A Bottom-up Approach to Vegetation Mapping of the Lake Tahoe Basin Using Hyperspatial Image Analysis

Jonathan A. Greenberg, Solomon Z. Dobrowski, Carlos M. Ramirez, Jahalel L. Tuil, and Susan L. Ustin

Abstract

Increasing demands on the accuracy and thematic resolution of vegetation community maps from remote sensing imagery has created a need for novel image analysis techniques. We present a case study for vegetation mapping of the Lake Tahoe Basin which fulfills many of the requirements of the Federal Geographic Data Committee base-level mapping (FGDC, 1997) by using hyperspatial Ikonos imagery analyzed with a fusion of pixel-based species classification, automated image segmentation techniques to define vegetation patch boundaries, and vegetation community classification using querying of the species classification raster based on existing and novel rulesets. This technique led to accurate FGDC physiognomic classes. Floristic classes such as dominance type remain somewhat problematic due to inaccurate species classification results. Vegetation, tree and shrub cover estimates (FGDC required attributes) were determined accurately. We discuss strategies and challenges to vegetation community mapping in the context of standards currently being advanced for thematic attributes and accuracy requirements.

Introduction

Advances in the techniques and technology of hyperspatial image analysis are beginning to narrow the discrepancy between the fields of ground-based forestry and terrestrial remote sensing. With the need for increasingly larger scales of study to understand landscape level processes, pressure has been mounting for more accurate outputs from remote sensing given the large expense associated with field campaigns. However, to date, remote sensing has not approached the degree of accuracy and precision in measuring vegetation that an investigator on the ground achieves. In recent years, the U.S. Forest Service has been developing standards for vegetation classification and mapping using remote sensing imagery (Franklin *et al.*, 2000; USDA, 2002). This paper details novel techniques by which most of the requirements of base-level mapping can be fulfilled using a combination of hyperspatial image analysis including individual plant mapping, automated image segmentation, and vector-raster querying techniques.

A base-level map must contain the following information: FGDC physiognomic classifications of order, class, and subclass; floristic classifications of cover types, dominance types and alliances; total vegetation, tree, and shrub cover classes in increments of 10 percent; and mean tree diameter classes ranging from 0 to 50+ inches. The minimum accuracies of these attributes are 80 percent for physiognomy, 65 percent for floristics, 65 percent for cover class, and 65 percent for mean tree diameter class. The minimum mapping unit (MMU) for a base-level map is defined as "the smallest polygon feature to be mapped at a given map level" (Warbington *et al.*, 2002). We note the use of the term "polygon": the concept of vegetation mapping units is based on the spatial extent of soil orders at different scales (Warbington *et al.*, 2002) and were historically performed by manual digitization of aerial photographs outlining patches of vegetation and soil. We hereafter refer to these polygon mapping units as "patches" consistent with FGDC terminology.

While few base-level maps have been produced to date, coarser thematic-scale maps commonly employ one of two approaches to produce maps: (a) pixel-based classifications of medium- and coarse-scale imagery (ground resolution >1 m, e.g., CALVEG, Parker and Matayas, 1979), and (b) object-oriented analysis (OOA) of (typically) hyperspatial imagery (ground resolution ≤ 1 m, e.g., Lobo *et al.*, 1998). Pixel-based classifications of medium and coarse-scale imagery typically rely on training data of field plots, which have been classified to a given physiognomic or floristic level. Techniques such as maximum likelihood and classification and regression trees (CART) are then used to produce maps with the appropriate classification. Higher taxonomic levels can be mapped by merging the classes into the appropriate coarser floristic or physiognomic level, or by reclassifying the field data into the coarser classes and then re-training the classifier. OOA is a relatively new technique for mapping which consists of two steps: (a) generation of a vector layer of vegetation patches by automated image segmentation algorithms (e.g., Baatz *et al.*, 2003) and (b) classification of patches using spectral and textural data from raster pixels that fall within these patches. OOA techniques are garnering more attention for vegetation

Jonathan A. Greenberg is at the NASA Ames Research Center, MS 242-4, Moffett Field, CA 94035 (jgreenberg@arc.nasa.gov) and formerly at CALSPACE, University of California at Davis.

Solomon Z. Dobrowski, Jahalel L. Tuil, and Susan L. Ustin are with CALSPACE, University of California at Davis, 1 Shields Ave., Davis, CA 95616.

Carlos M. Ramirez is with the USDA Forest Service, Region 5, State and Private Forestry – Forest Health Protection, 3237 Peacekeeper Way, Suite 207, McClellan, CA 95652.

Photogrammetric Engineering & Remote Sensing
Vol. 72, No. 5, May 2006, pp. 581–589.

0099-1112/06/7205–0581/$3.00/0

mapping because they can, to some extent, replicate photointerpretive digitization techniques to produce more realistic patch shapes and sizes than direct pixel-based classifications. Both of these approaches are "top-down" or "direct" classification techniques; they attempt to generate a classification ruleset which directly relates a vegetation class to image spectral data.

Directly classifying vegetation communities using coarse-scale pixels or image objects has a number of major challenges which may prevent accurate base-level mapping. For one, vegetation patches are composed of a complex mixture of vegetation, shadow, ground-cover, and other materials in heterogeneous two- and three-dimensional spatial arrangements. Since rules for classes from dominance type to order are based on *cover*, often across a wide range of cover values, difficulty arises in establishing a relationship between the target vegetation classes and image spectra in mixed communities. For instance, the FGDC (1997) definition of a tree-dominated order is a region in which tree cover is greater than or equal to 25 percent. A region with 20 percent tree and 80 percent shrub would be classified as "Shrub Dominated." A region with 30 percent tree and 70 percent shrub would be classified as "Tree Dominated" as would a region with 80 percent tree and 20 percent shrub. A difference in 10 percent of tree cover is unlikely to consistently produce a significant change in spectral signal, so as tree cover approaches the lower limit of the class; they are frequently misclassified using direct classification techniques. We illustrate this problem in Figure 1.

The floristic levels of alliance pose an even more significant problem which, in some cases, can be impossible to circumvent using optical sensors. Alliances are defined as: "A grouping of associations with a characteristic physiognomy and habitat and which share one or more diagnostic species typically found in the uppermost or dominant stratum of the vegetation" (Jennings *et al.*, 2003). While dominance types and alliances have a similar definition (indeed, dominance types are also referred to as "provisional alliances"), alliances can be defined by subcanopy species. Diagnostic vegetation which is found under a tree or shrub canopy cannot be directly detected by an optical sensor (although it can be inferred by non-remote sensing information).

In addition to the problems of linking spectral information to vegetation classes, the spatial extent of a vegetation patch makes collecting both training and test data difficult. Many medium and coarse-scale pixel analyses use 30 m (e.g., Landsat) or larger pixels. Collecting training or test data for these pixels usually requires subsampling the pixel itself, which may cause errors if the subsampling is not representative of the entire pixel. This problem becomes even more pronounced when using OOA techniques to generate vegetation patches. As seen in Figure 1a through 1c, vegetation patches can be large and irregularly shaped, making subsampling extremely difficult.

Image analysis techniques that utilize hyperspatial imagery can provide a new way to map vegetation at thematic levels from dominance type to division. We approach the problem from a bottom-up perspective: first, we create a pixel-based map of species and non-vegetated cover at high-resolution; second, we define patch boundaries in a separate analysis; and last, we merge the two using rulesets for naming vegetation classes from dominance types to division. This approach allows for: (a) repeatable vegetation mapping over time, (b) flexibility in modifying patch and thematic class definitions, and (c) an easily validated dataset, since individual plants or homogenous species at a 1 m do not require subsampling of the pixels.

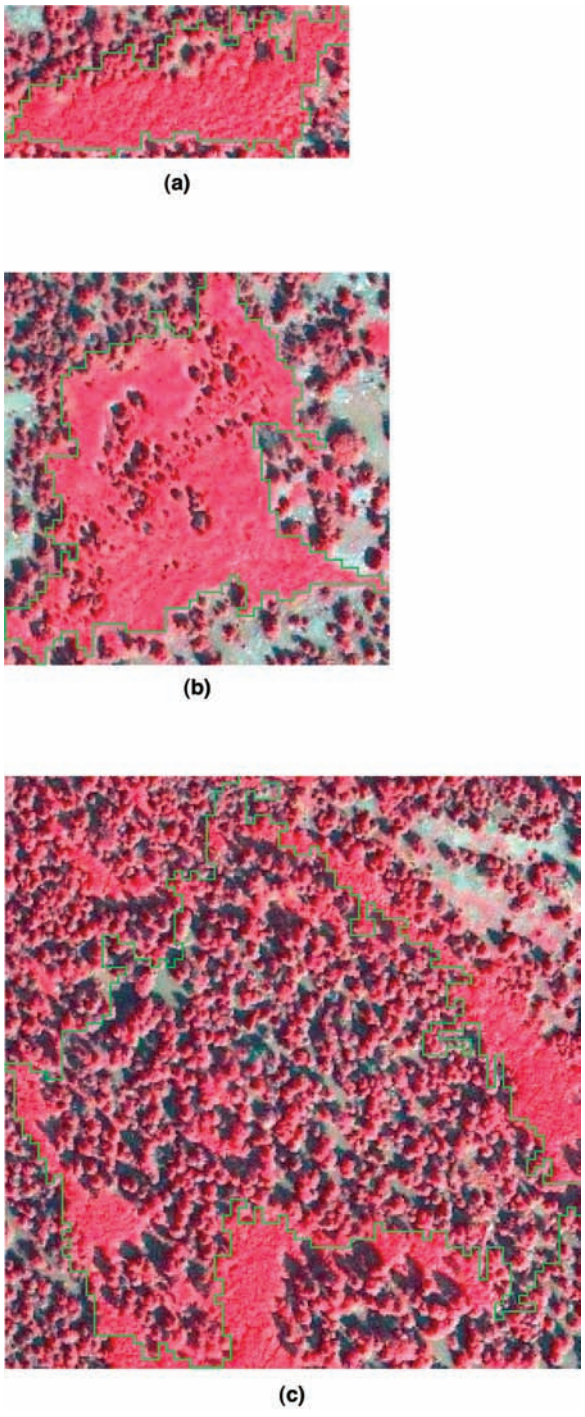


Figure 1. Illustration of problem of relationship between information class and spectral information. (a) has approximately 20 percent tree cover and is labeled "shrub dominated," (b) has approximately 30 percent tree cover and is labeled "tree dominated," and (c) has approximately 80 percent tree cover and is labeled "tree dominated." Spectrally and texturally (a) and (b) are far more similar to one another than (b) and (c).

Methods

Our analysis consists of the following steps: (a) create a pixel level (1 m) raster map of vegetated and non-vegetated

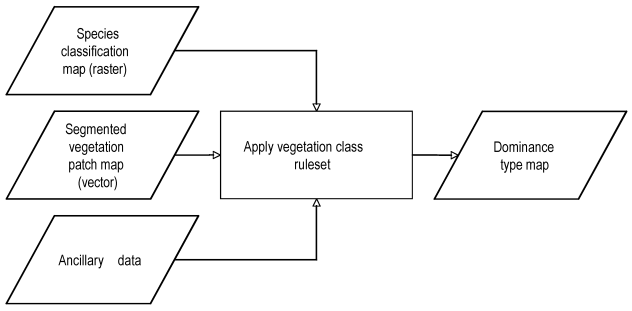


Figure 2. Workflow of the dominance type mapping process.

TABLE 1. TREE/NOT TREE CLASSIFICATION ACCURACY FOR ALL COVER CLASSES USED

Species/Cover Class, Class Code	Fraction of Pixels Near Shadow	Accuracy
Trees		
Abies concolor, ABCO	0.98	98%
Abies magnifica, ABMA	0.92	92%
Juniperus occidentalis, JUOC	0.70	70%
Pinus albicaulis, PIAL	0.80	80%
Pinus contorta, PICO	0.92	92%
Pinus jeffreyi, PIJE	0.93	93%
Pinus monticola, PIMO	0.97	97%
Populus tremuloides, POTR	0.96	96%
Tsuga mertensiana, TSME	0.97	97%
Shrubs		
Alnus incana, ALIN	0.77	77%
Arctostaphylos patula, ARPA	0.08	92%
Artemesia tridentate, ARTR	0.00	100%
Ceanothus cordulatus, CECO	0.09	91%
Ceanothus velutinus, CEVE	0.13	87%
Quercus vaccinifolia, QUVA	0.30	70%
Salix sp., SASP	0.76	76%
Herbs		
Dry Graminoid, DRGR	0.00	100%
Green Graminoid, GRGR	0.05	95%
Carex sp., CASP	0.42	58%
Non-vegetated classes		
Bright Impermeable, BRIM	0.00	100%
Dark Impermeable, DAIM	0.00	100%
Water, WATER	0.00	100%

cover classes at the species level, (b) independently create vector polygons which are used to define the extent of a vegetation patch, (c) generate a formal ruleset that follows FGDC naming conventions for physiognomic classes and the creation of a dominance-type ruleset for floristic mapping, and (d) use the patch geodata as the basis for querying the raster class map using this ruleset. Figure 2 shows an overall workflow of this analysis. Table 1, column 1 contains the cover classes and their respective codes which are referenced in the following sections.

Site Information

The Lake Tahoe Basin falls along the California and Nevada border. The 82,000 ha basin is surrounded by the Carson and Sierra Nevada mountain ranges. Elevation ranges between 1,900 m above sea level (ASL) to 3,400 m ASL, spanning the montane (2,187 m to 2,656 m ASL) and subalpine (>2,656 m ASL) elevation zones. Vegetation types are varied and include diverse meadow and fen habitats, evergreen and deciduous shrublands, and conifer dominated forests. The vegetation communities have experienced significant anthropogenic disturbances, including significant recent development and urbanization within the basin. Moreover, roughly two-thirds of the forests were clear-cut during the latter third of the 19th century (Elliot-Fisk *et al.*, 1997).

Image Data

The imagery used in this analysis was from the Ikonos polar-orbiting sensor. Ikonos has four 4 m ground resolution multispectral bands (blue, green, red, and near-infrared) and one 1 m panchromatic band. Four images swaths were collected in July 2002 that covered the entire Lake Tahoe Basin. The imagery was orthorectified and then radiometrically corrected using (a) topographic shade correction, (b) atmospheric correction, (c) empirical line calibration using ground spectra, and (d) image-to-image normalization. We performed principle components (PC) pan-sharpening (Welch and Ahlers, 1987) on the radiometrically corrected 4 m multispectral images substituting the 1 m panchromatic images as PC band 1, resulting in a set of 1 m multispectral images. The images were then mosaiced. Greenberg *et al.* (2005) describes the preprocessing of this image dataset in more detail. All subsequent analyses with the exception of the image segmentation were performed on the 1 m pan-sharpened multispectral image. The image segmentation was performed on the 4 m multispectral image.

In order to assist classification, grey-level co-occurrence matrix (GLCM) texture images were produced using the mosaicked near-infrared band. Eight texture images were produced using the mean, contrast, entropy, and angular second moment features (Haralick *et al.*, 1973) with 3 m and 9 m windows. All GLCM calculations were performed using 32 grey-level quantization levels within ENVI® image analysis software (Research Systems Inc., Boulder, CO).

Following the production of the texture images, a principle components transform was conducted on the 12 image bands (4 pansharpened spectral bands and the 8 GLCM texture images). The transform was conducted in order facilitate classification by reducing the effects of collinearity between image bands and to improve the multivariate normality of the classification vectors used in the maximum likelihood algorithm. Eigenvalues and eigenvectors were calculated from the entire mosaicked image using the covariance matrix.

Field Data

Training data for image classification and map accuracy test data were acquired for 19 vegetation classes representing prevalent species or genera, and three non-vegetated classes including water, dark impermeable surfaces (volcanic substrates, asphalt), and bright impermeable surfaces (granite, concrete). The vegetation classes included nine tree species, six shrub species and one shrub genera, and three herbaceous classes. Field data were collected using differential GPS linked in real-time to geo-referenced color-infrared Ikonos imagery using SOLO Field software (Tripod Data Systems Inc., Corvallis, OR). The goal of the field data collection was to identify individual trees and patches of shrub and herbaceous species. This was accomplished by using in-field digitization techniques. GPS was used to navigate to a cover class target visually identified both within the image and on the ground. A pixel point was manually selected on the imagery, vectorized, and converted to a point shapefile. This technique allowed for much higher spatial precision in pixel selection since there was no reliance on GPS and image orthorectification accuracy.

Approximately 1,700 point features were delineated representing individual tree crowns, patches of homogenous

vegetation, and non-vegetated regions. Points were taken across a wide range of elevations and geographic regions within the basin to capture intra-class variability. Additionally, the geomorphology of the basin allowed for several sampling forays that traversed elevation profiles thus maximizing exposure to different vegetation classes.

Along with the point vector data, crown radius and diameter-at-breast-height (DBH) were collected and recorded for all tree species. Non-tree cover classes were assigned a patch radius (the largest circle which could be inscribed in the particular patch). In cases where trees appeared as multi-stemmed individuals (common in subalpine species), crown radius was recorded, but not DBH. In cases where trees could not be distinguished as individuals, but the cluster of trees was homogenous in species composition, patch radius and species was recorded (common in hardwoods and dense stands of small trees). Both crown radius and patch radius was then subsequently used in a GIS to create a circular polygon buffer around each point feature by multiplying the recorded radius by 75 percent. This ensured that pixels found within the buffer were in the cover class in question. In summary, this data was used to create a shapefile with approximately 1,700 polygon features that was subsequently used for image classification and accuracy assessment.

Species/Lifeform Classification Map

Vegetation classification was performed at two thematic resolutions: lifeform and species. Lifeform classification allowed "access" to all FGDC hierarchical levels from Division to Class. Species cover was used for lower-level community classifications from subclass to dominance type, and used for lifeform classification to reduce the classification complexity. Figure 3 shows a flowchart of the pixel-level classification process, which is described below.

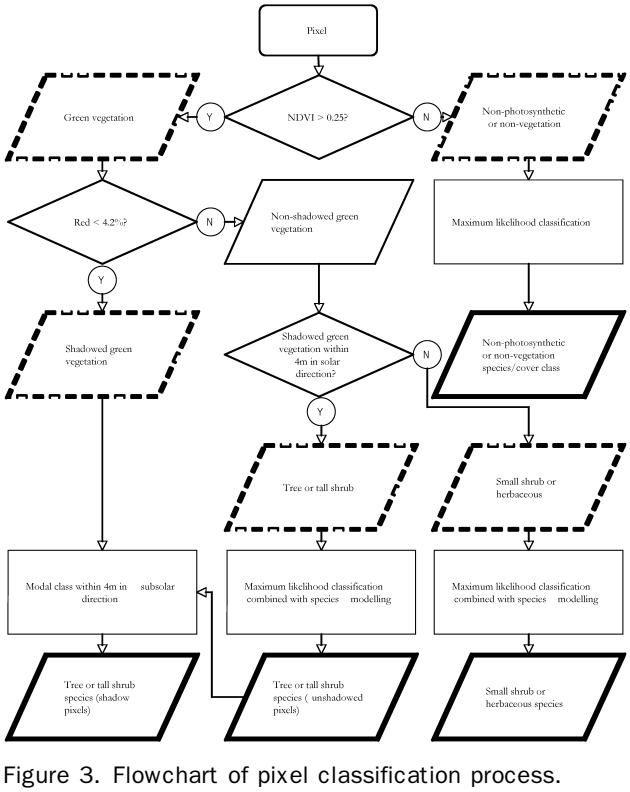


Figure 3. Flowchart of pixel classification process. Hyphenated outputs are used for the lifeform level classification, and boldfaced outputs are used for species level classification.

Green Vegetation versus Non-photosynthetic Vegetation and Non-vegetation Classification

As detailed in Greenberg *et al.* (2005), we first split the lifeforms into green vegetation (GV) and non-photosynthetic vegetation/non-vegetation (collectively, NV) classes using an NDVI threshold. Brewer *et al.* (2004) define a vegetated region as being >1 percent (we slightly modify the published definition of non-vegetation to be ≤1 percent as opposed to the published <1 percent). The threshold was determined by regressing aerial cover estimates from plots acquired from local resource agencies against the mean normalized difference vegetation index (NDVI) values for pixels found in these plots. The regression yielded:

$$\text{NDVI} = 0.0033 \times \% \text{ aerial cover} + 0.2511, \text{ RMSE } 0.15$$

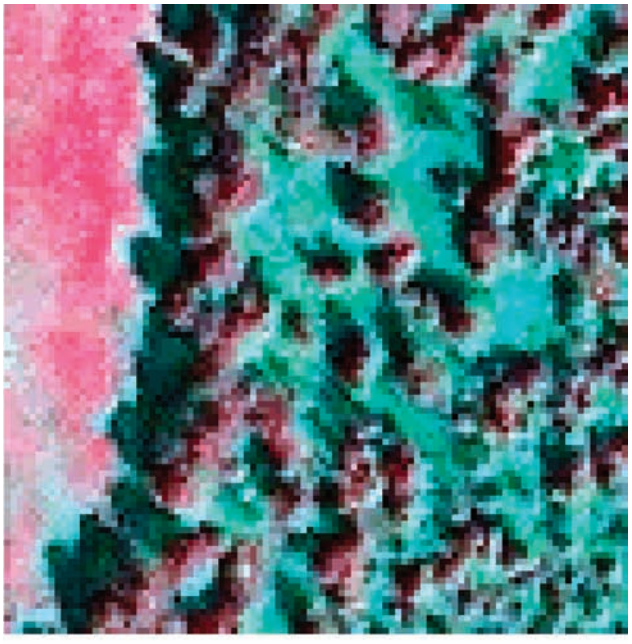
Substituting in 1 percent aerial cover, we found an NDVI threshold of 0.25, above which a pixel was considered GV, and below which it was considered NV.

Tree versus Non-tree Green Vegetation

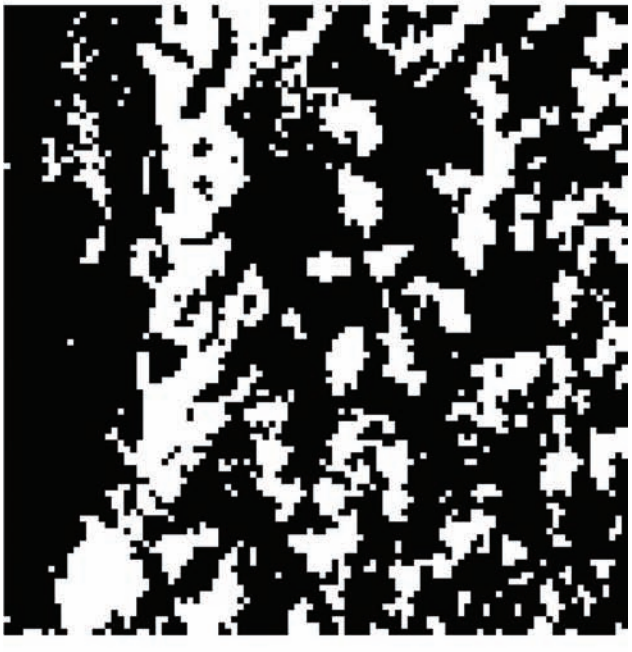
The second step was to differentiate trees and tall shrubs (hereafter all are referred to as "trees") from all other vegetation. At 1 m ground resolution, trees and tall shrubs are visually identifiable as multi-pixel objects. Currently, there are two methods to identify a cluster of pixels as being components of a crown: direct classification using standard algorithms and image segmentation techniques to vectorize individual tree crowns into unique polygons. Initial analysis and results from other studies suggested individual pixel classifications are fraught with errors, as a single tree crown can have more spectral variation than between species (Leckie *et al.*, 1992). Segmentation of tree crowns has been found to be useful in these cases, but previous studies suggested that the spatial resolution was too coarse to accurately delineate the borders of tree crowns. Using a novel approach, we isolated tree crown pixels from pixels belonging to other cover classes based on significant differences in shadows. Crown pixels, therefore, can be defined as those pixels that are within a certain distance from a crown shadow pixel in the solar direction.

To define a crown shadow, we follow the technique of Greenberg *et al.* (2005, see Figure 4) to determine a red band threshold that differentiates "shadow" from "not shadow." We calculated the mean + 2 standard deviations (stdev) of red reflectance (=4.2 percent reflectance) from 100 randomly chosen pixels that fell in a crown shadow. Values higher than this threshold were classified as "not shadow" and those below the threshold were classified as "shadow." To separate dark, non-vegetated pixels (such as asphalt and water) from vegetation shadows, we joined the shadow/not-shadow mask with the vegetation/not-vegetation mask previously described. A crown shadow, therefore, is defined as a pixel having a red reflectance of <4.2 percent and an NDVI >0.25. A mask was created using this rule, so crown shadows were assigned a value of 1 and all other pixels a value of 0.

A filter was designed such that all pixels that were within the approximate mean radius of trees found in the basin (4 m) in the solar direction were defined as being a part of a tree crown. We then applied the moving filter shown in Figure 5 to the crown shadow mask image. If a pixel had a value of >0.0 (it contained at least one shadowed vegetation pixel in the solar direction), it was classified as a crown. If the resultant filtered pixel had a value of 0.0, it was not near a crown shadow and was classified as a non-crown.



(a)



(b)

Figure 4. From Greenberg et al. (2005). (a) is a color infrared image subset of a wet meadow to forest transition, and (b) is the resultant mask showing shadowed green vegetation (white).

0.077	0.077	0	0	0	0	0	0	0
0.077	0.077	0.077	0	0	0	0	0	0
0	0.077	0.077	0.077	0	0	0	0	0
0	0	0.077	0.077	0.077	0	0	0	0
0	0	0	0.077	**0.077**	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Figure 5. Moving window used to determine if a pixel is near shadowed vegetation (and therefore is part of a tree). The center cell (boldfaced) indicates the reference pixel.

Species Classification

Training and test data was extracted from the 12 band PCA image using STARSPAN (Rueda *et al.*, 2004) and the vector coverage developed from the field sampling data (see above). For each pixel located within one of the 1,700 train and test vector polygon features (>90,000 pixels in total), the following data was recorded in a database: (a) the feature ID of the polygon that the pixel intersected, (b) the map class associated with the polygon, and (c) all 12 PCA image band values. A multivariate outlier analysis was performed on the dataset using a jackknife technique with a Mahalanobis distance statistic. The Mahalanobis distance was calculated for each observation with estimates of the mean, standard deviation, and correlation matrix that did not include the observation itself. Observations falling outside of the 99 percent quantile were dropped from the analysis. This dataset was then split into training and test data by stratifying the polygons by map class and then randomly assigning 60 percent of the polygon features (and their respective pixels) to a training set and the remaining 40 percent to a test set. Pattern vectors for classification were calculated from this training data set for the 22 map classes.

We conducted Gaussian maximum likelihood classification (MLC) using all 12 PCA bands for feature vectors, resulting in a 22 band likelihood image with each band corresponding to one of the 22 map classes. The 22 band likelihood image was spectrally subset by the 19 bands representing vegetation classes. For this image, we assumed that all pixels belonged to one of the described vegetation classes, and as such, normalized the likelihood values to sum to 1. This resulted in a maximum likelihood probability image for each species/cover class. Species/land-cover classification was performed by dividing the entire set of classes into three categories: NV, non-crown vegetation and trees/tall shrubs. Figure 3 details the decision tree used to perform the classification.

NV classification was performed by choosing the highest ML value from the four NV classes: water, bright impermeable, dark impermeable, and dry graminoids. In the case of a tie, the pixel was labeled "unclassified."

The non-tree vegetation and tree branches of the ML classification were combined with species prediction surfaces developed using general additive modeling (GAM) using a consensus theoretic approach. Dobrowski *et al.* (2006) describe in detail the production of the species prediction

surfaces using GAM, and the process by which the ML/species prediction surfaces were weighted.

Lifeform Classification
We recognized four lifeforms in our analysis: non-vegetation, trees, shrubs, and herbaceous vegetation. The fifth lifeform recognized by the FGDC, nonvascular plants, did not appear in sufficiently large covers to warrant inclusion in the classification. Each pixel was assigned a lifeform as follows: if the pixel was non-vegetated, it was classified as non-vegetation unless it was determined to be dry graminoid, in which case it was classified as herbaceous. If the pixel was part of a crown, it was classified as a tree unless the species classification was *Alnus incana* or *Salix* sp., in which case the pixel was classified as a shrub. If the pixel was found to be vegetated, but not part of a tree or tall shrub crown, it was assigned to the shrub or herbaceous class depending on the particular species designation.

Accuracy Assessment
We used the 40 percent of remaining field collected data to perform an accuracy assessment on the decision tree. We determined the accuracy for the vegetation/non-vegetation split, crown/non-crown vegetation, species classification and lifeform classification. A confusion matrix was generated and user's, producer's, overall, and Kappa accuracies were calculated.

Vegetation Patch Map
Automated image segmentation was conducted on the Ikonos 4 m resolution mosaiced image. We used a region-based segmentation algorithm (Baatz *et al.*, 2003) in which a seed of pixels are grown into an object, continuing until a homogeneity defined threshold was reached (see Benz *et al.*, 2004 for further details). The selection of the parameters that define this threshold were iteratively determined using subsets of the imagery and a qualitative assessment by an analyst (scale = 160, shape = 0.63, smoothness/compactness = 0.7). This resulted in a vector layer of approximately 21,000 polygons that were spatially continuous over the entire study area and represented vegetation patches identified primarily on the basis of unique textural and tonal information.

Attributing Vegetation Patches With Species Cover
Despite the need for providing efficient attributing of vector geodata from raster data, no software package was capable of dealing with a dataset the size of the one generated. As such, we developed a program, STARSPAN (Rueda *et al.*, 2004), which quickly attributes a vector layer with either raw or summarized pixel data for a given polygon's extent. For each vegetation patch, we determined the pixel coverage of each species/cover class used in the analysis. From this, we generated relative cover of each species/cover class, as well as relative physiognomic cover (tree, shrub, herbaceous and non-vegetated). In addition, we determined the mean elevation for each patch for use in distinguishing subalpine from montane dominance types.

Applying Rulesets to Generate Vector Based Vegetation Maps

Division to Subclass
Using these patch data, we followed the FGDC existing vegetation hierarchy down to the level of subclass. Division, order and class require relative lifeform covers (tree, shrub, herbaceous, and non-vegetated). We found no evidence of large areas of nonvascular vegetation cover, and no cover class was identified in the field of this lifeform, so the nonvascular vegetation class was not used. Subclass requires two additional pieces of information: determining whether a pixel is evergreen or deciduous for tree and shrub dominated classes, and the inundation status of herbs. These data were extracted from our species raster layer.

Dominance Type
Dominance types have a range of definitions, and no official ruleset had been generated for the basin. Jennings *et al.* (2003) refers to a dominance type as "one or more species which are usually the most important ones in the uppermost or dominant layer of the community, but sometimes of a lower layer of higher coverage." To formalize this definition, we generated the following ruleset by which to generate a dominance type name. First, the dominance status of the dominant lifeform for each vegetated polygon was determined (e.g., the FGDC Division). For each polygon the following attributes were used (note: we did not include all of these attributes in the final database):

- Relative cover rank (e.g. percent covered by Jeffrey Pine divided by total tree cover, or percent covered by Huckleberry Oak divided by total shrub cover) for each species used in the analysis.
- Mean elevation for each polygon.

If the top ranked species (by relative cover) of the dominant life form has >10 percent cover than the second ranked species, the polygon is *single-species dominant*. If the top two ranked species have less than a 10 percent difference in relative cover, but rank 2 has a relative cover >10 percent higher than rank 3, the polygon is dual-species dominant. If the top three ranked species are within 10 percent relative cover of one another, the polygon is labeled mixed-dominant. Single species dominated polygons are given the dominant species as the label. Dual species dominated polygons are given the top two ranked species (alphabetically ordered and hyphenated) as the label. Mixed dominant polygons were assigned the subclass name with an elevation modifier: "montane" if the mean elevation for that polygon is <7,000 ft. (2,133.6 m), "subalpine" if mean elevation ≥7,000 ft. (2,133.6 m). For tree-dominant polygons, we substituted the word "conifer" for "evergreen."

Polygons with no dominant lifeform were labeled "No Dominant Life Form." Non-vegetated polygons were labeled "Water" if water ≥50 percent of the cover, "Impermeable" otherwise.

Results

Species/Lifeform Classification Accuracy

The NDVI threshold resulted in an accurate GV/NPV split: no non-vegetation pixels were labeled vegetation and only a single vegetation pixel out of 32,016 pixels was labeled non-vegetation. The proximity to vegetation filter resulted in ≥90 percent tree/not-tree accuracy for 15/22 classes, and ≥70 percent accuracy for every class except for the inundated herbaceous class, which was 58 percent accurate (see Table 1 for a summary of results). Lifeform overall accuracy was 86.2 percent, kappa = 0.81 (see Table 2). Producer's accuracy for lifeform ranged from 73.8 percent (shrubs) to 97.1 percent (non-vegetated). User's accuracy ranged from

TABLE 2. LIFEFORM ACCURACY RESULTS: OVERALL ACCURACY 86.2; KAPPA 0.81

Life Form	Producer's	User's
Non-vegetated	97.1%	100.0%
Trees	81.3%	94.2%
Shrubs	73.8%	47.6%
Herbs	82.0%	92.6%
Average	83.6%	83.6%

TABLE 3. COVER CLASS ACCURACY

Species/Cover Class, Class Code	Producer's Accuracy	User's Accuracy
Trees		
Abies concolor, ABCO	69.2%	52.4%
Abies magnifica, ABMA	36.3%	29.5%
Juniperus occidentalis, JUOC	9.6%	38.3%
Pinus albicaulis, PIAL	27.1%	54.9%
Pinus contorta, PICO	32.6%	35.6%
Pinus jeffreyi, PIJE	47.4%	48.9%
Pinus monticola, PIMO	21.0%	20.8%
Populus tremuloides, POTR	44.5%	70.4%
Tsuga mertensiana, TSME	18.2%	36.7%
Shrubs		
Alnus incana, ALIN	31.3%	93.5%
Arctostaphylos patula, ARPA	74.6%	26.3%
Artemesia tridentate, ARTR	65.0%	27.0%
Ceanothus cordulatus, CECO	47.6%	11.4%
Ceanothus velutinus, CEVE	47.3%	70.3%
Quercus vaccinifolia, QUVA	57.4%	8.8%
Salix sp., SASP	37.5%	32.9%
Herbs		
Dry Graminoid, DRGR	84.9%	68.4%
Green Graminoid, GRGR	80.4%	27.0%
Carex sp., CASP	63.8%	98.8%
Non-vegetated classes		
Bright Impermeable, BRIM	97.3%	98.9%
Dark Impermeable, DAIM	80.0%	99.8%
Water, WATER	100.0%	100.0%
Average	53.3%	52.3%

47.6 percent (shrubs) to 100 percent (non-vegetated). Mean Producer's accuracy for species was 53.3 percent, and User's accuracy or 52.3 percent. Producer's accuracy for species ranged from 9.6 percent (JUOC) to 84.9 percent (DRGR). User's accuracy for species ranged from 8.8 percent (QUVA) to 98.8 percent (CASP). Table 3 details species accuracy results and Table 4 details the confusion matrix. We note that accuracy results presented here are different than those published in Dobrowski *et al.* (2006) as they pertain to a different stage of the workflow.

Vegetation Cover

Tree dominance types accounted for 67 percent of the total vegetated area, followed by shrub dominated (31 percent), not dominant (1 percent) and herb dominated (1 percent). In total, 67 vegetation dominance types were identified. Of these, 17 account for 90 percent of the total vegetated area of the basin. Jeffrey Pine Forest, White Fir Forest, and Red Fir Forest are the most common dominance types found in the basin, and account for over 50 percent of the total vegetated area. Table 5 lists the dominance type names and their covers, ranked by most coverage to least coverage.

Map and Map Attributes

The resulting Tahoe Basin Existing Vegetation Map (TBEVM) contained nearly all base level attributes required to meet FGDC mapping standards. All FGDC physiognomic levels, dominance type, and vegetation, tree and shrub cover could be generated from these techniques. Greenberg *et al.* (2005) describe the generation of mean DBH classes for all vegetation patches. In addition to the required attributes, TBEVM also contains species cover, herbaceous and non-vegetated cover,

TABLE 4. COVER CLASS CONFUSION MATRIX (UNITS ARE NUMBERS OF PIXELS). COLUMNS ARE MAP CLASSES, AND ROWS ARE GROUND CLASSES

Class	ABCO	ABMA	ALIN	ARPA	ARTR	CECO	CEVE	DRGR	GRGR	JUOC	PIAL	PICO	PIJE	PIMO	POTR	QUVA	SASP	TSME	CASP	BRIM	DAIM	WATER
ABCO	985	116	73	0	1	0	2	0	1	131	32	187	133	37	94	0	34	41	13	0	0	0
ABMA	94	315	0	0	0	0	0	0	0	120	76	136	107	107	16	0	2	86	10	0	0	0
ALIN	0	0	546	0	0	2	0	0	0	0	0	0	0	0	0	0	36	0	0	0	0	0
ARPA	10	3	0	349	2	37	102	0	10	6	43	4	5	1	9	11	0	0	737	0	0	0
ARTR	0	1	0	13	445	29	66	289	0	7	2	0	0	0	0	0	0	0	739	50	6	0
CECO	16	16	3	24	42	139	29	74	72	70	0	29	43	10	32	27	50	5	536	0	0	0
CEVE	0	2	0	2	3	17	353	0	0	9	28	3	15	1	0	30	29	5	5	0	0	0
DRGR	0	0	0	0	131	10	0	3024	3	7	0	0	0	0	0	0	0	0	1247	0	0	0
GRGR	0	0	566	6	32	0	62	15	842	14	0	0	20	0	0	5	285	0	1261	0	0	0
JUOC	13	17	0	0	0	0	0	0	0	90	12	25	27	42	2	0	0	4	3	0	0	0
PIAL	0	9	0	0	0	0	0	0	0	11	178	8	0	10	0	0	0	99	9	0	0	0
PICO	49	119	0	0	0	0	0	0	0	61	17	342	135	64	31	5	2	104	32	0	0	0
PIJE	145	56	0	0	0	1	1	0	0	23	32	109	657	39	116	1	2	7	154	0	0	0
PIMO	8	65	0	0	0	0	0	0	0	109	52	76	21	119	7	1	0	115	0	0	0	0
POTR	5	17	34	1	0	1	0	0	0	7	25	16	2	17	387	2	35	0	1	0	0	0
QUVA	79	60	171	70	4	44	67	4	51	180	112	30	99	24	68	147	124	40	298	0	0	0
SASP	0	0	349	1	0	11	43	0	68	2	10	27	39	2	106	25	360	4	47	0	0	0
TSME	0	49	0	0	0	0	0	0	0	28	20	22	0	85	0	0	0	118	0	0	0	0
CASP	8	2	0	0	0	0	21	0	0	2	7	10	50	1	1	2	2	2	9131	0	0	0
BRIM	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2853	31	0
DAIM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1947	4
WATER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	53939

TABLE 5. DOMINANCE TYPES OF THE LAKE TAHOE BASIN ORDER BY COVER

Dominance Type	Cover (hectares)
Jeffrey Pine	15346
White Fir	14927
Red Fir	14146
Sagebrush Scrub	10215
Huckleberry Oak	5250
Mixed Subalpine Conifer	1864
Lodgepole Pine	1725
Greenleaf Manzanita	1655
Non-inundated Herbaceous	1380
Willow Scrub	1324
Jeffrey Pine-Red Fir	1299
Mixed Subalpine Evergreen Shrubland	1269
Whitebark Pine	1142
Quaking Aspen	1083
Huckleberry Oak-Sagebrush Scrub	923
Sparse vegetation	837
Huckleberry Oak-Mountain Whitethorn	702
Lodgepole Pine-Red Fir	679
Mixed Montane Evergreen Shrubland	531
Mixed Montane Conifer-Deciduous	427
Jeffrey Pine-Lodgepole Pine	379
Mountain Whitethorn-Sagebrush Scrub	372
Tobacco Brush	341
Jeffrey Pine-Quaking Aspen	334
Western White Pine	289
Red Fir-White Fir	288
Lodgepole Pine-White Fir	165
Mixed Montane Conifer	156
Mountain Whitethorn	145
Lodgepole Pine-Western White Pine	144
Mountain Hemlock	141
Huckleberry Oak-Tobacco Brush	130
Mixed Subalpine Conifer-Deciduous	123
Inundated Herbaceous	114
Red Fir-Whitebark Pine	113
Red Fir-Western White Pine	103
Greenleaf Manzanita-Huckleberry Oak	101
Mountain Juniper-Red Fir	97
Quaking Aspen-Red Fir	94
Lodgepole Pine-Whitebark Pine	86
Jeffrey Pine-White Fir	84
Sagebrush Scrub-Tobacco Brush	72
Mountain Alder	53
Greenleaf Manzanita-Mountain Whitethorn	34
Huckleberry Oak-Willow Scrub	34
Mountain Hemlock-Western White Pine	34
Mixed Herbaceous	32
Mixed Subalpine Evergreen-Deciduous Shrubland	30
Quaking Aspen-White Fir	28
Mixed Montane Evergreen-Deciduous Shrubland	27
Mountain Juniper	26
Jeffrey Pine-Mountain Juniper	19
Western White Pine-White Fir	18
Lodgepole Pine-Mountain Juniper	15
Western White Pine-Whitebark Pine	14
Tobacco Brush-Willow Scrub	13
Mountain Hemlock-White Fir	13
Mountain Hemlock-Red Fir	12
Mountain Alder-Willow Scrub	12
Mountain Whitethorn-Willow Scrub	8
Greenleaf Manzanita-Sagebrush Scrub	8
Huckleberry Oak-Mountain Alder	7
Lodgepole Pine-Quaking Aspen	7
Mountain Juniper-Whitebark Pine	6
Lodgepole Pine-Mountain Hemlock	5
Greenleaf Manzanita-Tobacco Brush	4
Mountain Juniper-Western White Pine	2

mean crown size, basal area, and aboveground biomass (see Greenberg *et al.*, 2005), and crosswalks to other major vegetation classifications including CALVEG (Parker and Matayas, 1979) and California Wildlife Habitat Relationships (CWHR, Mayer and Laundenslayer Jr., 1988). TBEVM (version 4.1 at the time of this publication) is hosted under the Open Content License at **http://casil.ucdavis.edu/projects/ tbevm**.

Discussion

We have demonstrated the ability of hyperspatial remote sensing data to produce vegetation maps at different thematic scales (from dominance type to division) by fusing species mapping, automated vegetation patch generation, and intelligent querying. A focus on species-level cover mapping using hyperspatial image analysis has several strengths over a direct vegetation community classification techniques:

1. Bottom-up vegetation class mapping provides a flexible framework in which changes in the vegetation classification rules, be it cover class percentages or the size/shape of stands, do not require a time-consuming re-classification. Instead it requires a requerying/attributing step which can be performed rapidly.
2. There are frequently a smaller number of cover species than dominance types, so the classification problem at this thematic scale is reduced.
3. Hyperspatial data allows for complex spectral and textural relationships to be decoupled (e.g., removing shadows from a spectral analysis).
4. Species-to-dominance type mapping can provide a larger, continuous dataset of community attributes (e.g., percent cover by species) in addition to the single dominance type designation.
5. Ecological information characterizing species distributions can be incorporated into the classifier, whereas modeling vegetation classes is problematic.
6. Accuracy assessment is far easier to perform on a 1 m^2 of species cover than it is in a large, irregularly spaced dominance type polygon.

Hyperspatial data appears to be giving accurate estimations of lifeform cover, but species classifications remain problematic. As discussed in Dobrowski *et al.* (2006), the incorporation of modeled species distribution data can improve classification accuracy substantially, but there remains significant error at this level of spatial and thematic resolution. This likely stems from several issues:

1. Complex phenological characteristics of herbs and shrubs which can cause significant variation in spectral signals within a single species.
2. Variations in vegetation density (e.g., LAI and biomass), the spectral effects of which will be pronounced in herbaceous and shrub cover classes, and
3. Plant crown variation in density and shadowing, particularly in trees and large shrubs.

In addition, species level information tends to overpredict common species and underpredict rare species. These problems will have profound effects on the accuracy of our dominance type layer, and need to be solved in order to better map floristic classes. Despite our approach to subsetting the image using shadow proximity to define tree pixels, the approach remained a pixel-level classification. A more object-oriented approach to classifying tree crowns is likely to significantly improve classification (e.g., Gougeon, 1995). In addition, hyperspatial sensors with higher spectral resolution such as CASI, Hymap, and SpecTIR may allow for more accurate species classifications to be made than lower spectral resolution sensors such as Ikonos and Quickbird.

Accuracy assessment of hyperspatial pixels is relatively simple to perform, since subsampling is not needed to identify a 1 m^2 region. With that said, an accuracy assessment of the vegetation community classes is more complicated, and techniques to link species level accuracies and

vegetation community class accuracies are needed. We can make the statement that if our species layer was 100 percent accurate, by definition all vegetation classes based on querying this layer would be 100 percent accurate. We have begun to develop techniques to simulate vegetation class accuracies based on these inputs. In addition to vegetation class accuracy, there remain issues of whether or not our patch layer is accurate: is our segmentation method producing patches of more or less homogenous vegetation? Natural communities do not necessarily have clean spatial divisions (ecotones being an extreme example of this), so homogenous patches of repeatable composition may not truly exist in a given landscape. More techniques need to be developed to determine if the patches represent "ecologically meaningful" landscape units.

Hyperspatial data has unique attributes that separates it from coarser level image products in that individual plants can be readily identified as unique objects. This creates an ability to directly map individual plants (or small homogenous patches of smaller plants), rather than classifying complex mixtures of plants of different species and structure, ground cover, and shadow. Dominance type and coarser level community designations can be directly generated from species maps. In addition, species maps provide a wealth of information including vegetation, tree and shrub cover, and species distributions (which can be related to continuous climate maps). Using more complex individual tree crown recognition techniques (e.g., Gougeon, 1995; Brandtberg and Walter, 1998), tree crown attributes such as size and shape can be determined on an individual tree basis. In Greenberg *et al.* (2005) we demonstrated that we can link crown level information to tree characteristics that can not be directly measured such as DBH, leaf area index (LAI), and biomass. The parameters, particularly biomass, allow for a direct link between hyperspatial image analysis outputs and regional climate models.

Currently, FGDC attribute and accuracy requirements for base-level maps have not been fulfilled by any published effort. One issue is that alliances, as they are currently defined, pose an essentially insurmountable problem to mapping: the inability to remotely detect subcanopy vegetation which may be part of an alliance definition. We propose that base-level mapping should target dominance types as their finest floristic unit, rather than alliances, as dominance types are restricted to canopy species which can be directly detected through remote sensing. Hyperspatial image analysis techniques can provide improvements over previous attempts to perform base level mapping, but there remains significant uncertainty in the products that can lessen their usefulness to managers. We see several research topics which are likely to improve these mapping efforts: (a) improved species mapping using OOA and hyperspectral techniques, (b) formalization of vegetation patch definitions and validation of these patches, (c) development of vegetation mapping rulesets which are informed by the remote sensing community, (d) allometric links between canopy attributes and attributes which cannot be directly estimated (e.g., DBH, LAI, and biomass), and (e) continuing research into the integration of ecological information into mapping efforts.

References

Baatz, M., U. Benz, M. Dehghani, A. Höltje, P. Hofmann, I. Lingenfelder, M. Mimler, M. Sohlbach, M. Weber, and G. Wilhauck, 2003. *eCognition User Guide 3*, Definiens Imaging GmbH, München.

Benz, U.C., P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen, 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information, *ISPRS Journal of Photogrammetry and Remote Sensing*, 58:239–258.

Brandtberg, T., and F. Walter, 1998. Automated delineation of individual tree crowns in high spatial resolution aerial images by multiple-scale analysis, *Machine Vision & Applications*, 11:64–73.

Brewer, C.K., B. Schwind, R.J. Warbington, W. Clerke, P.C. Krosse, L.H. Suring, and M. Sehanta, 2004. Section 3: Existing vegetation mapping protocol (R.J. Brohman and L.D. Bryant, editors), *Existing Vegetation Classification and Mapping Technical Guide*, USDA Forest Service, Washington, D.C.

Dobrowski, S.Z., J.A. Greenberg, C.M. Ramirez, and S.L. Ustin, 2006. Improving image derived vegetation maps with regression based distribution modeling, *Ecological Modelling*, 192(1–2):126–142.

Elliot-Fisk, D.L., T.A. Cahil, O.K. Davis, L. Duan, C.R. Goldman, G.E. Gruell, R. Harris, R. Kattlemann, R. Lacey, D. Leisz, S. Lindstrom, D. Machida, R.A. Rowntree, P. Rucks, D.A. Sharkey, S.L. Stephens, and D.S. Ziegler, 1997. *Lake Tahoe Case Study*, University of California, Center for Water and Wildland Resources, Davis, California.

FGDC, 1997. *Vegetation Classification Standard: FGDC-STD-005*, U.S. Geological Survey, Reston, Virginia.

Franklin, J., C.E. Woodcock, and R. Warbington, 2000. Digital vegetation maps of forest lands in California: Integrating satellite imagery, GIS modeling, and field data in support of resource management, *Photogrammetric Engineering & Remote Sensing*, 66(10):1209–1217.

Gougeon, F., 1995. A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images, *Canadian Journal of Remote Sensing*, 21:274.

Greenberg, J.A., S.Z. Dobrowski, and S.L. Ustin, 2005. Shadow allometry: Estimating tree structural parameters using hyperspatial image analysis, *Remote Sensing of Environment*, 97:15–25.

Haralick, R.M., K. Shanmugam, and I. Dinstein, 1973. Texture feature for image classification, *IEEE Transactions on Systems, Man, and Cybernetics*, 2:610–621.

Jennings, M., O. Loucks, D. Glenn-Lewin, R. Peet, D. Faber-Langendoen, D. Grossman, A. Damman, M. Barbour, R. Pfister, M. Walker, S. Talbot, J. Walker, G. Hartshorn, G. Waggoner, M. Abrams, A. Hill, D. Roberts, and D. Tart, 2003. *Guidelines for Describing Associations and Alliances of the U.S. National Vegetation Classification*, The Ecological Society of America Vegetation Classification Panel.

Leckie, D.G., X. Yuan, D.P. Ostaff, H. Piene, and D.A. MacLean, 1992. Analysis of high resolution multispectral MEIS imagery for spruce budworm damage assessment on a single tree basis, *Remote Sensing of Environment*, 40(2):125–136.

Lobo, A., K. Moloney, and N. Chiariello, 1998. Fine-scale mapping of a grassland from digitized aerial photography: An approach using image segmentation and discriminant analysis, *International Journal of Remote Sensing*, 19(1):65–84.

Mayer, K.E., and W.F. Laundenslayer Jr., 1988. *A Guide to Wildlife Habitats of California*, Department of Fish and Game, Sacramento, California.

Parker, I., and W. Matayas, 1979. *CALVEG: A Classification of Californian Vegetation*, Region 5 Ecology Group, USDA Forest Service, San Francisco, California.

Rueda, C., J.A. Greenberg, and S.L. Ustin, 2004. STARSpan, University of California, Davis, URL: *http://casil.ucdavis.edu/projects/starspan* (last date accessed: 22 February 2006).

USDA, 2002. *Existing Vegetation Classification and Mapping Technical Guide (Draft)*, USDA, Forest Service, Washington, D.C.

Warbington, R., B. Schwind, R. Brohman, K. Brewer, and W. Clerke, 2002. *Requirements of Remote Sensing and Geographic Information Systems to Meet the New Forest Service Existing Vegetation Classification and Mapping Standards*, USDA Forest Service.

Welch, R., and W. Ahlers, 1987. Merging multiresolution SPOT HRV and Landsat TM data, *Photogrammetric Engineering & Remote Sensing*, 53:301–303.