SUMMER AND WINTER LANDSAT TM/ETM IMAGE DATA (30-METER RESOLUTION)

| | |
|---|---|
| tmsum_b1 | Band 1 (blue: 0.45–0.52 $\mu$m) from 05 July 2000 TM image |
| tmsum_b2[1,2] | Band 2 (green: 0.52–0.60 $\mu$m) from 05 July 2000 TM image |
| tmsum_b3[1] | Band 3 (red: 0.63–0.69 $\mu$m) from 05 July 2000 TM image |
| tmsum_b4[1] | Band 4 (infrared: 0.76–0.9 $\mu$m) from 05 July 2000 TM image |
| tmsum_b5[1] | Band 5 (mid- IR: 1.55–1.75 $\mu$m) from 05 July 2000 TM image |
| tmsum_b6[1,2] | Band 7 (mid- IR: 2.08–2.35 $\mu$m) from 05 July 2000 TM image |
| tmsum_pc1[1,3] | Principal component 1 from 05 July 2000 TM image |
| tmsum_pc2 | Principal component 2 from 05 July 2000 TM image |
| tmsum_pc3 | Principal component 3 from 05 July 2000 TM image |
| tmsum_tc1[1] | Tasseled cap brightness from 05 July 2000 TM image |
| tmsum_tc2[2,3] | Tasseled cap greenness from 05 July 2000 TM image |
| tmsum_tc3 | Tasseled cap wetness from 05 July 2000 TM image |
| tmsum_ndvi[1,2] | NDVI (IR − Red/ IR + Red) obtained from 05 July 2000 TM image |
| tmsum_ndmi[1] | NDMI (IR − MIR/IR + MIR) obtained from 05 July 2000 TM image |
| tmwin_b1 | Band 1 (blue: 0.45–0.52 $\mu$m) from 13 March 2002 ETM image |
| tmwin_b2[1,2] | Band 2 (green: 0.52–0.60 $\mu$m) from 13 March 2002 ETM image |
| tmwin_b3[1] | Band 3 (red: 0.63–0.69 $\mu$m) from 13 March 2002 ETM image |
| tmwin_b4[1] | Band 4 (infrared: 0.76–0.9 $\mu$m) from 13 March 2002 ETM image |
| tmwin_b5[1,2] | Band 5 (mid-IR: 1.55–1.75 $\mu$m) from 13 March 2002 ETM image |
| tmwin_b6[1] | Band 7 (mid-IR: 2.08–2.35 $\mu$m) from 13 March 2002 ETM image |
| tmwin_pc1[1] | Principal component 1 from 13 March 2002 ETM image |
| tmwin_pc2[1,3] | Principal component 2 from 13 March 2002 ETM image |
| tmwin_pc3 | Principal component 3 from 13 March 2002 ETM image |
| tmwin_tc1 | Tasseled cap brightness from 13 March 2002 ETM image |
| tmwin_tc2[1,2] | Tasseled cap greenness from 13 March 2002 ETM image |
| tmwin_tc3[2] | Tasseled cap wetness from 13 March 2002 ETM image |
| tmwin_ndvi[1,2] | NDVI (IR − Red/ IR + Red) obtained from 13 March 2002 ETM image |
| tmwin_ndmi[1,2] | NDMI (IR − MIR/IR + MIR) obtained from 13 March 2002 ETM image |

1 = variable used in the discriminant analysis classification for the Hills and Breaks landscape strata
2 = variable used in the discriminant analysis classification for the Bottomland landscape strata
3 = variable used in the discriminant analysis classification for the Igneous Knobs landscape strata

TABLE 5. DISCRIMINATING VARIABLES OBTAINED FROM TOPOGRAPHIC DATA DERIVATIVES TAKEN FROM 10-METER RESOLUTION DEM DATA

| | |
|---|---|
| slope[1,2,3] | Slope angle in degrees – values range from 0 to 90 |
| beer100[1,2,3] | Beers-transformed slope aspect; index ranges from 0 to 200, with 0 being a grid cell that faces southwest, 100 indicating either northwest or southeast and 200 being northeast; the formula is (cos(aspect − 45) + 1)*100 |
| trnbeer[1,3] | Slope multiplied by Beers transformed aspect |
| rsp[1,2,3] | Relative slope position; ranges from 0 to 100, where 0 is a valley bottom and 100 is a ridgetop. |
| tci[1] | Topographic convergence index (TCI) is a measure of potential wetness; the formula is $\ln\alpha/\tan\beta$ where $\ln\alpha$ is the log of the upslope contributing area of a grid cell and $\tan\beta$ is the tangent of the slope of a grid cell |
| trmi[3] | Terrain relative moisture index; computed as slope + aspect + curvature + slope position; ranges in value from 0 to 60 (Parker, 1982) |
| rise[1,2,3] | Elevation rise in meters from Current or Jacks Fork Rivers; river channels have an elevation value of 0 |

1 = variable used in the discriminant analysis classification for the Hills and Breaks landscape strata
2 = variable used in the discriminant analysis classification for the Bottomland landscape strata
3 = variable used in the discriminant analysis classification for the Igneous Knobs landscape strata

curvature (Parker, 1982). These topographic gradients were used as independent variables similar to the remote sensing data to help differentiate NVCS vegetation associations based on the gradient affinities of their component species.

### Map Production

*Pilot Area Test of Statistical Methods*

A pilot area was delineated within the overall ONSR mapping area in which the capacity of supervised parallelepiped decision rule and classification and regression tree (CART) statistical approaches could be tested for the ONSR mapping area. The accuracy of maps produced using these two classification approaches were compared so that both their relative utility, as well as the inherent discernability of the different NVCS vegetation associations in the mapping area could be assessed. The CART approach has been applied in a number of remote sensing data classification problems, and has been demonstrated to produce robust results for land-cover classification in a number of environments and over a spectrum of scales (Hansen *et al.*, 2000; Joy *et al.*, 2003; de Colstoun *et al.*, 2003). The CART approach also provides a valuable data mining tool, in that the most relevant independent variables are chosen for the separate nodes (Venables and Ripley, 1994). The structure of a tree generated through the CART approach is therefore heuristically valuable, providing insight into the predictive structure of the support data as well as the discernability of the target categories, even in cases where they are not homogenous over the measurement space (Breiman *et al.*, 1984).

For the parallelepiped decision rule classification, ERDAS Imagine®, version 8.7 was used to create feature space objects that function as nonparametric signatures. These signatures were produced by drawing polygonal training areas around field survey plot locations, so that the spatial data in the immediate vicinity could be associated with those locations. A subset of less than half of the available survey points was used to train the classification, and the remaining data were set aside to validate the resulting map. The signatures were evaluated to determine which of the inputs should be included to optimize the classification by examining the transformed divergence, a measure of statistical distance between signatures. The nonparametric parallelepiped decision rule was used to assign individual training data cases to vegetation association types based on patterns in the independent variables, and the maximum likelihood decision rule was subsequently used to assign cases that either fell into more than one class (overlap), or were not within any of the parallelepiped class boundaries (unclassified) into one of the NVCS association classes. The choice to perform a wall-to-wall classification was motivated by the desire to emulate as closely as possible the greedy algorithm employed by the regression tree classification also performed for this pilot study, thus enhancing the comparability of the results of the two classification approaches.

The Recursive Partitioning and Regression Trees (RPART) package in the R Statistical Software, version 1.8.1 was used to develop a regression tree based on the training data described above. The class method was used to create a regression tree model to classify plant community associations, using the Gini index of impurity as a splitting rule to maximize impurity reduction during data splitting at tree nodes with prior probabilities proportional to observed class frequencies. No pruning was applied to the greedy outcome of the regression tree to maximize the number of NVCS vegetation associations classified by the tree model. A set of if-then statements based on the parameters in the regression tree was coded using the ArcGIS® software to produce a map of NVCS association types for the pilot area.

Unsatisfactory overall accuracy results were obtained from the two classification approaches tested in this pilot study. The low overall classification accuracies obtained from the CART model (39.2 percent; kappa = 0.367) and the parallelepiped decision rule classification (34.5 percent; kappa = 0.314) point out the limitations associated with the application of a single statistical classification model for a complex land-cover mapping problem, and provided a compelling argument for the application of a more complex hybrid mapping approach for the ONSR mapping area. The structure of the CART model did however provide a heuristically valuable tool, which was used to stratify the mapping area into relatively homogenous regions within which separate classification models could be applied. The primary splits identified by the CART model (Figure 3) separate the pilot study area into vegetation communities that occur on igneous knobs, in bottomland areas (including old fields), and those that occur on the remainder of the upland hills and breaks. Further, it was decided that because of the numerous categories in the NVCS association-level classification scheme and the availability of copious independent variables available for statistical classification model input, discriminant analysis represented a more appropriate method to pursue, as this statistical approach has shown promise in utilizing large hyperspectral remote sensing input datasets to produce detailed classification maps (Clark *et al.*, 2005; Karimi *et al.*, 2005).

*Discriminant Analysis Statistical Model*

A discriminant analysis statistical model was developed to discern differences between the various NVCS vegetation association types. Discriminant analysis is a statistical approach in which all of the cases in a dataset can be assigned to the group to which they most closely resemble by discerning differences between two or more groups using several variables simultaneously (Klecka, 1980). In this research, discriminant analysis was used to classify all of the pixels in the ONSR mapping area as the NVCS association type to which they have the highest probability of membership using linear combinations of the large set of discriminating variables derived from remote sensing and ancillary data (Figure 4). A collection of 92 discriminating variables obtained from remote sensing and topographic data sources served as potential training data for this statistical classification approach (Tables 2 through 5). Separate discriminant analysis classifications were performed for bottomland areas, hills and breaks, and igneous knobs as delineated using an ELT classification. The progression of the classification approach entailed: (a) variable selection using stepwise discriminant analysis (STEPDISC procedure in SAS), (b) reduction of input variable dimensionality using canonical discriminant analysis (CANDISC procedure in SAS) to produce canonical discriminant functions from linear combinations of the discriminating variables chosen in step (a), (c) classification based on the probabilities generated using discriminant analysis (DISCRIM procedure in SAS) with the canonical functions created in step (b), and (d) conversion of the canonical functions and discriminant analysis classification results to raster image datasets.

The ERDAS Imagine® Model Maker was used to apply the results from the discriminant analysis to create raster image datasets (see Figure 4). This included mathematically combining the various image surfaces representing the input discriminating variables to produce images containing continuous representations of the canonical discriminant functions used in the classification models run for the three landscape strata (bottomland, hills and breaks, and igneous knobs). The discriminant function coefficients generated in SAS were next incorporated into a second set of models wherein they were combined to produce images for each NVCS association with

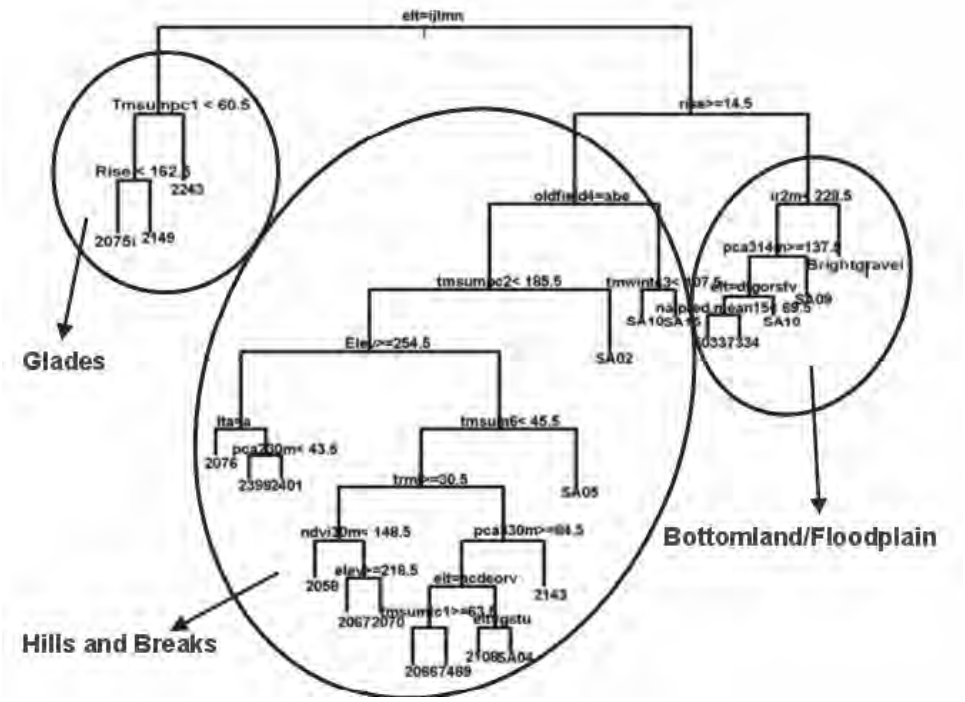


Figure 3. Classification and regression tree results, with structural features that suggested spatial stratification of the ONSR mapping area into igneous glades, hills and breaks, and bottomland/floodplain regions to enable a separate classification for each of these regions.

continuous spatial representations of classification scores for the three landscape strata. From these classification scores, posterior probabilities of group membership for all of the NVCS association types were computed (Klecka, 1980). A final set of models created thematic raster images for the three landscape strata using a conditional statement to assign the NVCS association type with the highest membership probability to each cell in the ONSR mapping area. A second set of raster images was created to represent the membership probability value for each raster cell.

*Decision Rule Classification for River Channel and Gravel Bars*

A non-parametric parallelepiped decision rule classification was performed using the ERDAS Imagine® software to discriminate the river channel and non-vegetated (bright) gravel bar features in the ONSR mapping region. Infrared, red, and green grey-level values combined with the first and third principal components and a vegetation index (IR/R) from the 2-meter 2002 CIR photo image data served as input data for this classification. These six input variables were chosen from the numerous data derivations created from the 2-meter CIR photo image data for this classification based on the results of a transformed divergence analysis of the statistical distance between signatures, using the guidelines proposed by Jensen (1996). The classification was performed within an area limited to a spatial mask that contained river channel and bright gravel bar land-cover features. This separate classification permitted a detailed characterization of the locations of the river channel and bare gravel bars, which should be regarded as a *snapshot* of the location of the river channel and gravel bar complex at the time of the October 2002 image acquisition, because their spatial expression is ephemeral and change with each high river discharge event (Jacobson and Gran, 1999).

*Photointerpretation and Digitizing*

Portions of the ONSR mapping area that have been significantly altered, such as roads, fields, logged areas, utility corridors, urban and residential areas, and quarries were photointerpreted and digitized from the 0.25-meter resolution October 2002 aerial photo images. Other land-cover features such as surface water and glade-woodland complex areas were also digitized from these images. The decision to digitize these features rather than attempting to discriminate them using the automated statistical classification approach was based on the low accuracy results obtained in the pilot area classifications. Agricultural forested woodlots (SA37) were digitized and identified as woodlots rather than classified along with other forested areas because they are expected to function differently than more contiguous forested areas with respect to species composition and habitat characteristics (Boutin and Jobin, 1998; Nupp and Swihart, 2000) and occur on private lands, thereby prohibiting the characterization of their composition and structure. Historic agricultural fields and pastures were also photointerpreted and digitized from a 1-meter resolution aerial photo mosaic obtained in October 1964 over the ONSR mapping region. This was done in an attempt to identify historic land-use patterns during this period, as more than two-thirds of the fields apparent in this imagery that are located in what is now the ONSR park have been abandoned and are now in various successional stages of reforestation. The geospatial representation of these successional communities was produced using a spatial overlay of the 1964 fields over areas that were classified as forested vegetation community classes through discriminant analysis.

The final map for the ONSR area was assembled by superimposing the digitized human-dominated land-cover features, glade complexes, historic fields, and the classified

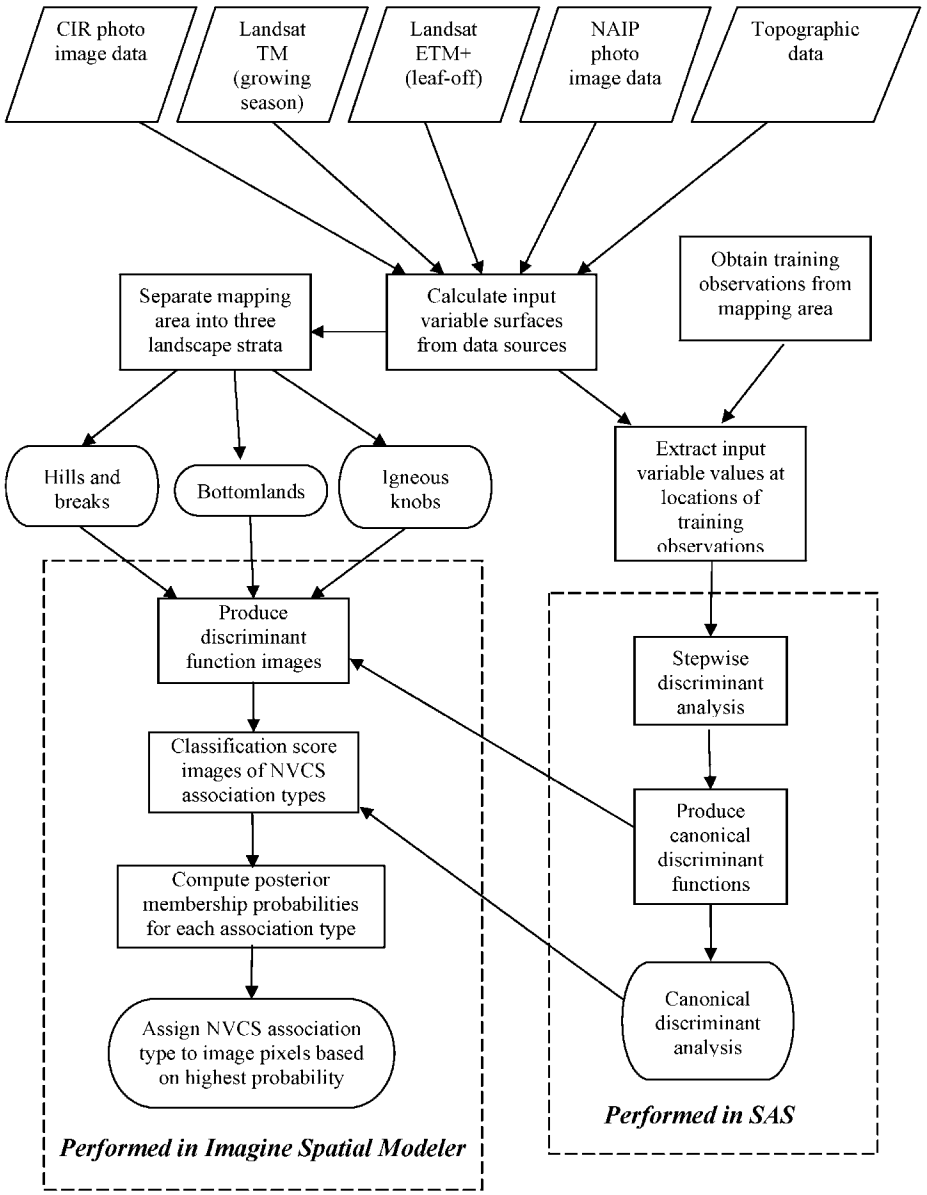


Figure 4. Flow chart describing the data processing steps performed to operationalize the discriminant analysis classification.

river channel and bare gravel bars over the separate discriminant analysis results obtained for the stratified hills and breaks, bottomland, and igneous glade regions. The logic of overlay order follows the manner in which anthropogenic land-cover modifications are superimposed over natural vegetation communities.

*Validation Sample Location*

A stratified sample of points randomly located in each land-cover polygon was generated within a sampling universe that was limited to public lands to ensure that access would be possible for field personnel. The randomly located field sampling points located within 30-meters of classified polygon boundaries were replaced with a point located at the calculated polygon centroid to minimize both GPS spatial registration errors and the potential for sampling in transitional and/or ecotonal areas. A goal was set to visit at least 30 validation points for every vegetation community type, and emphasis was placed to visit under-sampled community types. A total of 2,057 validation points were applied to assess overall and class-wise map accuracy. This validation data set was composed of 1697 field-visited points (Figure 1), which included some previously visited points withheld from training data, augmented by 360 photointerpretation observations located in human dominated land-cover categories. The NVCS association type and the percent cover of dominant trees in the canopy stratum were noted at the locations of field-visited points to enable the calculation of species importance values (average cover of the individual tree species recorded multiplied with their relative frequency) within the NVCS associations, which were subsequently used to better understand some of the causes of inter-association confusion in the resulting map during the error analysis phase of this research.

## Results

### Vegetation Community Maps

Using a combination of statistical methods and photointerpretation, natural and altered vegetation communities as well as other land-cover classes in the ONSR mapping area were mapped at a 10-meter spatial resolution to two levels of classification detail. The primary mapping level, which includes 49 classes, conforms with the highly-detailed association level of the NVCS (Figure 5). Overall accuracy is 63 percent (k = 0.596) for this detailed map. Twenty-four map classes relate to NVCS vegetation associations, while the 25 remaining classes include cultural and non-vegetated features, and ruderal communities on abandoned agricultural lands. The discriminant analysis statistical method described here (Figure 4) used a large number of the overall set of input variables to map NVCS vegetation communities, which cover 84 percent of the total spatial extent of the ONSR mapping area (Chastain *et al.*, 2006). Photointerpretation methods were applied to map the cultural and non-vegetated land-cover classes. The second mapping level combines vegetation assemblages which appear similar or are ecologically related into locally-identified Community Types (Figure 6). The overall accuracy of this 33-class map is 77.5 percent (k = 0.716), just slightly below the USGS-NPS vegetation mapping standard of 80 percent.

Various types and scales of input data were chosen in the stepwise inclusion component of the discriminant analysis statistical approach implemented in this research. A total of 58 out of 92 input variables (63 percent, Tables 2 through 5) were chosen through stepwise discriminant analysis for the hills and breaks region of the mapping area. These 58 variables included six topographic variables, four texture variables, and 21 greenness variables. The remote sensing variables were obtained at spatial resolutions of 2- ($n$ = 8), 15- ($n$ = 11), and 30- ($n$ = 33) meters and capture different phenological conditions (growing season, leaf-off, and autumnal browning-down). A total of 24 out of 92 variables (26.1 percent, Tables 2 through 5) were chosen for the igneous regions, including six topographic variables, one texture variable, 11 greenness variables, and with remote sensing variables obtained at spatial resolutions of 2- ($n$ = 2), 15- ($n$ = 10), and 30- ($n$ = 6) meters. Finally, a total of 38 out of 92 variables (41.3 percent, Tables 2 through 5) were chosen for the bottomland/floodplain portion of the mapping area, including four topographic variables, three texture variables, 12 greenness variables, and with remote sensing variables obtained at spatial resolutions of 2- ($n$ = 7), 15- ($n$ = 14), and 30- ($n$ = 13) meters. These results point to the fact that high-resolution textural information, color and reflectance information at various scales, and topographic gradients modeled from elevation data, all provided valuable information for vegetation mapping at the vegetation association level of specificity in the ONSR area.

### Map Accuracy

The overall accuracy of the 49-class NVCS association level map for the ONSR mapping region is 63 percent, with a kappa statistic of 0.596. This estimate of overall accuracy is based on 2,057 validation observations obtained from field-sampled points withheld from the statistical classification, field observations obtained in the summer of 2005, and from photointerpretation observations located within human-dominated, land-cover categories. The error matrix (Table 6) indicates areas where confusion occurred between mapping categories, and was used to calculate user's and producer's accuracies for the various categories (Table 7). Additional accuracy measures were obtained by computing receiver operating characteristic (ROC) curves from the class membership probabilities generated using the discriminant analysis classifier applied in this research, and provide a useful class-wise assessment of the results of remote sensing classifications (Pontius and Schneider, 2001; Gardner and Urban, 2003).

The producer's accuracy (total correctly classified observations divided by the total observations of one ground reference type classified into any category) indicates the probability that a reference location is correctly classified, and is a measure of omission error or how well a land-cover type can be classified given the available training data (Jensen, 1996). User's accuracy (total correctly, classified observations divided by the total ground reference observations classified into a category that are correctly or incorrectly classified) is the probability that a location on a classification map actually represents that category on the ground, and is referred to as commission error (Lillesand and Kiefer, 1994). The USGS-NPS Vegetation Mapping program standards set a goal of 80 percent for both within class and overall accuracy. Where this goal is not attained, aggregation of classes into broader categories is permitted. Class aggregation is facilitated in this research by mapping locally-identified Community Types after NatureServe's scological systems approach.

Receiver operating characteristic (ROC) curves computed from the class membership probabilities produced using discriminant analysis represent an integration of the uncertainty information contained in the user's and producer's accuracy figures obtained for a given map class. An ROC curve is generated by plotting the probability of detection against the probability of false positives (the probability of false alarm) over increasing discrimination thresholds. The area under a ROC curve (AUC) functions as a comparative measure among mapping classes of discrimination strength or accuracy, and can be interpreted as the probability that in randomly chosen positive and negative cases, a higher score will be assigned to the positive rather than the negative case by the classifier (Mason and Graham, 2002). AUC measures were calculated from ROC curves to obtain per-class measures of classification accuracy for the NVCS vegetation association map classes for which there were sufficient support validation data (Table 6, Figure 7). The AUC measures for all except the dolomite Post Oak-Blackjack Oak Forest and Post Oak-Blackjack Oak/Little Bluestem Woodland (2075 and 2149) NVCS associations are well above the no-discrimination line formed by a horizontal line from bottom left to top right (AUC = 0.5). The low accuracy associated with these two classes is related to the rareness of these vegetation associations, with validation sample sizes of only 5 and 4, respectively.

The 33-class community-type map produced through class aggregation guided by the ecological systems approach achieved an overall accuracy of 77.5 percent (kappa = 0.716), and thus conforms closely with USGS-NPS Vegetation Mapping Program accuracy goal of 80 percent. Riverine shrublands and gravel bars are aggregated upward into an "Active Channel/Gravel Bar Complex" and bottomland forests are combined into a "Riverfront and Bottomland Forests" type in this map. Dry-mesic, white oak dominated forests are aggregated into "White Oak Forests," oak and hickory forests with greater than 25 percent pine cover are aggregated into "Pine and Pine-Oak Forests," and oak and hickory dominated forests lacking significant pine are combined into "Mixed Oak-Hickory Forests." Glades and associated woodlands on both igneous and dolomite substrates are combined into "Igneous Glade/Woodland Complex" and "Dolomite Glade/Woodland Complex," respectively.

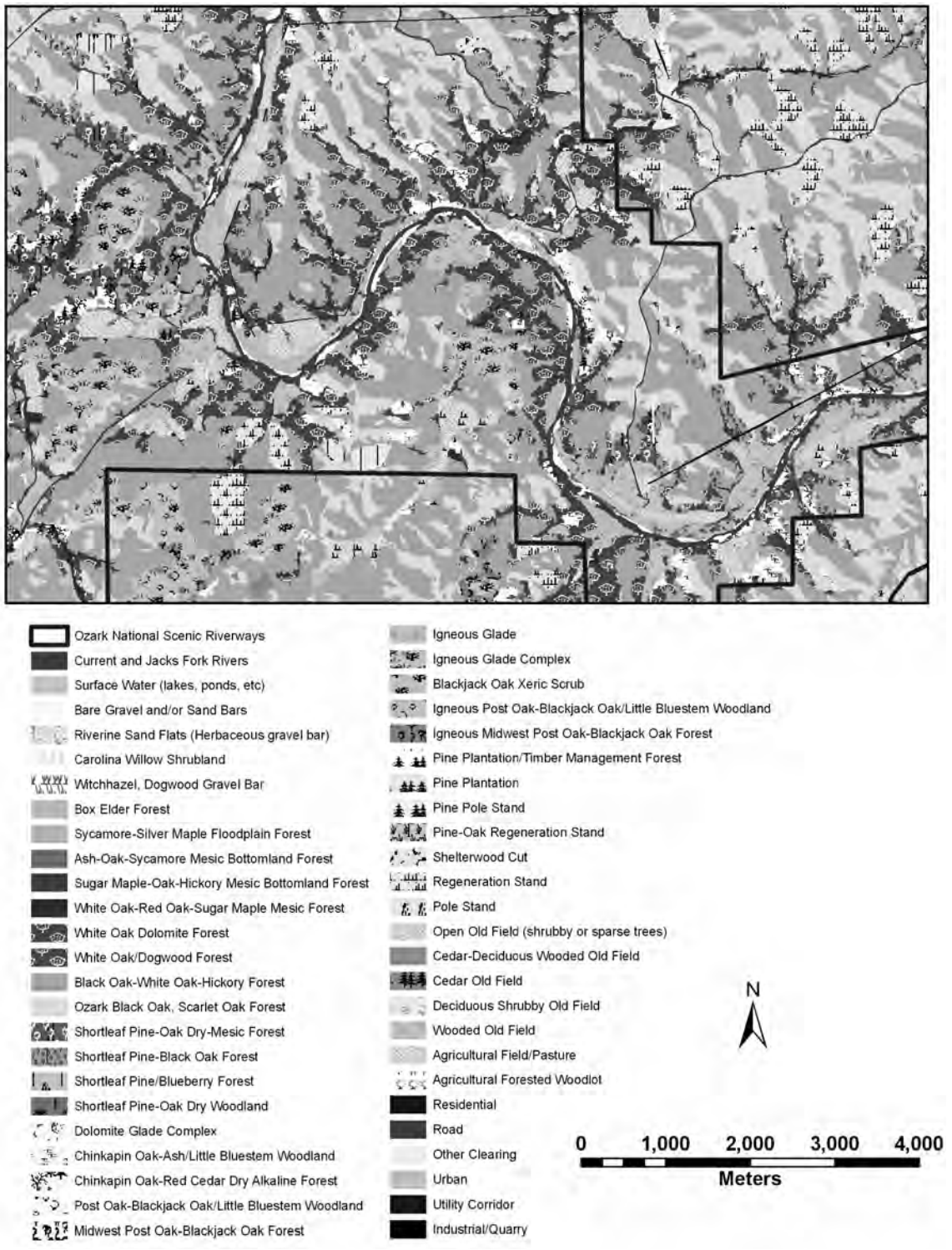


Figure 5. Detail of the 49-class nvcs vegetation association map. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

Improvement of class-wise accuracies in the 33-class vegetation community map was substantial, further enhancing the conformity of this map to Vegetation Mapping Program standards (Tables 6 and 7). For example, when examined individually, the user's accuracies for the Ozark Black Oak-Scarlet Oak Forest and the Black Oak-White Oak-Hickory Forest were 36 and 83 percent, respectively. When combined into the Mixed Oak-Hickory Forest Community Type, user's accuracy improved to 84 percent. Grouping the White Oak/Dogwood Forest (42 percent user's accuracy) and the White Oak-Mixed Oak Dry-Mesic Alkaline Forest (30 percent) into the White Oak Forest Community Type improved user's accuracy to 59 percent (Table 8). Aggregation of the Shortleaf Pine/Blueberry Forest, the Shortleaf

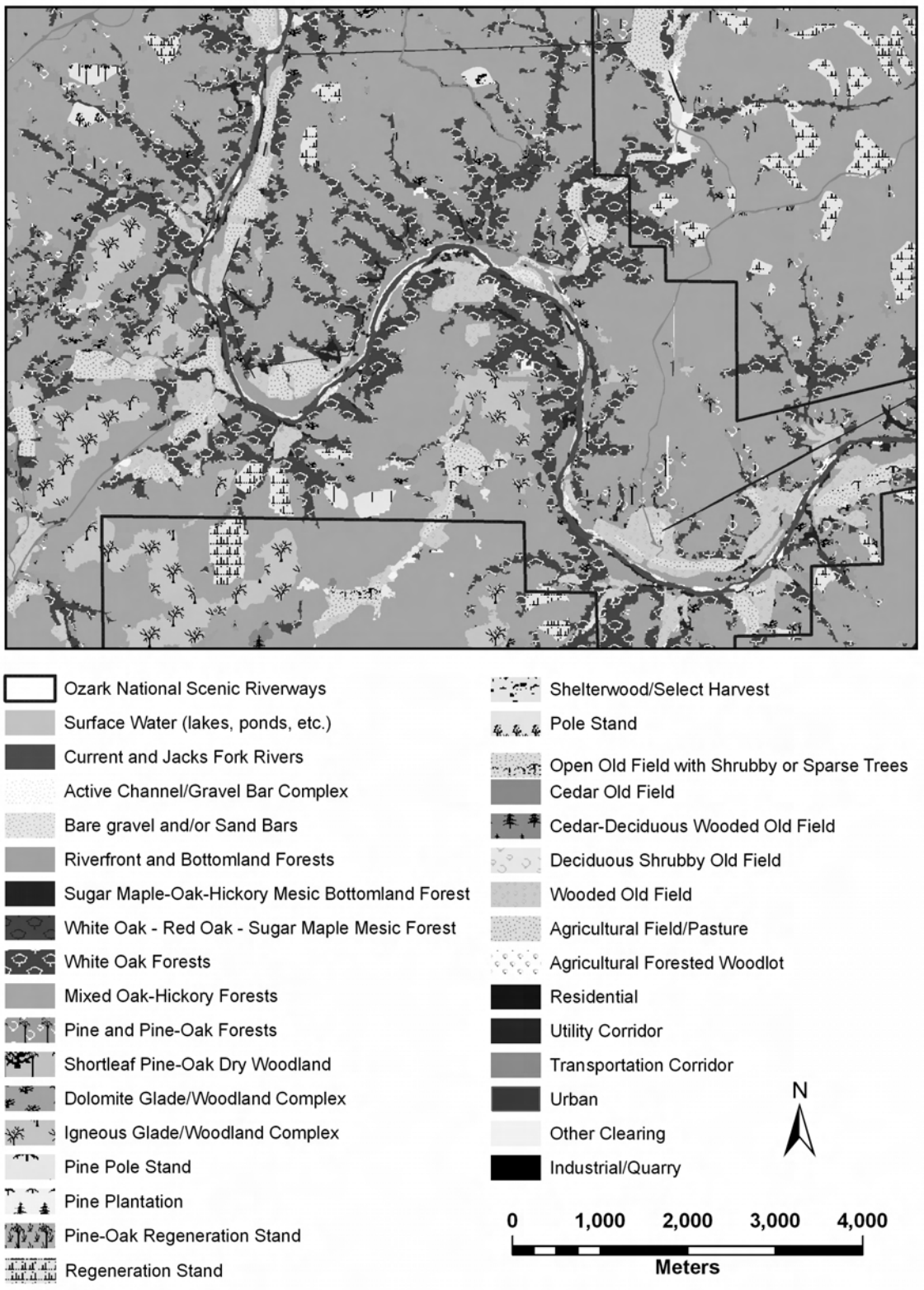


Figure 6. Detail of the 33-class Community Type map in which NVCS associations were aggregated based on to improve accuracy and aid in resource management planning. A color version of this figure is available at the ASPRS website: **www.asprs.org**.

TABLE 6. ERROR MATRIX OF NVCS ASSOCIATIONS AND OTHER LAND-COVER CATEGORIES. BOXES ALONG DIAGONAL DEMARCATED BY BOLD LINES INDICATE NATURESERVE COMMUNITY TYPES (CLASSES FROM THE 49-CLASS MAP)

| | Reference Data (ground truth) | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Map Classification | 2058 | 2066 | 2070 | 2076 | 2399 | Dolomite Glade Complex | 2108 | 2143 | 2149 | Igneous Glade Complex | 2243 | 2425 | 2075 | 2393 | 2400 | 2401 | 7489 |
| White Oak-Red Oak-Sugar Maple Mesic Forest (2058) | 18 | 14 | 8 | 3 | | | 1 | | | | | | | | | | 1 |
| White Oak/Dogwood Forest (2066) | 3 | 57 | 37 | 27 | | | 2 | | | | | | | | | | 5 |
| White Oak Dolomite Forest (2070) | 4 | 17 | 32 | 28 | 2 | | 4 | 4 | | | | | 2 | 2 | | | 1 |
| Black Oak-White Oak-Hickory Forest (2076) | | 15 | 3 | 242 | 18 | | | 1 | | | | | 5 | | | 3 | 2 |
| Ozark Black Oak, Scarlet Oak Forest (2399) | | 2 | | 47 | 47 | | 1 | | | | | | 5 | 3 | | 16 | 5 |
| Dolomite Glade Complex | | | | | | 10 | 5 | 1 | 5 | | | | | 2 | | | |
| Chinquapin Oak-Red Cedar Dry Alkaline Forest (2108) | | | 5 | 1 | | | 69 | 9 | 1 | | | | | 8 | | 6 | 5 |
| Chinquapin Oak-Ash/Little Bluestem Woodland (2143) | | | 1 | 4 | | 2 | | 18 | | | | | 1 | 1 | | 1 | |
| Post Oak-Blackjack Oak/Little Bluestem Woodland (2149) | | 1 | | | 1 | | | | 4 | 1 | | | 6 | | | | 1 |
| Ozark Igneous Glade Complex (includes 2075i) | | | | | | | | | 4 | 42 | 1 | 4 | | | | | |
| Ozark Igneous Glade (2243) | | | | 1 | | | | | 3 | | 3 | 3 | 3 | 1 | | 1 | |
| Blackjack Oak Xeric Scrub (2425) | | | | | | | | | 2 | 1 | | 2 | 1 | | | | |
| Midwest Post Oak-Blackjack Oak Forest (2075) | | | | | | | | | 2 | | | | 7 | | | 1 | |
| Shortleaf Pine-Oak Dry Woodland (2393) | | | | 3 | 1 | | | 1 | 1 | 1 | | | | 7 | | 1 | |
| Pine/Blueberry Forest (2400) | | | | | 1 | | | | | | | | | 1 | 1 | 1 | |
| Shortleaf Pine-Black Oak (2401) | | | | 5 | 1 | | | | | | | | 2 | 3 | 1 | 42 | 20 |
| Shortleaf Pine-Oak Dry-mesic Forest (7489) | | 1 | | 8 | 1 | | 5 | 1 | | | | | | | | 4 | 35 |
| Bare Gravel Bar | | | | | | | | | | | | | | | | | |
| Riverine Sand Flats (2049) | | | | | | | | | | | | | | | | | |
| Witchhazel, Dogwood Gravel Wash (3898) | | | | | | | | | | | | | | | | | |
| Carolina Willow Shrubland (3899) | | | | | | | | | | | | | | | | | |
| Ash-Oak-Sycamore Mesic Bottomland Forest (2410) | 1 | | 2 | | | | | | | | | | | | | | |
| Box Elder Forest (5033) | | | | | | | | | | | | | | | | | |
| Sycamore-Silver Maple Floodplain Forest (7334) | | 1 | 1 | | | | | | | | | | | | | | |
| Sugar Maple-Oak-Bitternut Hickory Mesic Bottomland Forest (2060) | 2 | | 4 | 1 | | | | | | | | | | | | | |
| Wooded Old Field | | 2 | | 1 | | | 2 | 1 | | | | | | | | | |
| Open Old Field | | | 2 | | | | | 1 | | | | | | | | | |
| Shelterwood Harvest | | | | | | | | | | | | | | | | | 1 |
| Regeneration/Pole Stand | | | | | 1 | | | | | | | | | | | | |
| Pine Plantation | | | | 1 | | | | | | | | | | | | | |
| River | | | | | | | | | | | | | | | | | |
| Roads | | | | | | | | | | | | | | | | | |
| Utility Corridor | | | | | | | | | | | | | | | | | |
| Agricultural Field/Pasture/ Other Clearing | | | | | | | | | | | | | | | | | |
| Residential/Urban/Industrial | | | | | | | | | | | | | | | | | |
| Totals | 28 | 110 | 95 | 372 | 73 | 12 | 89 | 37 | 22 | 45 | 4 | 9 | 32 | 28 | 2 | 76 | 76 |

Pine-Oak Dry-Mesic Forest, and the Shortleaf Pine-Black Oak Forest into the Upland Pine and Pine-Oak Forest Community Type improved user's accuracy from a mean value of 42 percent to a collective value of 70 percent. Aggregation of communities into the Dolomite and Igneous Glade/Woodland Complexes improved user's accuracy from mean values of 49 and 51 percent to 66 and 82 percent, respectively. The Active Channel/Gravel Bar Complex had an aggregated user's accuracy of 95 percent, compared with a mean accuracy value of 65 percent for the three classes included in that Community Type. Finally, aggregation of riverfront and floodplain forests into the Riverfront and Bottomland Forests Community Type improved accuracy from a mean value of 53 percent for individual communities to 76 percent overall.

Examination of tree species importance values (computed from field observation data) within the NVCS association types elucidated some of the errors contributing to the

TABLE 6. (CONTINUED)

| Map Classification | Reference Data (ground truth) | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Gravel Bar | 2049 | 3898 | 3899 | 2410 | 5033 | 7334 | 2060 | Wooded Old Field | Open Old Field | Shelterwood | Regen. Stand | Pine Plant. | River | Road | Utility Corr. | Field/Pasture | Urban | Total |
| White Oak-Red Oak-Sugar Maple Mesic Forest (2058) | | | | | | | 6 | 2 | 7 | | 2 | 1 | | 1 | | | | | 64 |
| White Oak/Dogwood Forest (2066) | | | | | | | | 4 | 1 | | | | 1 | | | | | | 137 |
| White Oak Dolomite Forest (2070) | | | | | | | | 1 | 8 | 1 | | | | | | | | | 106 |
| Black Oak-White Oak-Hickory Forest (2076) | | | | | | | | | | | 1 | | | | | | | | 290 |
| Ozark Black Oak, Scarlet Oak Forest (2399) | | | | | | | | | 2 | | | 1 | | | | | | | 129 |
| Dolomite Glade Complex | | | | | | | | | | | | | | | | | | | 23 |
| Chinquapin Oak-Red Cedar Dry Alkaline Forest (2108) | | | | | 1 | | | | 3 | | | | | | 2 | | | | 110 |
| Chinquapin Oak-Ash/Little Bluestem Woodland (2143) | | | | | | | 2 | | 1 | | 2 | 1 | | | | | | | 34 |
| Post Oak-Blackjack Oak/Little Bluestem Woodland (2149) | | | | | | | | | 1 | | 7 | | | | | | | | 22 |
| Ozark Igneous Glade Complex (includes 2075i) | | | | | | | | | | | | | | | | | | | 51 |
| Ozark Igneous Glade (2243) | | | | | | | | | | | | | | | | | | | 15 |
| Blackjack Oak Xeric Scrub (2425) | | | | | | | | | | | | | | | | | | | 6 |
| Midwest Post Oak-Blackjack Oak Forest (2075) | | | | | | | | | | | | | | | | | | | 10 |
| Shortleaf Pine-Oak Dry Woodland (2393) | | | | | | | | | | | | | | | | | | | 15 |
| Pine/Blueberry Forest (2400) | | | | | | | | | 1 | | | 1 | | | | | | | 6 |
| Shortleaf Pine-Black Oak (2401) | | | | | | | | | | | 1 | | | | | | | | 75 |
| Shortleaf Pine-Oak Dry-mesic Forest (7489) | | | | | | | | | 7 | 1 | 1 | | | | 2 | | | 1 | 67 |
| Bare Gravel Bar | 53 | | | | | | | | | | | | | | | | | | 53 |
| Riverine Sand Flats (2049) | | 8 | 2 | 2 | | | | | | | | | | | | | | | 12 |
| Witchhazel, Dogwood Gravel Wash (3898) | | 2 | 5 | 4 | | | | | | | | | | | | | | | 11 |
| Carolina Willow Shrubland (3899) | | | 2 | 28 | 2 | | 1 | | | | | | | | | | | | 33 |
| Ash-Oak-Sycamore Mesic Bottomland Forest (2410) | | | 1 | 2 | 23 | 10 | 25 | 6 | 16 | 2 | | | | | 2 | 2 | | | 92 |
| Box Elder Forest (5033) | | | | | 4 | 25 | 8 | 4 | 2 | | | | | | | | | | 43 |
| Sycamore-Silver Maple Floodplain Forest (7334) | | | | 3 | 3 | 4 | 55 | | 3 | | | | | | | | | 1 | 71 |
| Sugar Maple-Oak-Bitternut Hickory Mesic Bottomland Forest (2060) | | | | | 1 | 2 | 8 | 17 | 6 | | | | | | | | | | 41 |
| Wooded Old Field | | | | | 1 | 1 | | | 144 | 6 | | 3 | | | | | | 2 | 163 |
| Open Old Field | | | | 1 | | | | | 19 | 12 | | | | | | 1 | | 3 | 39 |
| Shelterwood Harvest | | | | | | | | | | | 3 | | | | | | | | 4 |
| Regeneration/Pole Stand | | | | | | | | | | | 1 | 18 | | | 2 | | | | 22 |
| Pine Plantation | | | | | | | | 1 | | | | | | | | | | | 2 |
| River | | | | 1 | | | | | | | | | | 51 | | | | | 52 |
| Roads | | | | | | | | | | | | | | | 72 | 2 | | | 74 |
| Utility Corridor | | | | | | | | | | | | | | | | 38 | | | 38 |
| Agricultural Field/Pasture/ Other Clearing | | | | | | | | | | 2 | | | | | | | 68 | 4 | 74 |
| Residential/Urban/Industrial | | | | | | | | | | | | | | | | | | 73 | 73 |
| Totals | 53 | 10 | 10 | 41 | 35 | 42 | 105 | 35 | 221 | 24 | 18 | 25 | 1 | 52 | 80 | 43 | 68 | 84 | 2057 |

lower user's accuracy obtained in some of the NVCS and aggregated classes. These errors can be attributed to (a) similarities in species composition among vegetation classes, arising from overlaps inherent in the class definitions with respect to diagnostic and dominant tree species, and (b) the prevalence of transitional areas between *pure* expressions of vegetation community classes, within which field validation observations are sometimes made. The White Oak Forest Community Type, which obtained a user's accuracy of only 59 percent, is an example of the former source of classification error. Confusion occurs between its component associations (2066 and 2070) and the most abundant vegetation class, the Black Oak-White Oak-Hickory Forest (2076), which shares many dominant tree species, particularly white oak. Because of the abundance of these classes in the mapping area, confusion between these vegetation classes has a somewhat inordinate effect in suppressing both class-wise and overall accuracy. The 22 cases of confusion between White Oak-Red Oak-Sugar Maple Mesic Forest (2058) and white oak forest associations (2066 and 2070)

TABLE 7. PRODUCER'S ACCURACY, USER'S ACCURACY AND THE AREA UNDER THE RECEIVER OPERATING CHARACTERISTIC CURVE (AUC) FOR NVCS VEGETATION ASSOCIATIONS (49-CLASS MAP) AND OTHER LAND-COVER CATEGORIES

| NVCS Association/Land-cover Type | Producer's Accuracy | User's Accuracy | ROC Curve AUC |
|---|---|---|---|
| White Oak-Red Oak-Sugar Maple Mesic Forest (2058) | 0.64 | 0.28 | 0.84 |
| White Oak/Dogwood Forest (2066) | 0.52 | 0.42 | 0.77 |
| White Oak Dolomite Forest (2070) | 0.34 | 0.30 | 0.60 |
| Black Oak-White Oak-Hickory Forest (2076) | 0.65 | 0.83 | 0.83 |
| Ozark Black Oak, Scarlet Oak Forest (2399) | 0.64 | 0.36 | 0.83 |
| Dolomite Glade/Woodland Complex | 0.83 | 0.43 | – |
| Chinquapin Oak-Red Cedar Dry Alkaline Forest (2108) | 0.78 | 0.63 | 0.90 |
| Chinquapin Oak-Ash/Little Bluestem Woodland (2143) | 0.49 | 0.53 | 0.74 |
| Post Oak-Blackjack Oak/Little Bluestem Woodland (2149) | 0.18 | 0.18 | 0.47 |
| Igneous Glade/Woodland Complex | 0.93 | 0.82 | – |
| Ozark Igneous Glade (2243) | 0.75 | 0.20 | – |
| Blackjack Oak Xeric Scrub (2425) | 0.22 | 0.33 | – |
| Midwest Post Oak-Blackjack Oak Forest (2075) | 0.22 | 0.70 | 0.46 |
| Shortleaf Pine-Oak Dry Woodland (2393) | 0.25 | 0.47 | 0.63 |
| Pine/Blueberry Forest (2400) | 0.50 | 0.17 | – |
| Shortleaf Pine-Black Oak (2401) | 0.55 | 0.56 | 0.80 |
| Shortleaf Pine-Oak Dry-mesic Forest (7489) | 0.46 | 0.52 | 0.80 |
| Bare Gravel Bar | 1.00 | 1.00 | – |
| Riverine Sand Flats (2049; herbaceous gravel bars) | 0.80 | 0.67 | – |
| Witchhazel, Dogwood Gravel Wash (3898) | 0.50 | 0.45 | – |
| Carolina Willow Shrubland (3899) | 0.68 | 0.84 | 0.69 |
| Ash-Oak-Sycamore Mesic Bottomland Forest (2410) | 0.66 | 0.25 | 0.79 |
| Box Elder Forest (5033) | 0.60 | 0.58 | 0.76 |
| Sycamore-Silver Maple Floodplain Forest (7334) | 0.52 | 0.77 | 0.70 |
| Sugar Maple-Oak-Hickory Mesic Bottomland Forest (2060) | 0.49 | 0.41 | 0.70 |
| Wooded old field | 0.65 | 0.88 | – |
| Open Old Field with Shrubby or Sparse Tree | 0.50 | 0.31 | – |
| Shelterwood cut | 0.17 | 0.75 | – |
| Regeneration/Pole stand | 0.72 | 0.82 | – |
| Pine Plantation | 0.00 | 0.00 | – |
| River (non-vegetated) | 0.98 | 0.98 | – |
| Road | 0.90 | 0.97 | – |
| Utility corridor | 0.88 | 1.00 | – |
| Agricultural Field/Pasture/Other Clearing | 1.00 | 0.92 | – |
| Residential/Urban/Industrial | 0.87 | 1.00 | – |

represent an example of classification error due to transitional community prevalence. Classification confusion here is caused by the fact that the five most important species identified in these 22 cases are also the most important in the correctly classified validation observations. These misclassified cases can be said to be compositionally transitional communities that share similarities with both the mixed hardwood mesic forest and white oak forest associations. Similarly, 16 cases of confusion exist in the validation data between transitional expressions of Ozark Black Oak-Scarlet Oak Forest (2399) and Shortleaf Pine-Black Oak Forest (2401) associations, caused by field validation observations of class 2401 in locations where black oak was actually more important than the pine that is diagnostic of this class.

## Discussion

The integrated method employed in this research, merging (a) statistical classifications performed within separate landscape strata, (b) photointerpeted landscape elements, and (c) information gleaned from historical aerial photography, yielded a very detailed map at a level of accuracy required to guide the management of a large National Park area. Comparison of the results of the limited pilot classifications with the final classification of the entire ONSR mapping region indicates that a hybrid combination of statistical methods and photointerpretation is needed to obtain overall and class-wise accuracy levels adequate for USGS-NPS Vegetation Mapping Program standards. While the discriminant analysis approach presented here was applied to map 84 percent of the ONSR mapping area (Chastain *et al.*, 2006), complementary methods proved valuable to accurately map additional land-cover classes. The search for repeated patterns that drives a statistical classifier tends to be undermined by the anomalous land-use patterns that produce highly altered vegetation communities and human-dominated land-cover types. For example, harvested and regenerating timber stands and the old field association types can occur in numerous landscape positions and have varying reflectance signatures, but were readily photointerpreted, digitized, and superimposed over the mapped results of the discriminant analysis classifier. This form of spatial overlay is also logically consistent with the manner in which anthropogenic modifications are superimposed over a landscape matrix of relatively natural plant communities. Likewise, dolomite and igneous glades (which are sparsely treed savanna-like communities that contain bright rocky areas, green vegetation, and dark shadows) represent reflectively mixed targets that defy accurate discernment via statistical classification using remote sensing image data (Smith, 2001), and were more accurately mapped through photointerpretion.

Discriminant analysis proved to be an effective approach in which the large set of input variables that was available for this research could be fused to statistically classify a landscape with a complex pattern of vegetation communities. The input training data included both topographic

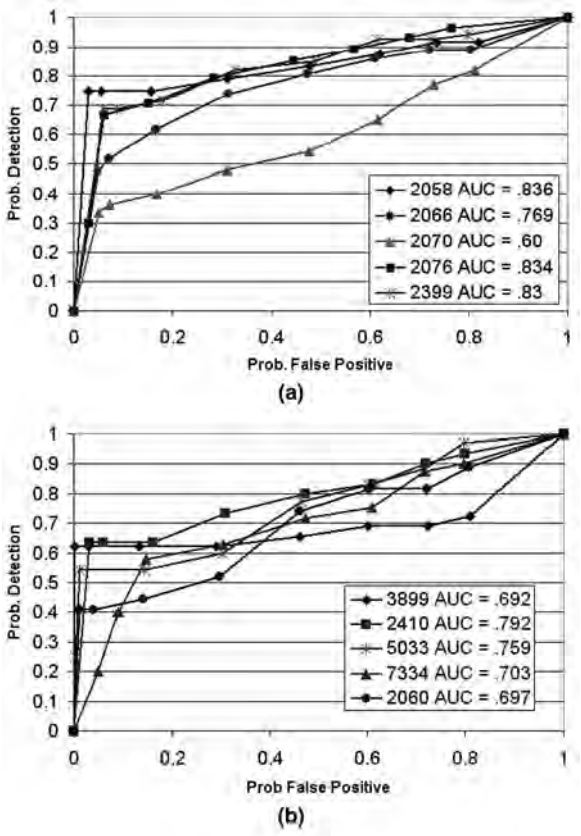


Figure 7. ROC curves obtained for (a) mesic and upland oak and hickory forest classes, and (b) bottomland and floodplain vegetation classes. Class codes listed in the graph legends can be related to their NVCS association names in Table 6.

(representing potential vegetation patterns mediated by resource gradients) and remote sensing (expressing differences in the color, reflectance, greenness, brightness, and texture of actual vegetation patterns) variables. These input variables were obtained at scales ranging from 2- to 30-meters, with spatial transformations performed to employ intermediate scales of the photogrammetric image data. These scales of remote sensing input data correspond to commonly applied remote sensing data sources, so our approach and results may thus be compared to other studies in which QuickBird (2.4 m), ASTER (15 to 30 m), or Landsat (30 m) data are employed for vegetation classification and mapping. The approach described here represents a valuable data mining method when the types and scales of variables that should be included in a classification model are not known *a priori*, because all of the variables that are hypothesized to be relevant are candidates for input into the statistical classification.

Application of a remote sensing approach similar to that described here (wherein input variables derived from multitemporal data obtained at multiple resolutions are fused to enhance the ability to extract information about vegetation characteristics) should improve the results of vegetation mapping efforts in other areas, especially those located within a temperate deciduous biome (e.g., Huang *et al.*, 1997). Through strategic selection of remote sensing image data obtained during certain time periods, the multitemporal dimension of these data can be exploited to enhance classification accuracies in vegetation studies by capturing contrasts in plant phenology that occur seasonally between different vegetation cover types (Wolter *et al.*, 1995; Lawrence *et al.*, 2006). These contrasts can be particularly acute in temperate forested regions with vegetation communities containing both deciduous and evergreen components. Additionally, it has been shown that fusion of multiresolution remote sensing data yields higher classification results in complex landscapes, with some land-cover classes proving to be more identifiable using coarser resolution data and others more amenable to finer scaled data (Chen and Stow, 2003). The approach developed for the ONSR mapping area represents a method for utilizing the information content present in both the multitemporal and multiresolution dimensions of a remote sensing data set to improve classification accuracy. This approach is transferable to other regions with novel vegetation classes and image data sets through calibration of model inputs using available training data and input variables obtained from the image data obtained for that location.

Given the highly detailed nature of this classification scheme, the overall accuracy (63 percent) obtained for the 49-category NVCS association-level map can be considered quite acceptable. Aggregation of the 49-category map to produce a 33-category map brought substantial improvement to the overall and class-wise accuracies, with an overall accuracy (77.5 percent) that is only just below the 80 percent USGS-NPS vegetation mapping program accuracy standard. Only minor aggregations would be needed to bring the overall accuracy up to this standard. For example, only two class merges, i.e., merging NVCS association 2058 (White Oak-Red Oak-Sugar Maple Mesic Forest) with the "White Oak Forest" community type as represented in the 33-class

TABLE 8. PRODUCER'S AND USER'S ACCURACY OBTAINED FOR AGGREGATED NATURESERVE COMMUNITY TYPES (33-CLASS MAP)

| Community Type | Producers Accuracy % | Producers Validation N | Users Accuracy % | Users Validation N |
|---|---|---|---|---|
| Dolomite Glade/Woodland Complex | 78 | 160 | 66 | 189 |
| Igneous Glade/Woodland Complex | 74 | 90 | 82 | 82 |
| Active Channel/Gravel Bar Complex | 87 | 61 | 95 | 56 |
| Bottomland Forest | 86 | 182 | 76 | 206 |
| White Oak Forest | 70 | 205 | 59 | 243 |
| Mixed Oak-Hickory Forest | 80 | 445 | 84 | 419 |
| Upland Pine and Pine-Oak Forest | 68 | 154 | 70 | 148 |

map and merging the wooded and open old field land-cover classes, would be needed to bring the overall accuracy up to 80.1 percent. Because the Ecological Systems and Community Types used in this study were designed to address management issues as well as to address mapping difficulties arising from community similarities and spatial arrangement issues, no further aggregation of mapping classes was performed beyond the 33-class map, as this would diminish the utility of the vegetation map for resource management planning by the National Park Service. The integration of multiple mapping approaches as described here yielded a vegetation community map with a substantially higher thematic resolution than similar vegetation classification maps produced at this level of overall accuracy, in that the USGS-NPS vegetation mapping program accuracy standard was by-and-large met with a 33-class map compared to the six-class (Joy *et al.*, 2003; 74.5 percent overall accuracy), 11-class (de Colstoun *et al.*, 2003; 82 percent), or 13-class (Hansen *et al.*, 1996; 82 percent) maps produced for other vegetation mapping projects using a CART approach alone.

Implementation of the integrated mapping approach applied in this research should be considered when a very detailed vegetation community map is needed for management or inventory purposes. Segmentation of the ONSR mapping area into more homogenous regions to simplify the classification problem and application of a discriminant analysis classifier to calculate per-pixel probabilities of NVCS association type membership proved to be a good alternative to supervised decision rule classification approaches that are typically applied with remote sensing and ancillary data. While the decision rule classification approach is robust, it lacks an objective input variable selection method. The mapping approach applied here also provided superior results compared to the CART approach applied during pilot investigations, as limitations associated with missing classification output categories were avoided. The CART approach objectively selects input variables and provides a heuristic tool for landscape stratification for complex landscapes, but considers only one input variable at a time for recursive splitting. The multidimensionality of a large set of input variables is therefore not exploited as thoroughly using CART compared to the discriminant analysis approach, in which large amounts of input variables are fused into canonical discriminant functions. Furthermore, spatially continuous representations of class membership probabilities produced through the discriminant analysis approach described here can be used to extend the information contained in a thematic classification map such that spatially continuous representations of uncertainty can be expressed.

## Acknowledgments

The authors wish to dedicate this research to the memory of Daniel E. Tenaglia, whose love for the flora of Missouri guided his life and served to inspire us. We gratefully acknowledge Phil Townsend and Jane Foster for their assistance in the application of the discriminant analysis method. Many thanks also to Jessie Bebb, Sara Bellchamber, Keith Grabner, Yunjeong Jang, Benjamin King, Tim Nigh, Raun Oberman, Jean Perry, Geovanne Stighen, Esther Stroh, Dan Tenaglia, Dan Twyman, and Todd Weakly for their valuable contributions to this work.

## References

Allen, T.R., 2000. Topographic normalization of Landsat Thematic Mapper data in three mountain environments, *Geocarto International*, 15(2):13–19.

Austin, M.P., and T.M. Smith, 1989. A new model for the continuum concept, *Vegetatio*, 83:35–47.

Austin, M.P., and P.C. Heyligers, 1989. Vegetation survey design for conservation: Gradsect sampling of forests in north-eastern New South Wales, *Biological Conservation*, 50,13–32.

Avers, P.E., D.T. Cleland, W.H. McNab, M.E. Jensen, R.G. Bailey, T. King, C.B. Goudey, and W.E. Russell, 1994. *The National Hierarchical Framework of Ecological Units*, ECOMAP, USDA Forest Service, Washington, D.C.

Beers, T.W., P.E. Dress, and L.C. Wensel, 1966. Aspect transformation in site productivity research, *Journal of Forestry*, 64:691–692.

Boutin, C., and B. Jobin, 1998. Intensity of agricultural practices and effects on adjacent habitats, *Ecological Applications*, 8(2):544–557.

Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone, 1984. *Classification and Regression Trees*, Chapman and Hall, New York.

Carleer, A., and E. Wolff, 2005. Exploitation of very high-resolution satellite data for tree species identification, *Photogrammetric Engineering & Remote Sensing*, 70(1):135–140.

Chander, G., and B. Markham, 2003. Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges, *IEEE Transactions on Geoscience and Remote Sensing*, 41(11):2674–2677.

Chastain, R.A., M.A. Struckhoff, K.W. Grabner, E.D. Stroh, H. He, D.R. Larsen, T.A. Nigh, and J. Drake, 2006. *Mapping Vegetation Communities in Ozark National Scenic Riverways-Final Technical Report to the National Park Service*, U.S. Geological Survey Open-File Report 2006–1354, 90 p., plus appendices.

Chen, D., and D. Stow, 2003. Strategies for integrating information from multiple spatial resolutions into land-use/land-cover classification routines, *Photogrammetric Engineering & Remote Sensing*, 69(11):1279–1287.

Clark, M.L., D.A. Roberts, and D.B. Clark, 2005. Hyperspectral discrimination of tropical tree species at leaf to crown scales, *Remote Sensing of Environment*, 96(3–4):375–398.

Crist, E.P., and R.J. Kauth, 1986. The tasseled cap de-mystified, *Photogrammetric Engineering & Remote Sensing*, 8(2):81–86.

Cunningham, B., and C. Hauser, 1992. The decline of the Missouri Ozark forest between 1880 and 1920, (A.R. Journet, and H.G. Spratt, editors), *Proceedings of Towards a Vision for Missouri's Public Forests*, 27–28 March. Southeast Missouri State University, Cape Girardeau, Missouri, pp. 14–19.

De Colstoun, E.C.B., M.H. Story, C. Thompson, K. Commiso, 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(3):316–327.

Federal Geographic Data Committee, 1996. *National Vegetation Classification Standard*, Version 2-Working Draft, FGDC-STD-005, Version 2, FGDC Secretariat, USGS, Reston, Virginia, 62 p.

Frank, T.D., 1988. Mapping dominant vegetation communities in the Colorado Rocky Mountain front range with Landsat Thematic Mapper and digital terrain data, *Photogrammetric Engineering & Remote Sensing*, 54(12):1727–1734.

Franklin, J., 1995. Predictive vegetation mapping: Geographic modelling of biospatial patterns in relation to environmental gradients, *Progress in Physical Geography*, 19(4):474–499.

Gardner, R.H., and D.L. Urban, 2003. Model validation and testing: Past lessons, present concerns, future prospects, (C.D. Canham, J.C. Cole, and W.K. Lauenroth, editors), *The Role of Models in Ecosystem Science*, Princeton University Press, Princeton, New Jersey.

Guyette, R.P., and B.E. Cutter, 1991. Tree ring analysis of fire history in a post oak savanna in the Missouri Ozarks, *Natural Areas Journal*, 11:93–99.

Hansen, M., R. Dubayah, and R. DeFries, 1996. Classification trees: An alternative to traditional land cover classifiers, *International Journal of Remote Sensing*, 17(5):1075–1081.

Hansen, M.C., R. DeFries, J.R.G. Townshend, and R. Sohlberg, 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach, *International Journal of Remote Sensing*, 21:1331–1364.

Hay, G.J., K.O. Niemann, and G.F. McLean, 1996. An object-specific image-texture analysis of H-resolution forest imagery, *Remote Sensing of Environment*, 55:108–122.

Huang, C., B. Wylie, C. Homer, L. Yang, and G. Zylstra, 2002. Derivation of a tasseled cap transformation based on Landsat 7 at-satellite reflectance, *International Journal of Remote Sensing*, 23(8):1741–1748.

Huang, T., Z. Liu, Y. Pan, and Y. Zhang, 1997. Land-cover survey in northeast China using remote sensing and GIS, *Chinese Geographical Science*, 8(3):264–270.

Irish, R.R., 2000. *Landsat-7 Science Data User's Handbook*, Chapter 11: Data Products, Report 430-15-01-003-0, National Aeronautics and Space Administration, URL: http://ltpwww.gsfc.nasa.gov/IAS/handbook/handbook_htmls/chapter 11/chapter11.htm (last date accessed: 07 November 2007).

Ivits, E., B. Koch, T. Blaschke, M. Jochum, and P. Adler, 2005. Landscape structure assessment with image grey-values and object-based classification at three spatial scales, *International Journal of Remote Sensing*, 26(14):2975–2993.

Jacobson, R.B., and K.B. Gran, 1999. Gravel sediment routing from widespread, low-intensity landscape disturbance, Current River basin, Missouri, *Earth Surface and Processes and Landforms*, 24:897–917.

Jensen, J.R., 1996. *Introductory Digital Image Processing: A Remote Sensing Perspective*, Prentice Hall, Upper Saddle River, New Jersey, 316 p.

Joy, S.M., R.M. Reich, and R.T. Reynolds, 2003. A non-parametric, supervised classification of vegetation types on the Kaibob National Forest using decision trees, *International Journal of Remote Sensing*, 24(9):1835–1852.

Karimi, Y., S.O. Prasher, H. McNairn, R.B. Bonnell, P. Dutilleul, and R.K. Goel, 2005. Classification accuracy of discriminant analysis, artificial neural networks, and decision trees for weed and nitrogen stress detection, *Transactions of the ASAE*, 48(3):1261–1268.

Klecka, W.R., 1980. *Discriminant Analysis*, Sage University Paper Series on Quantitative Applications in the Social Sciences, 07–019, Beverly Hills, California, Sage Publications.

Lawrence, R., R. Hurst, T. Weaver, and R. Aspinall, 2006. Mapping prairie pothole communities with multitemporal Ikonos satellite imagery, *Photogrammetric Engineering & Remote Sensing*, 72(2):169–174.

Lillesand, T.M., and R.W. Keifer, 1994. *Remote Sensing and Image Analysis*, Third edition, Wiley, New York.

Markham, B.L., and J.L. Barker, 1986. *Landsat MSS and TM Post-calibration Dynamic Ranges, Exo-atmospheric Reflectance, and At-satellite Temperatures*, Landsat Technical Notes, EOSAT, URL: http://ltpwww.gsfc.nasa.gov/IAS/handbook/pdfs/L5_cal_document.pdf (last date accessed: 07 November 2007).

Mason, S.J., and N.E. Graham, 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation, *Quarterly Journal of the Royal Meteorological Society*, 128:2145–2166.

Meyer, P., K.I. Itten, T. Kellenberger, S. Sandmeier, and R. Sandmeier, 1993. Radiometric corrections of topographically induced effects on Landsat data in an alpine environment, *ISPRS Journal of Photogrammetry and Remote Sensing*, 48(4):17–28.

NatureServe, 2006. *International Ecological Classification Standard: Terrestrial Ecological Classifications*, NatureServe Central Databases Arlington, Virginia, URL: www.natureserve.org (last date accessed: 07 November 2007).

Nigh, T., C. Buck, K. Grabner, J. Kabrick, and D. Meinert, 2000. *An Ecological Classification System for the Current River Hills*, Working Draft, Missouri Department of Conservation, Jefferson City, Missouri.

Nupp, T.E., and R.K. Swihart, 2000. Landscape-level correlates of small-mammal assemblages in forest fragments of farmland, *Journal of Mammology*, 81(2):512–526.

Ohmann, J.L., and M.J. Gregory, 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest neighbor imputation in coastal Oregon, USA, *Canadian Journal of Forest Research*, 32(4):725–741.

Parker, A.J., 1982. The topographic relative moisture index: An approach to soil-moisture assessment in mountain terrain, *Physical Geography*, 3(2):160–168.

Pontius, R.G., and L.C. Schneider, 2001. Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA, *Agriculture, Ecosystems, and Environment*, 85:239–248.

SAS Institute, Inc., 2004. *Statistical Analysis Software* version 8, The SAS Institute, Inc., Cary, North Carolina.

Smith, A., 2001. *Delineation of Glade Complexes in the Southeast Missouri Ozarks Using Multi-temporal Multi-spectral Satellite Imagery*, M.S. thesis, Southern Illinois University.

Strahler, A.H., C.E. Woodcock, and J.A. Smith, 1986. On the nature of models in remote sensing, *Remote Sensing of Environment*, 20(2):121–139.

The Nature Conservancy, 1994. *Field Methods for Vegetation Mapping*, Report to U.S. Department of Interior, The Nature Conservancy, Arlington, Virginia and Environmental Systems Research Institute, Redlands, California, URL: http://biology.usgs.gov/npsveg/field-methods/index.html (last date accessed: 07 November 2007).

The Nature Conservancy Ozarks Ecoregional Assessment Team, 2003. *Ozarks Ecoregional and Conservation Assessment*, Minneapolis, Minnesota, TNC Midwestern Resource Office, 48 p.

Trietz, P., and P. Howarth, 2000. High spatial resolution remote sensing data for forest ecosystem classification: An examination of spatial scale, *Remote Sensing of Environment*, 72(2):268–289.

USGS-NPS, 2006. USGS – NPD Vegetation Mapping Program, URL: http://biology.usgs.gov/npsveg/ (last date accessed: 07 November 2007).

Venables, W.N., and B.D. Ripley, 1994. *Modern Applied Statistics with S-Plus*, Springer Verlag, New York.

Wang, L., W.P. Sousa, and P. Gong, 2004. Integration of object-based and pixel-based classification for mapping mangroves with Ikonos imagery, *International Journal of Remote Sensing*, 25(24):5655–5668.

Warner, T.A., J.Y. Lee, and J.B. McGraw, 1998. Delineation and identification of individual trees in the eastern deciduous forest, *Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry* (D.A. Hill and D.G. Leckie, editors), 10–12 February, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, pp. 81–91.

Wolter, P.T., D.J. Mladenoff, G.E. Host, and T.R. Crow, 1995. Improved forest classification in the Northern Lake States using multi-temporal Landsat imagery, *Photogrammetric Engineering & Remote Sensing*, 61(9):1129–1143.

Wulder, M.A., R.J. Hall, N.C. Coops, and S.E. Franklin, 2004. High spatial resolution remotely sensed data for ecosystem characterization, *Bioscience*, 54(6):511–521.