



Improving image derived vegetation maps with regression based distribution modeling

S.Z. Dobrowski*, J.A. Greenberg, C.M. Ramirez, S.L. Ustin

Center for Spatial Technologies and Remote Sensing, Department of Land Air and Water Resources, University of California, Davis, USA

Received 5 January 2005; received in revised form 7 September 2005; accepted 7 September 2005
Available online 7 November 2005

Abstract

Incorporating ecological information into image-based vegetation mapping remains a challenge. Much attention has been placed on the use of ancillary information layers in image classification (e.g. slope, aspect, elevation) in that they provide indirect links to information that is ecologically relevant to species distributions. The objective of this study was to assess the utility of incorporating regression-based distribution model surfaces with image classification results using a consensus theoretic approach. We used spatially explicit non-parametric regression modeling in order to incorporate ancillary information in the production of an existing vegetation map for the Lake Tahoe basin. Probability surfaces for 19 prevalent species or genera were produced using generalized additive modeling (GAM). Models were fit to plot data obtained from multiple resource agencies using land-form based explanatory variables derived from a digital elevation model. Model evaluation was assessed by examining species response curves and through cross-validation, resulting in a range of accuracies for individual species (ROC values from 0.58 to 0.85). Probability surfaces for the study area were subsequently generated within a GIS. These surfaces were spatially re-sampled and used in conjunction with IKONOS imagery for use in vegetation mapping. The GAM surfaces were combined with maximum likelihood image classification results using consensus theory and a simple iterative weighting scheme. Results from the analysis demonstrate that the inclusion of the GAM surfaces improved individual class accuracies and suggests the need for implementing standardized and objective species modeling techniques for improving vegetation maps.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Generalized additive model (GAM); Species prediction; Distribution modeling; IKONOS; Ancillary data; Multi-source; Consensus theory; Vegetation mapping; Species mapping; Lake Tahoe

1. Introduction

Mapping vegetation at high thematic resolutions remains a significant challenge for the remote sensing

community. Advances in the area of data-fusion have improved image classification results (Le Hégat-Masclé et al., 2000; Solberg, 1999) and has focused on fusing image data from different sensors or different spatial scales (Pohl and Van Genderen, 1998; Prasad et al., 2001; Wald, 1999) while less attention has been

* Corresponding author.

paid to optimal methods for incorporating ecological data into remote sensing analysis. We begin this introduction by asking: How can ecological data best be used to improve remote sensing classification? A complementary and relevant problem to this is how to best incorporate image variables in gradient modeling, a common technique for predictive vegetation mapping.

The practice of including ecologically relevant terrain variables in image analysis has been with us for some time (Hutchinson, 1982). Similarly, image data has long been used in gradient modeling as predictor variables (Lees and Ritman, 1991) and is more recently referred to as a functional gradient given that it describes the response of the biota to direct, indirect, and resource gradients (Muller, 1998). Franklin (1995) reviews the applications, theory, data sources, and methodologies of predictive vegetation mapping highlighting the complementary nature of remotely sensed data and gradient analysis.

Advances in computing power, in conjunction with the availability of suitable statistical and GIS software, has led to a proliferation in analytical tools for examining and predicting species distributions. Franklin (1995) and Guisan and Zimmerman (2000) provide excellent reviews of the evolution of these tools, highlighting many of the prevalent statistical approaches that are available. Additionally, greater attention has been paid to ecological theory as it relates to distribution modeling. Austin (2002) provides an invaluable review of existing ecological theory as it relates to the expected shape of species response curves to environmental gradients.

Regression analysis, in its various forms, has become particularly prevalent in the literature supplanting more traditional ordination techniques. Generalized linear models (GLM, McCullagh and Nelder, 1989) and generalized additive models (GAM, Hastie and Tibshirani, 1990) using presence–absence survey data have garnered much attention. One of the primary strengths of these regression techniques is that they readily allow assessment of non-linear relationships between response and predictor variables. Non-parametric GAM models are particularly well suited to fitting complex non-linear response curves to environmental gradients. The GAM implementation utilizes a user defined smoothing function such as loess or spline smoothers to relate response to predictors (Chambers and Hastie, 1992). Consequently, the nature

of the response is data-defined and can be complex. Interpreting the shape of these responses provides a visual qualitative assessment of model suitability when compared with expert knowledge or known species autecology. This form of model assessment is not possible using other non-parametric modeling approaches such as classification and regression tree techniques (CART, Breiman et al., 1984) which use binary splits in explanatory variables and consequently do not provide response curves. Studies comparing CCA, GLM, GAM, and CART have demonstrated the rigor and flexibility of both GLM and GAM models using a wide array of response variables and across multiple scales (Franklin, 1998; Guisan et al., 1999; Moisen and Frescino, 2002; Wilfried et al., 2003).

Incorporating the results of regression based distribution models into image classification needs further exploration. Logistic forms of GLM and GAM models produce spatially explicit probability surfaces of species presence that are highly compatible with image classification techniques. These probability surfaces can be incorporated into parametric classifiers such as the maximum likelihood classifier as prior probabilities (Pedroni, 2003; Strahler, 1980) or into non-parametric classifiers such as CART (McIver and Friedl, 2002) in order to improve classification results. However, a common dilemma with using prior probabilities in supervised classification algorithms is that priors can incorrectly bias the posterior probability towards classes predicted using the ancillary information (Strahler, 1980). This is most problematic when there are large uncertainties associated with the estimate of the priors. For example, low prior probabilities can preclude the selection of a class even if it is spectrally unique from other classes (Strahler, 1980). Consequently, more flexible and exploitable techniques are needed for incorporating modeled ecological data into image classification.

A technique for combining multiple independent information sources that has received little attention in the ecological and remote sensing literature is consensus theory (Benediktsson and Swain, 1992). Consensus theory attempts to combine single probability distributions to summarize estimates from multiple experts (information sources) with the assumption that each expert makes decisions based on Bayesian decision theory. It posits that a group decision is better in terms of mean squared error, then a decision based on a single expert (information source) (Benediktsson et al., 1997).

Consensus theory can flexibly accommodate a wide array of information sources. Ancillary data is incorporated directly using source-specific density estimation that characterizes that source (Benediktsson and Swain, 1992). Additionally, information from independent classifiers or other modeling approaches can be combined to improve the final results (Briem et al., 2002).

In consensus theory, each information source is considered separately and then subsequently combined using a global membership function. The global membership function is used in classification by applying the maximum a posteriori solution (MAP) in which the combined estimate is calculated for all classes and a pattern vector is assigned the class with the highest posterior probability. The combination formula, known as a consensus rule, incorporates a source-specific weighting scheme for which the weights can be arbitrarily derived, heuristically based on the quality of the individual information sources, or optimized for the particular classification problem. Moreover, more sophisticated weighting schemes can be incorporated that not only weight the individual information sources, but can also weight the individual classes within an information source using linear and non-linear optimization routines (Benediktsson and Sveinsson, 2003; Benediktsson et al., 1997; Briem et al., 2002). The most commonly used consensus rule is the linear opinion pool (LOP), which is a weighted sum of the posterior probabilities from each data source. Another consensus rule, the logarithmic opinion pool (LOGP) is based on the weighted product of the posterior probabilities

$$\text{LOP} \quad C_j(Z) = \sum_{i=1}^n \lambda_i p(w_j|z_i) \quad (1)$$

$$\text{LOGP} \quad L_j(Z) = \prod_{i=1}^n p(w_j|z_i)^{\lambda_i} \quad (2)$$

where $p(w_j|z_i)$ is a source specific posterior probability, λ_i 's ($i = 1, \dots, n$) are source specific weights which control the relative influence of that data source, and, $C_j(Z)$ and $L_j(Z)$ are global membership functions. Consensus theoretic classifiers have the potential of providing more accurate classification results compared to traditional multivariate methodologies.

Given this, the objective of this study is to utilize regression based distribution models to inform image-based species mapping in a transparent systematic

manner. More explicitly, we intend to integrate GAM species distribution surfaces with IKONOS image classification results and characterize the improvements, if any, of utilizing a consensus theoretic approach.

2. Methods

2.1. Study area

The Lake Tahoe Basin spans California and Nevada between the Carson and Sierra Nevada mountain ranges. The total land area in the basin is approximately 82,000 hectares, roughly half of which is public lands including the Tahoe, Toiyabe, and Eldorado National Forests. The elevation of the basin ranges between 1900 m (a.s.l.) at lake level and 3400 m (a.s.l.) at the highest peaks. Annual precipitation is 500–1500 mm depending on elevation and location, two-thirds of which falls between December and March as snow (Rogers, 1974). The geology of the basin is primarily granitic along the eastern, southern, and western shores. The north shore of the basin has volcanic substrates. These bedrocks have produced primarily shallow Entisols, Inceptisols, and more developed Alfisols (Rogers, 1974).

The majority of the basin spans the montane (2187–2656 m) and sub-alpine (>2656 m) elevation ranges. Vegetation types are diverse and include conifer dominated forests, both evergreen and deciduous shrublands, as well as diverse meadow and fen habitats. The vegetation communities have had significant anthropogenic disturbances. Roughly two-thirds of the basin forests were clear-cut during the latter third of the 19th century (Elliot-Fisk et al., 1997). More recently, there has been significant development and urbanization within the basin.

2.2. GAM analysis

2.2.1. Vegetation modeling data

Plot data for species prediction modeling was primarily acquired from resource agencies including the Natural Resource Conservation Service and the USDA Forest Service. This plot data was acquired following standards put forth by the Terrestrial Ecological Unit Inventory Technical Guide (Winthers et al., 2003). Existing data was augmented by a proportional strati-

Table 1
Summary of environmental stratification (frequency in study area | frequency in field sample) used for augmenting existing data sources

Aspect	Slope (°)	Elevation	East/west
Groups			
Flat (0.04 0.07)	0–10 (0.23 0.32)	Lower montane, <2187 m (0.22 0.20)	East (0.25 0.20)
North (0.31 0.26)	11–30 (0.36 0.39)	Upper montane, 2187–2656 m (0.69 0.64)	West (0.75 0.80)
East (0.75 0.80)	31–60 (0.34 0.25)	Sub-alpine, >2656 m (0.09 0.16)	
South (0.31 0.35)	>60 (0.7 0.5)		
West (0.18 0.16)			

Unique combinations of the stratification variables were identified and the frequency of cells within these eco-regions for the entire study area as well as for the vegetation plots were calculated in order to facilitate locating vegetation plots. Given the large number of unique eco-regions, we present the frequencies for the original stratified variables.

fied sampling campaign that focused on missing vegetation classes and coverage of previously non-sampled geographic regions. In order to facilitate sampling efforts, the basin was split into east and west regions and then stratified by elevation, slope, and aspect. One hundred and twenty unique combinations of groupings were identified and are referred to as eco-regions. We calculated the frequency of cells within these eco-regions for the entire Tahoe basin as well as for the vegetation plots. Given the large number of potential eco-regions, we provide frequencies for both the study area and the plot sample for the four original stratified variables as opposed to all 120 eco-regions (Table 1). This form of stratification indicates that our plot samples were roughly representative of the study area with discrepancies arising in three principle areas: (1) we under-sampled steep slopes and over sampled flat areas; (2) we under sampled the eastern portion of the basin; (3) and lastly, we over sampled sub-alpine areas.

The minimum consistent suite of measurements that occurred across all the datasets from the various agencies included GPS coordinates, basic site characteristics (elevation, slope, aspect), ocular estimates of all woody species cover, and cover estimates of prevalent herbaceous species. This set of measurements represents the core suite of measures taken in 442 plots.

2.2.2. Explanatory variables

Five topographic and three climate variables were assessed as explanatory variables in this analysis. The topographic variables were produced using a 10 m USGS digital elevation model. These variables included both direct and indirect gradients including elevation, slope, potential annual direct radiation (RAD), topographic relative moisture index (TRMI), and distance to water. RAD was calculated as described

by McCune and Keon (2002) and uses a set of empirical equations based on slope, aspect, and latitude to estimate potential solar radiation. The TRMI was calculated as described by Parker (1982) and quantifies available soil moisture as a function of aspect, slope, topographic position, and slope configuration. The TRMI provides a scalar value (scaled between 0 and 60) that helps identify potentially xeric and mesic sites. Distance to water was based on Euclidean distance to existing stream courses. Fig. 1 shows the topographically derived predictor surfaces used in the study. The three climate variables included minimum annual temperature, maximum annual temperature, and average precipitation, and were acquired from PRISM (www.ocs.orst.edu/prism/) at 4 km resolution.

2.2.3. Data screening

Prior to fitting the GAM model for each species, the dataset was limited to plots that occurred along a data-defined range of each of the environmental gradients characterized by the explanatory variables. This data screening was performed in order to reduce the leverage of large numbers of species absence values on a given GAM model. It also reduces the complexity of the model and the likelihood of over predicting species distributions (Austin and Meyers, 1996; Austin et al., 1994). For example, Jeffrey Pine was identified in the plot data between 1898 and 2174 m. A statistical model was built on plot data within this range of elevations plus 10 observations that included Jeffrey Pine absences beyond the upper and lower limit of this elevation range. This procedure was conducted for all explanatory variables producing a hyper-volume in environmental space for each species.

Also prior to model fitting, the co-linearity of predictor variables for each species was assessed. We

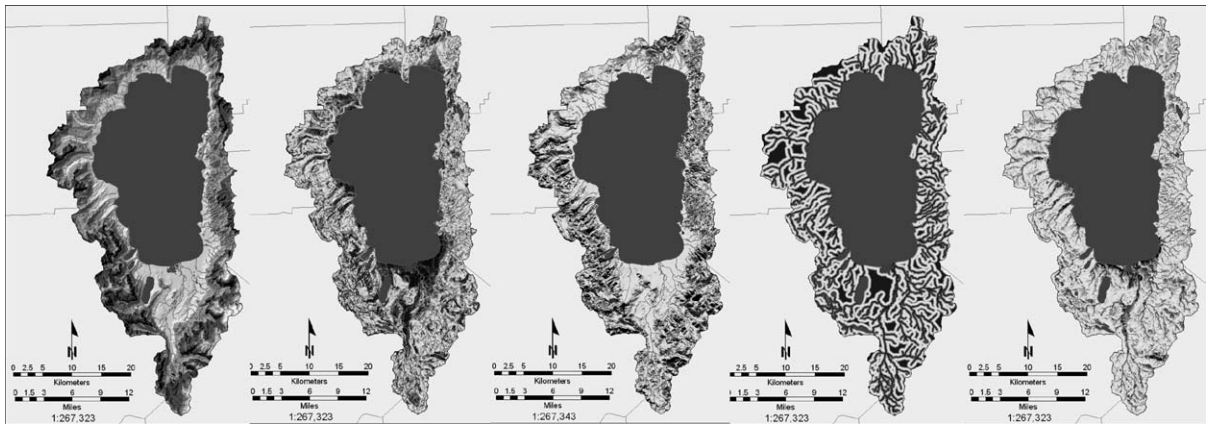


Fig. 1. Explanatory variables used in GAM. From left to right: elevation, slope, potential annual direct radiation (RAD), distance to water, topographic relative moisture index (TRMI).

calculated pair-wise correlation coefficients between all predictor variables in order to quantify co-linearity. We chose a threshold of $r = 0.70$ as a cutoff for discarding a given explanatory variable from the analysis.

2.2.4. Model fitting

GAM was implemented to predict the distribution of 19 prevalent species and genera using a group of S-plus (Insightful Inc.) tools called generalized regression analysis and spatial prediction (GRASP, Lehman et al., 2002). The vegetation classes included nine tree species, six shrub species one shrub genera, and three herbaceous classes. A logit link function and a quasi-binomial error term were used in the GAM modeling. A cubic spline smoothing function was used to relate response and predictor variables. Model selection was based on a step-wise procedure using the χ^2 change in deviance at the 5% level. A starting model was fitted with all predictor variables smoothed with four degrees of freedom. The significance of dropping the smoothed terms or converting them to linear terms was then tested. The updated, more parsimonious, model was kept and acted as the starting point for the next step-wise test. The step-wise procedure was conducted until no further significant reduction in deviance occurred due to adding or dropping a given term.

2.2.5. Model assessment

Qualitative assessment of the modeling results was conducted using partial response curves and plots of

response versus predictors. This allows for assessment of the model by comparing the results to known autecological species characteristics. A quantitative assessment of the modeling results was conducted using a five-fold cross-validation procedure. The cross-validation procedure partitions the plot data into five groups, four for training and one for assessing model quality, and then sequentially assigns each of the remaining four groups to the test group, while recording model quality at each iteration. The area under the curve (AUC, Fieldings and Bell, 1997) statistic was used to quantify model quality for the binomial model. The AUC is a graphical method that does not require a fixed threshold for converting probability estimates to nominal classes, but instead assesses a range of thresholds.

2.2.6. Spatial prediction

Regression models were used to produce a lookup table for each of the 19 species and genera using the GRASP tools in S-plus. These lookup tables were implemented in a GIS the predictor surfaces for inputs. The output was a spatially explicit probability surface of species presence for each of 19 vegetation classes. Elevation thresholds derived from both the plot data and published sources were then used to mask the probability surfaces to zero values outside of known species elevation ranges.

We assumed that all pixels belonged to one of the described vegetation classes. Consequently we normal-

ized the probability values of all 19 vegetation classes to sum to one using:

$$\pi_{i,x,y} = \frac{p_{i,x,y}}{\sum_{i=1}^m p_{i,x,y}} \quad (3)$$

where $p_{i,x,y}$ is the GAM probability value, i the class ID, m the maximum number of classes, x the x coordinate of the cell, and y the y coordinate of the cell. We then re-sampled this 10 m spatial resolution image to a 1 m cell size and spatially co-registered it to the output of the IKONOS image classification (described below).

2.3. IKONOS image classification

The following section provides a general overview of the methods used in the image-based classification. We employed an H-resolution discrete scene model (Strahler et al., 1986) in which the scene elements are tree crowns and vegetation patches. For further detail see Dobrowski (2005) and Greenberg et al. (2005).

2.3.1. Maximum likelihood classification

Data for image classification and map accuracy were acquired for the 19 vegetation classes described above, and three non-vegetated classes including water, dark impermeable surfaces (volcanic substrates, asphalt), and bright impermeable surfaces (granite, concrete). About 1700 polygon features were delineated representing individual tree crowns, patches of homogenous vegetation, and non-vegetated regions. Polygons were taken across a wide range of elevations and geographic regions within the basin to capture intra-class variability. Field data were collected using in-field digitization techniques (differential GPS linked in real-time to geo-referenced color-infrared IKONOS imagery). Maximum likelihood classification (MLC) was conducted using pattern vectors of 12 principal component transforms (PCA) developed from four IKONOS pan-sharpened spectral bands and eight grey level co-occurrence matrix (GLCM) texture variables. For each pixel located within one of the 1700 features, the following data was recorded in a database: (1) the feature ID of the polygon that the pixel intersected; (2) the map class associated with the polygon; (3) and all 12 PCA image band values. A multivariate outlier analysis was performed on the dataset using a jackknife technique

with a Mahalanobis distance statistic. Observations falling outside of the 99% quantile were dropped from the analysis. This dataset was then split into a training data set and a validation data set by stratifying the polygons by map class and then randomly assigning 60% of the polygon features (and their respective pixels) to a training set and the remaining 40% to a test set.

After applying the MLC classification to the entire image, we spectrally subset the 22 band likelihood image 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 using Eq. (3). In a subsequent classification step, pixels were identified as non-vegetated classes and were given precedence over vegetated labels.

2.4. Multi-source analysis

GAM estimates of species probability distributions and IKONOS maximum likelihood estimates were combined using consensus theory. We tested both the LOP and LOGP consensus rules described in Eqs. (1) and (2). Two data sources were used in this analysis: the normalized MLC likelihood image, and the normalized GAM probability image. We combined them using a simple iterative weighting scheme in which we varied λ_i 's in 0.01 increments between 0 and 1 for each information source. At each weighting combination we recorded the accuracy of the resultant classification. Weights were chosen that maximized the overall kappa coefficient of the global membership function.

2.4.1. Accuracy assessment

In order to assess the accuracy of the image MLC, and the consensus theoretic (multi-source) output, we used the maximum a posteriori (MAP) solution to convert probability estimates to nominal vegetation classes. In other words, we allocated a given pixel to the class that had the largest posterior probability. As stated earlier, 40% of the ground data was set aside for assessing accuracy. This data was used to calculate class producers and users accuracy as well as the overall kappa coefficient for each information source. We also provide overall accuracy (%) for comparative purposes but do not endorse the use of this metric due to its known bias (Congalton, 1991, 2001).

3. Results

3.1. GAM model assessment

When constructing the species models, we found that in all instances, the three climatic variables (max temperature, min temperature, and precipitation) were

highly correlated with elevation ($r > 0.70$). Consequently, we dropped these predictor variables from the analysis due to their coarse resolution and to avoid model instability due to co-linearity.

We used predictor versus response curves and partial response curves to assess the suitability of the GAM models. Figs. 2 and 3 provide an example of

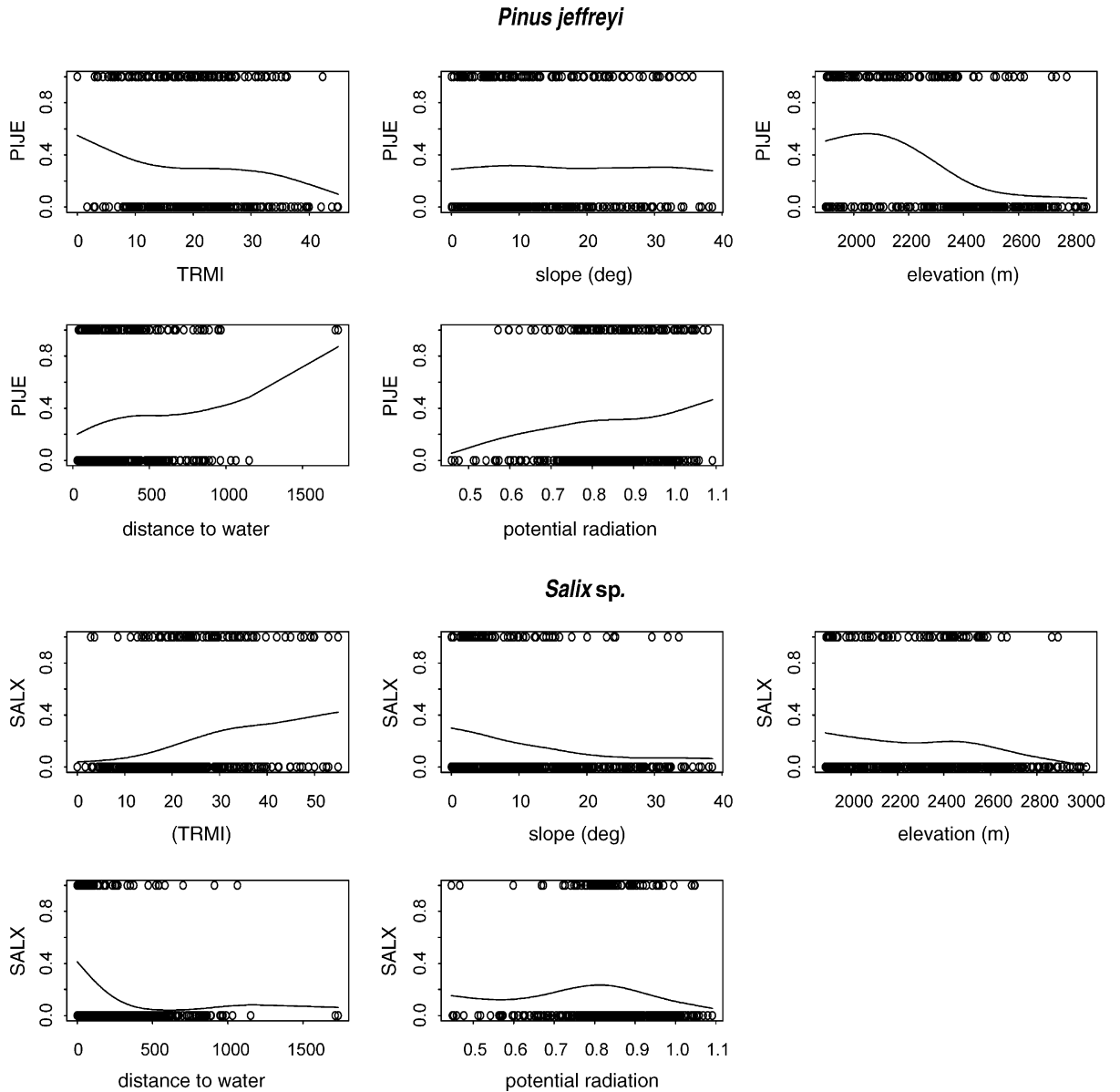


Fig. 2. Predictors vs. response curves for select species in the study. The y-axis represents the probability of species presence. The x-axis represents individual environmental predictor variables.

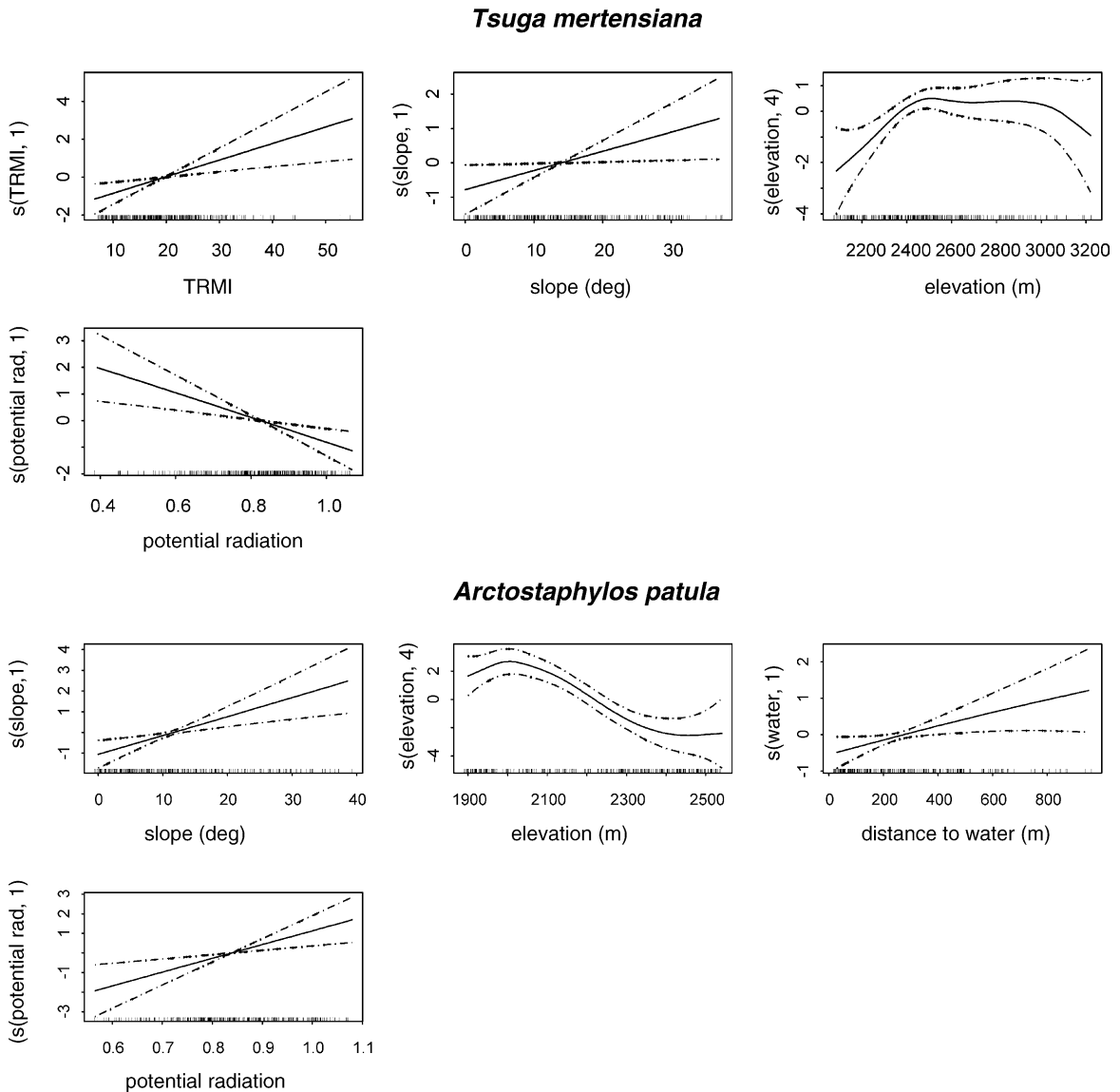


Fig. 3. Example of partial response curves for select species. The variables shown were found to be significant predictors in the relevant GAM model. The smoothed or linear functions represent the contribution of the individual explanatory variable to the fitted additive predictor in the GAM model. The dashed lines represent $2\times$ standard error bands. The rug plots show the density of samples across the range of the predictors.

the response and partial response curves for two conifer species (*Pinus jeffreyi* and *Tsuga mertensiana*) and two shrub types (*Salix* sp. and *Arctostaphylos patula*). The model results show that *T. mertensiana* is found on steep slopes at higher elevations in more mesic sites associated with north facing slopes whereas *P. jeffreyi* is found at lower elevations in more xeric sites. These

results are consistent with the known autecology of both *T. mertensiana*, and *P. jeffreyi*. Results from both *A. patula* and *Salix* sp. show consistency with known species characteristics as well. *A. patula* is found preferentially on steeper slopes at xeric sites across a wide range of lower elevations in the basin. These results are consistent with the natural history of this chaparral

species. In contrast, the strong response of *Salix* sp. to distance from water, slope, and TRMI is consistent with the requirements of riparian genera. In addition to the response patterns shown, Fig. 3 also includes both rug plots and standard error bands. This data provides an assessment of error at different regions of the predictor variables. For example, standard error estimates are large at high slope values due to insufficient sampling density in these regions of the predictor.

A more quantitative assessment of model quality was provided through cross-validation (Table 2). The AUC statistic is scaled between 0.5 (random relationship between predicted and observed classes) and 1.0 (perfect association). The AUC values for the species modeled in this study range between 0.58 and 0.84. A rough guide to interpreting these values is to follow the traditional academic grading system: 1.0–0.9 = excellent; 0.9–0.8 = good; 0.8–0.7 = fair; 0.7–0.6 = poor; 0.6–0.5 = fail. As expected, species modeled with smaller number of observations tended to have poorer cross-validation results (e.g. *J. occidentalis*, *Populus tremuloides*) suggesting that some classes were insufficiently sampled. We also noted that species with narrow environmental ranges (e.g. *Pinus albicaulis*) were better modeled than species

Table 2
Summary of cross-validation results

Class	Sample size (n)	AUC
Trees		
<i>Juniperus occidentalis</i>	332	0.58
<i>Abies concolor</i>	334	0.79
<i>Abies magnifica</i>	379	0.72
<i>P. albicaulis</i>	169	0.83
<i>P. contorta</i>	419	0.67
<i>P. jeffreyi</i>	334	0.82
<i>Pinus monticola</i>	322	0.74
<i>P. tremuloides</i>	236	0.63
<i>T. mertensiana</i>	295	0.78
Shrubs		
<i>Alnus incana</i>	202	0.63
<i>Arctostaphylos patula</i>	247	0.84
<i>Artemisia tridentata</i>	373	0.77
<i>Ceanothus cordulatus</i>	242	0.75
<i>Ceanothus velutinus</i>	140	0.66
<i>Quercus vaccinifolia</i>	343	0.84
<i>Salix</i> sp.	433	0.74
Herbs		
<i>Carex</i> sp.	438	0.73
<i>Juncus</i> sp.	417	0.71
Graminoids	442	0.74

The AUC statistic is the average of five “leave one out” iterations. The sample size represents the number of species presences within our dataset.

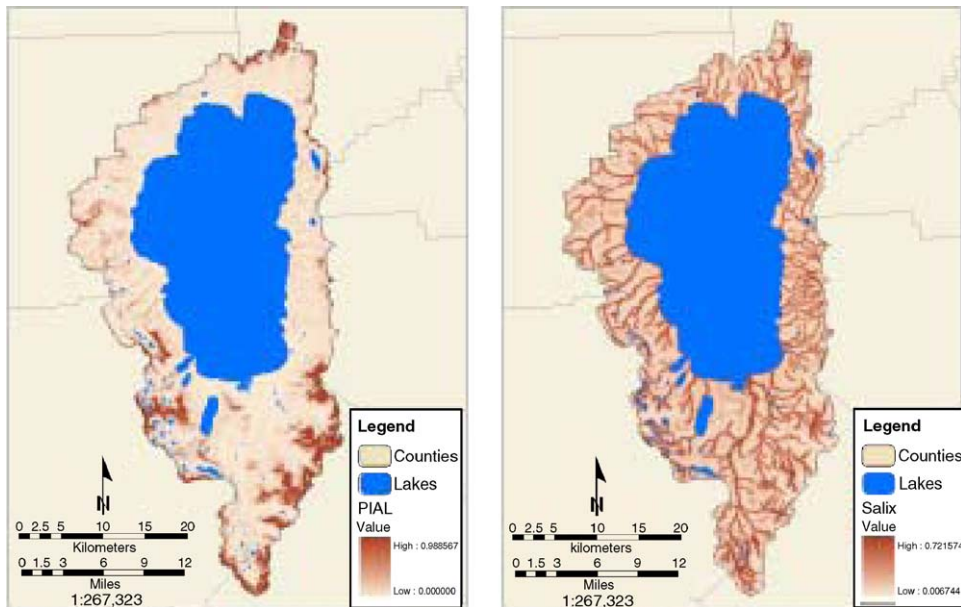


Fig. 4. An example of two GAM probability surfaces produced in the analysis for *P. albicaulis* (left) and *Salix* sp. (right).

showing broad environmental and geographic ranges (e.g. *Pinus contorta*).

We provide an example of the predicted model surfaces for *Salix* sp. and *P. albicaulis* (Fig. 4). These surfaces were also examined for consistency with known species distributions within the study area.

3.2. Multi-source classification and accuracy assessment

Without the inclusion of the GAM information ($MLC\lambda = 1$, $GAM\lambda = 0$) the overall kappa coefficient was 0.44 for the LOGP and LOP consensus rules. After performing the optimization, the maximum kappa coefficient was found to be 0.53 for LOGP and 0.48 for the LOP consensus rule. Consequently, inclusion of the GAM information resulted in a 0.09 and 0.04 improvements in the kappa coefficients for the LOGP and

LOP consensus rule respectively. The optimal weights (λ_i 's) for the consensus rules used in the analysis were 0.2 and 1.0 for the MLC and GAM information sources respectively. Without any form of optimization ($MLC\lambda = 1$, $GAM\lambda = 1$), the overall kappa coefficients were found to be 0.50 and 0.45 for the LOGP and LOP consensus rules respectively. The LOGP consensus rule appears to be a superior consensus rule under these circumstances and showed a higher overall kappa coefficient across all weights that included both information sources.

The inclusion of the GAM probabilities in the multi-source classification resulted in a significant improvement in a number of class level accuracies. Tables 3 and 4 summarize the image based and multi-source producer and user accuracies for both the LOP and LOGP consensus rules. The difference between the image based class accuracies and the multi-source class

Table 3
Comparison of accuracies (LOGP; $MLC\lambda = 0.2$, $GAM\lambda = 1.0$)

	Image based accuracy (%)		Multi-source accuracy (%)		Difference (%)	
	Producer	User	Producer	User	Producer	User
Trees						
<i>A. concolor</i>	51.6	47.3	73.3	63.9	21.7	16.6
<i>A. magnifica</i>	36.3	28.8	52.6	25.4	16.3	-3.4
<i>J. occidentalis</i>	40.3	33.3	29.8	54.9	-10.5	21.6
<i>P. albicaulis</i>	16.0	11.5	39.6	62.7	23.6	51.2
<i>P. contorta</i>	10.6	21.2	32.8	25.6	22.2	4.4
<i>P. jeffreyi</i>	34.4	31.4	50.7	41.4	16.3	10
<i>P. monticola</i>	21.4	16.5	20.2	18.9	-1.2	2.4
<i>P. tremuloides</i>	73.9	57.4	52.7	70.8	-21.2	13.4
<i>T. mertensiana</i>	19.5	12.9	22.5	35.3	3	22.4
Mean	33.8	28.9	41.6	44.3	7.8	15.4
Shrubs						
<i>A. incana</i>	59.7	88.2	54.0	93.3	-5.7	5.1
<i>A. patula</i>	60.3	22.5	71.4	48.9	11.1	26.4
<i>A. tridentata</i>	59.1	24.6	58.3	63.4	-0.8	38.8
<i>C. cordulatus</i>	42.8	10.1	48.3	10.2	5.5	0.1
<i>C. velutinus</i>	47.9	46.8	47.7	83.4	-0.2	36.6
<i>Q. vaccinifolia</i>	42.2	4.2	67.2	23.2	25	19
<i>Salix</i> sp.	44.3	38.1	63.1	36.6	18.8	-1.5
Mean	50.9	33.5	58.6	51.3	7.7	17.8
Herbs						
<i>Carex</i> sp.	53.0	99.8	62.2	99.9	9.2	0.1
<i>Juncus</i> sp.	N/A	N/A	N/A	N/A	N/A	N/A
Graminoids	78.0	60.0	83.4	47.2	5.4	-12.8
Mean	65.5	79.9	72.8	73.5	7.3	-6.3

Overall kappa coefficient = 0.53, overall accuracy = 60.2%.

Table 4
Comparison of accuracies (LOP; MLC λ = 0.20, GAM λ = 1.0)

	Image based accuracy (%)		Multi-source accuracy (%)		Difference (%)	
	Producer	User	Producer	User	Producer	User
Trees						
<i>A. concolor</i>	51.6	47.3	65.3	58.1	13.7	10.72
<i>A. magnifica</i>	36.3	28.8	42.0	29.8	5.7	1.0
<i>J. occidentalis</i>	40.3	33.3	36.4	40.7	-3.9	7.38
<i>P. albicaulis</i>	16.0	11.5	36.0	27.8	20	16.35
<i>P. contorta</i>	10.6	21.2	20.4	27.8	9.8	6.58
<i>P. jeffreyi</i>	34.4	31.4	41.7	33.6	7.3	2.19
<i>P. monticola</i>	21.4	16.5	16.1	18.1	-5.3	1.62
<i>P. tremuloides</i>	73.9	57.4	70.3	65.3	-3.6	7.91
<i>T. mertensiana</i>	19.5	12.9	18.8	20.0	-0.7	7.07
Mean	33.8	28.9	38.6	35.7	4.8	6.8
Shrubs						
<i>A. incana</i>	59.7	88.2	59.6	89.4	-0.1	1.2
<i>A. patula</i>	60.3	22.5	63.3	28.4	3	5.89
<i>A. tridentata</i>	59.1	24.6	56.3	25.2	-2.8	0.59
<i>C. cordulatus</i>	42.8	10.1	46.2	12.4	3.4	2.29
<i>C. velutinus</i>	48.0	46.8	45.4	59.7	-2.6	12.88
<i>Q. vaccinifolia</i>	42.2	4.2	47.3	6.0	5.1	1.71
<i>Salix</i> sp.	44.3	38.1	46.4	37.2	2.1	-0.87
Mean	50.9	33.5	52.1	36.9	1.2	3.4
Herbs						
<i>Carex</i> sp.	53.0	99.8	56.9	99.9	3.9	0.1
<i>Juncus</i> sp.	N/A	N/A	N/A	N/A	N/A	N/A
Graminoids	78.0	60.0	79.1	50.9	1.1	-9.1
Mean	65.5	79.9	68	75.4	2.5	-4.5

Overall kappa coefficient = 0.48, overall accuracy = 55.7%.

accuracies are also shown in these tables and represent the influence of incorporating the GAM information on the final map accuracy.

Image based accuracies for tree and shrub species were poor. Average user accuracies for tree and shrub species were 28.9% and 33.5% respectively, with individual class accuracies varying widely. Herbaceous classes showed significantly higher image based accuracies. These herbaceous classes represent common genera, or in the case of the graminoid class, a consistent physiognomy. Consequently, the higher accuracies are likely due to their thematic coarseness.

The inclusion of ecologically relevant ancillary data encapsulated in the GAM surfaces resulted in a marked improvement in class accuracies. For the LOGP consensus rule, inclusion of the GAM information resulted in an average improvement in producer accuracies of roughly 7% for tree, shrub, and herbaceous class types.

User accuracies showed an even larger improvement of 15% and 18% for the tree and shrub classes respectively whereas herbaceous classes suffered a reduction of 6%. The LOP consensus rule showed less marked changes in class accuracies with the inclusion of the GAM data. For the LOP consensus rule, changes in producer and user accuracy varied between -4.5% and 6.8%.

Accurate GAM models resulted in larger improvements in the multi-source class accuracies as compared to poor GAM models. Fig. 5 shows a positive linear relationship ($r^2 = 0.46$) between the GAM AUC values and change in producer accuracy due to inclusion of the GAM surface in the consensus theoretic classification. GAM models with low AUC values tended to show marginal improvements in producer accuracy, or in some cases, actually reduced the accuracy of the final multi-source classification.

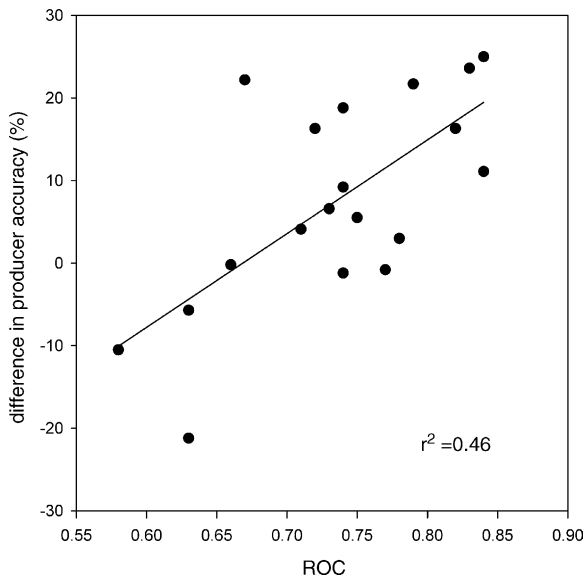


Fig. 5. GAM model accuracy (as characterized by a cross-validated ROC statistic) vs. change in the final classification accuracy by inclusion of the GAM data in the consensus theoretic classifier for all 19 vegetation classes.

4. Discussion

4.1. Species mapping

This study is a component of a larger project, the purpose of which was to produce a vector-based map of vegetation communities and vegetation structure for the Lake Tahoe basin using IKONOS imagery. We identified an ecological model (*sensu* Austin, 2002) that provided the starting point for the mapping project and was based on one fundamental principle: vegetation communities are comprised of *species* assembled along physical gradients in the environment. This is the familiar continuum concept of vegetation community ecology described by Whittaker (1951) among others, although we made no assumptions about the shape of the species – environmental gradient response curve.

By accepting this ecological model as the *modus operandi*, there were a number of positive implications for the development of the remote sensing methodologies subsequently employed: (1) the most important of these is that we would map species as opposed to communities, which would be subsequently identified

through data-mining and aggregation techniques (Greenberg et al., 2005); (2) when conducting a field campaign, sampling species was a more tractable task than sampling “communities” which can be difficult to define and have unclear boundaries; (3) lastly, we recognize that the spectral signature of a species has less variability than that of a vegetation community which can have ill-defined boundaries, can be comprised of a number of species with varying proportions, and thus have a highly variable spectral signature.

Despite these benefits, mapping species is challenging from a remote sensing perspective given that it requires both high spectral and spatial resolution to capture the target signal. Most satellite sensors provide high spatial resolution data or high spectral resolution data, not both. The IKONOS sensor, with 1 m panchromatic and 4 m multi-spectral data, has ushered in a new era of H-resolution (Strahler et al., 1986) remote sensing techniques that allow for the discernment of individual organisms within scenes. With these capabilities, come unique challenges as well.

Spatial resolution and spectral resolution play critical roles in our ability to distinguish unique floristic types. Our findings from the larger mapping project suggest that IKONOS imagery allows for accurate classifications at the lifeform level, which has been supported by other studies (Clark et al., 2004; Sawaya et al., 2003; Thenkabail et al., 2004). This is primarily a consequence of the unique image textures associated with different lifeforms that are readily identified in H-resolution imagery. Texture information has been widely exploited to estimate structural characteristics of vegetation and leaf area index using primarily H-resolution imagery (Wulder, 1998) and to a lesser extent L-resolution imagery (Franklin and Peddle, 1990). Its ability to inform species level mapping is contingent upon those species forming stands of unique density, spacing, or structural characteristics. We found little evidence of this for species within a given lifeform (e.g. conifers). Additionally, our poor image-based classification results at the species level attest to a lack of spectral resolution for the classification problem. In other words, the IKONOS sensor’s four broad spectral bands are not adequate for distinguishing between species with similar lifeforms.

Despite this, there are examples in the literature of direct species mapping using digital imagery. It

is mostly accomplished through the use of sensors that provide both high spectral and spatial resolution, such as AVIRIS or CASI. These sensors provide the foundation for most of the successful applications of direct species mapping in forests (Franklin, 1994; Leckie et al., 2005; Martin et al., 1998; Zarco-Tejada and Miller, 1999) and for invasive species (Underwood et al., 2003). Unfortunately, CASI and AVIRIS are aerial sensors with limited spatial extents and their widespread use is unlikely given the limited coverage of these sensors, the costs of acquiring data over large areas, and the challenge of analyzing it. Additionally, there are currently no space-borne sensors that combine both hyper-spatial and hyper-spectral resolution characteristics. The Landsat sensors, the mainstay of resource sciences over the past 30 years, do not have finely resolved spatial or spectral characteristics by today's standards and the most recent of these sensors, Landsat ETM+ has reduced operational capacity. Consequently, the role of ancillary information sources in vegetation mapping must be re-assessed as a means to improve map accuracies. Our approach addresses this explicitly and diverges from those hyper-spectral studies described above by our reliance on species distribution modeling to inform the species mapping process in the absence of high spectral resolution data.

4.2. Regression based distribution modeling

GAM regression techniques provide a systematic approach for distilling the large number of ancillary information sources that can be applied to the classification problem, into a single probability surface that can be incorporated into mapping efforts. Out of all the explanatory variables used in this study, not one of them fit a Gaussian distribution, with most distributions showing strong right or left skewness. GAM is non-parametric and data-specific. Consequently, it is readily applied to the non-Gaussian data sources used in this analysis without data transformations. Additionally, plots of the GAM results allowed for qualitative assessment of model adequacy. The response curves produced from the GAM analysis for the majority of the 19 vegetation classes showed response patterns that are consistent with known autecological characteristics of those species. Qualitative accuracy statistics also allowed for more refined and intuitive model assess-

ment. Those species that were poorly modeled shared a primary characteristic: they had the weakest representation in the plot sample with a number of these classes also showing geographically localized samples. Consequently, this suggests the need for improving the design and efficiency of the sampling strategy, such as utilizing an 'equal stratified sampling regime' (Hirzel and Guisan, 2002).

Regression techniques such as GAM also have distinct limitations that are relevant to vegetation mapping. GAM or GLM surfaces represent the realized niche of a particular species (Guisan and Zimmerman, 2000). The realized niche includes both abiotic and biotic influences on species distributions. Consequently, the relative role of biotic factors such as competition, as compared to abiotic factors, is difficult to ascertain. Examples of including biotic variables in GAM are rare but have been conducted (Leathwick and Austin, 2001) and need further exploration. More importantly, analysts must recognize that all of the empirical species distribution models discussed in this study are static and assume a pseudo-equilibrium in vegetation dynamics. However, the evidence in support of whether populations are near a stable equilibrium is sparse and difficult to obtain and this assumption may be unfounded (Loreau et al., 2001; McCann, 2000). Moreover, much of our public lands show evidence of anthropogenic and natural disturbance (e.g. logging and fire) that strains the equilibrium assumption. This limitation is particularly restrictive for species or communities that are short lived or quick to react to environmental change (e.g. riparian and grasslands). Little work has been conducted that includes disturbance information as predictors in regression-based distribution modeling. Some unique examples come from Leathwick (1998) and Leathwick and Mitchell (1992).

4.3. Effects of varying support sizes

Our approach comes with distinct challenges and limitations as well. There exists a mismatch in the size of support (*sensu* Dungan, 2001) between plot data collected to drive species distribution modeling (100's m²), ground data used to validate the map (individual tree crowns and vegetation patches at a few m² to 100's m²) and image data (1 m²). Support, a geostatistical term, refers to the shape, size, and orientation of an *n*-dimensional volume over which measurements are

made. Our image derived data, and ground validation data has a smaller support size than our modeled species distribution data. We re-sampled the GAM probability surfaces to the 1 m^2 support size of the IKONOS imagery recognizing that this is false precision. We retained this resolution in order to maintain the physical detail of the map for subsequent data querying and aggregation.

The effect of multiple support sizes is not entirely clear. The “support effect” (Olea, 1990) states that, with univariate distributions, the mean value of a regionalized variable will stay the same, the variance will decrease, and the distribution will become more symmetric with increasing support size. The effects of support size changes on bivariate or multivariate statistics remains difficult to ascertain, and when assessed on largely univariate problems, require the use of geostatistical techniques (Dungan, 2001). Of course support is not just represented by the area of the collected data, but also is a function of the geometrical shape and orientation of the data. In cases where irregularly shaped vector data is acquired (such as this study), the support problem is further confounded. Geostatistical models rely on a consistent support for each regionalized variable, and thus, are not amendable to variably shaped vector portrayals of spatial data used in GISs (Dungan, 2001).

The support effect likely has subtle and not-so-subtle consequences on the interpretation of the results of this study. For example, the accuracy of the GAM data alone ($\text{MLC}\lambda = 0$, $\text{GAM}\lambda = 1$) was 0.06 (kappa) for both the LOGP and LOP consensus rules. This value represents the probability that the GAM surfaces accurately predict the presence of one of 19 vegetation classes in a 1 m cell within the Tahoe basin. This accuracy value is extremely low and taken at face value suggests a complete failure of the GAM exercise. However, this value is misleading. It is partially a consequence of an incongruity in the size of support associated with GAM surfaces (i.e. 100 m^2) and ground truth data acquired from irregularly shaped groups of pixels, 1 m^2 to 10 s m^2 in size. It is also exemplary of the disjoint between the overlapping distributions of multiple species at a given location and the requirement for a nominal existing vegetation class needed for an accuracy statistic such as the kappa coefficient. The influence of varying support size in H-resolution mapping studies deserves further attention.

4.4. Combining multiple information sources

Despite these apparent complications, we found that the GAM data acts in a complementary fashion to the image data used in the analysis. We see from the class level results, that the inclusion of the GAM data in the consensus theoretic approach produced noteworthy improvements in mean class accuracies with greater than 50% improvements in user accuracies for specific classes. This was achieved using a simple iterative weighting scheme with only two information sources. These findings also demonstrate the importance of accurate modeling results. Our findings suggest that there is indeed a positive relationship between GAM model accuracy, as assessed through the AUC statistic, and an improvement in the final vegetation map accuracies. Vegetation classes that were poorly modeled resulted in a decrease in the final map accuracies. Consequently, significant attention needs to be paid to the efficacy of any species modeling approach.

The improvement in the mapping accuracy achieved by incorporating GAM results arises from the fact that the GAM distributions effectively limit the number of classes that can be considered in the consensus theoretic classification on a cell-by-cell basis. By reducing the number of classes that are addressed in the final classification, the likelihood of choosing the correct class increases, even if the choice is random. This is most readily apparent in the LOGP consensus rule. In the LOGP, zero probabilities act as vetoes, thus when the posterior probability equals zero ($p(w_j|z_i) = 0$) from any information source, the global membership function equals zero as well ($L_j(Z) = 0$). This may likely be the reason the LOGP showed superior results as compared to the LOP which is constructed on an additive basis.

This veto characteristic of consensus theory also highlights a potential strength of the method as well. Namely, consensus theory provides a flexible framework for not only weighting information sources directly, but also weighting the individual classes within an information source. If an individual class is modeled poorly, it is undue influence on the final map accuracy can be mitigated by the use of class-specific weights. Although, this aspect was not studied here, it has been shown to improve classification results using Landsat imagery and other ancillary data (Benediktsson and Sveinsson, 2003; Benediktsson et

al., 1997). Due to the large number of possible weighting combinations, class-specific weighting requires linear or non-linear optimization routines such as neural networks or genetic algorithms (Benediktsson and Sveinsson, 2003; Benediktsson et al., 1997; Briem et al., 2002). These optimization techniques need further consideration.

Consensus theory belongs to a larger group of statistical techniques aimed at addressing the “multi-source” classification problem; that of optimally combining independent sources of information in a manner that takes into account the reliability of each data source. Similar areas of research have been explored through the use of evidential reasoning and other applications of Dempster–Shafer’s theory (Kim and Swain, 1995; Peddle, 1995). Our intention here is not to compare and contrast consensus theory to these approaches but instead, to highlight the use of these techniques for combining information sources in an opportunistic manner. That said, we chose a consensus theoretic approach because of its simplicity and ease of use as compared to evidential reasoning approaches that utilize interval-valued probabilities and require measures of support, plausibility, and decision rules based on these metrics (Peddle, 1995).

Consensus theory provides a general framework for combining information sources using existing classifiers and modeling approaches. There is no theoretical upper limit in terms of the number of information sources that can be incorporated. For example, by including other image classifiers in the analysis we can take advantage of relative strengths and weaknesses of their approaches. Moreover, by including other distribution modeling information sources (e.g. CART) in the analysis, this may result in further improvements in class accuracies, given that other distribution modeling techniques have specific strengths (e.g. CART handles interaction terms between the predictors implicitly).

5. Conclusions

Consensus theoretic approaches have the potential to provide more accurate classification results compared to traditional multivariate methodologies. This is primarily due to the fact that convenient multivariate data models are not available for the disparate information sources used in mapping studies. Consensus

theory overcomes one of the primary limitations of traditional parametric classifiers such as maximum likelihood classification by reducing the number of features used in the classification and consequently reducing the training set requirements needed for proper estimation.

Consensus theoretic approaches also allow users to utilize statistical techniques that are better suited for the individual information sources needed. We posit that regression-based distribution modeling is a robust and flexible statistical technique for screening and consolidating the multitude of topographic variables that can be applied towards image-based vegetation mapping projects. Moreover, GLM and GAM techniques are supported by a large body of research which outlines their strengths and weaknesses, describes optimal sampling strategies, and most importantly addresses their link to a foundation of existing vegetation ecological theory. Our findings demonstrate the utility of implementing these standardized and objective species modeling techniques for improving image-derived vegetation maps.

Acknowledgements

We would like to thank the Tahoe Regional Planning Agency and the USDA Forest Service for support of this project. We would particularly like to thank Hugh Safford, Mike Volmer, Shane Romsos, and Marchelle Munnecke for their help and feedback. We also thank our field crew, Jennifer Buck and Upekala Wijayratne. Lastly, we acknowledge the comments provided by Janet Franklin and an anonymous reviewer that greatly improved this manuscript.

References

- Austin, M., Meyers, J., 1996. Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity. *For. Ecol. Manage.* 85 (1–3), 95–106.
- Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. *Ecol. Model.* 157, 101–118.
- Austin, M.P., Nicholls, A.O., Doherty, M.D., Meyers, J.A., 1994. Determining species response functions to an environmental gradient by means of a beta function. *J. Veg. Sci.* 5 (2), 215–228.
- Benediktsson, J.A., Sveinsson, J.R., 2003. Multisource remote sensing data classification based on consensus and pruning. *IEEE Trans. Geosci. Remote Sens.* 41 (4), 932–936.

- Benediktsson, J.A., Sveinsson, J.R., Swain, P.H., 1997. Hybrid consensus theoretic classification. *IEEE Trans. Geosci. Remote Sens.* 35 (4), 833–843.
- Benediktsson, J.A., Swain, P.H., 1992. Consensus theoretic classification methods. *IEEE Trans. Syst., Man Cybernet.* 22 (4), 688–704.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. *Classification and Regression Trees*. Chapman & Hall, New York.
- Briem, G.J., Benediktsson, J.A., Sveinsson, J.R., 2002. Multiple classifiers applied to multisource remote sensing data. *IEEE Trans. Geosci. Remote Sens.* 40 (10), 2291–2299.
- Chambers, J.M., Hastie, T.J., 1992. *Statistical Models in S*. Chapman & Hall, London.
- Clark, D.B., Read, J.M., Clark, M.L., Cruz, A.M., Dotti, M.F., Clark, D.A., 2004. Application of 1 and 4 m resolution satellite data to ecological studies of tropical rain forests. *Ecol. Appl.* 14 (1), 61–74.
- Congalton, R., 2001. Accuracy assessment and validation of remotely sensed and other spatial information. *Int. J. Wildland Fire* 10 (3–4), 321–328.
- Congalton, R.G., 1991. A review of assessing the accuracy of classification of remotely sensed data. *Remote Sens. Environ.* 37 (1), 35–46.
- Dobrowski, S.Z., 2005. *On the Integration of Ecology in Remote Sensing Science*. University of California, Davis, 440 pp.
- Dungan, J.L., 2001. Scaling up and scaling down: the relevance of the support effect on remote sensing of vegetation. In: Tate, N.J., Atkinson, P.M. (Eds.), *Modeling Scale in Geographic Information Science*. John Wiley and Sons, pp. 221–235.
- Elliot-Fisk, D.L., Rowntree, R.A., Cahill, T.A., Goldman, C.R., Gruell, G., Harris, R., Leisz, D., Lindström, S., Kattelmann, R., Machida, D., Lacey, R., Rucks, P., Sharkey, D.A., Ziegler, D.S., 1997. *Lake Tahoe Case Study*. Science Team, Sierra Nevada Ecosystems Project, Final Report to Congress. Centers for Water and Wildland Resources, University of California, Davis.
- Fieldings, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24 (1), 38–49.
- Franklin, J., 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Prog. Phys. Geogr.* 19 (4), 474–499.
- Franklin, J., 1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *J. Veg. Sci.* 9 (5), 733–748.
- Franklin, S.E., 1994. Discrimination of subalpine forest species and canopy density using digital CASI, SPOT, PLA, and Landsat TM data. *Photogram. Eng. Remote Sens.* 60 (10), 1233–1241.
- Franklin, S.E., Peddle, D.R., 1990. Classification of SPOT HRV imagery and texture features. *Int. J. Remote Sens.* 11, 551–556.
- Greenberg, J.A., Dobrowski, S.Z., Ramirez, C.M., Tuil, J., Ustin, S.L., 2005. A bottom up approach to vegetation mapping of the Lake Tahoe Basin using hyperspatial image analysis. *Photogram. Eng. Remote Sens.*, submitted for publication.
- Guisan, A., Weiss, S., Weiss, A., 1999. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol.* 143 (1), 107–122.
- Guisan, A., Zimmerman, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135, 147–186.
- Hastie, T.J., Tibshirani, R., 1990. *Generalized Additive Models*. Chapman & Hall, London.
- Hirzel, A., Guisan, A., 2002. Which is the optimal sampling strategy for habitat suitability modeling. *Ecol. Model.* 157, 331–341.
- Hutchinson, C.F., 1982. Techniques for combining Landsat and ancillary data for digital classification improvement. *Photogram. Eng. Remote Sens.* 48, 123–130.
- Kim, H., Swain, P.H., 1995. Evidential reasoning approach to multisource-data classification in remote sensing. *IEEE Trans. Syst., Man Cybernet.* 25 (8), 1257–1265.
- Leathwick, J., 1998. Are New Zealand's Nothofagus species in equilibrium with their environment? *J. Veg. Sci.* 9 (5), 719–732.
- Leathwick, J., Austin, M., 2001. Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology* 82 (9), 2560–2573.
- Leathwick, J.R., Mitchell, N.D., 1992. Forest pattern, climate, and vulcanism in central north-island, New Zealand. *J. Veg. Sci.* 3 (5), 603–616.
- Leckie, D.G., Gougeon, F.A., Tinis, S., Nelson, T., Burnett, C.N., Paradine, D., 2005. Automated tree recognition in old growth conifer stands with high resolution digital imagery. *Remote Sens. Environ.* 94, 311–326.
- Lees, B.G., Ritman, K., 1991. Decision-tree and rule-induction approach to integration of remotely sensed and GIS data in mapping vegetation in disturbed or hilly environments. *Environ. Manage.* 15, 823–831.
- Le Hegarat-Masclé, S., Quesney, A., Vidal-Madjar, D., Taconet, O., Normand, M., Loumagne, C., 2000. Land cover discrimination from multitemporal ERS images and multispectral Landsat images: a study case in an agricultural area in France. *Int. J. Remote Sens.* 21 (3), 15.
- Lehman, A., Overton, J.M., Leathwick, J.R., 2002. GRASP: generalized regression analysis and spatial predictions. *Ecol. Model.* 157, 187–205.
- Loreau, M., Naeem, S., Inchausti, P., 2001. Biodiversity and ecosystem functioning: current knowledge and future challenges. *Science* 294, 804–808.
- Martin, M.E., Newman, S.D., Aber, J.D., Congalton, R.G., 1998. Determining forest species composition using high spectral resolution remote sensing data. *Remote Sens. Environ.* 65, 249–254.
- McCann, K.S., 2000. The diversity–stability debate. *Nature* 405, 228–233.
- McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*. Chapman & Hall, London.
- McCune, B., Keon, D., 2002. Equations for potential annual direct incident radiation and heat load. *J. Veg. Sci.* 13, 603–606.
- McIver, D.K., Friedl, M.A., 2002. Using prior probabilities in decision tree classification of remotely sensed data. *Remote Sens. Environ.* 81, 253–261.
- Moisen, G.G., Frescino, T.S., 2002. Comparing five modeling techniques for predicting forest characteristics. *Ecol. Model.* 157, 209–225.
- Muller, F., 1998. Gradients in ecological systems. *Ecol. Model.* 108, 3–21.

- Olea, R.A., 1990. *Geostatistical Glossary and Multilingual Dictionary*. Oxford University Press, New York.
- Parker, A.J., 1982. The topographic relative moisture index: an approach to soil moisture assessment in mountain terrain. *Phys. Geogr.* 3, 160–168.
- Peddle, D.R., 1995. Knowledge formulation for supervised evidential classification. *Photogram. Eng. Remote Sens.* 61, 409–417.
- Pedroni, L., 2003. Improved classification of Landsat thematic mapper data using modified prior probabilities in large and complex landscapes. *Int. J. Remote Sens.* 24 (1), 91–113.
- Pohl, C., Van Genderen, J.L., 1998. Multisensor image fusion in remote sensing: concepts, methods and applications. *Int. J. Remote Sens.* 19 (5), 823–854.
- Prasad, N., Saran, S., Kushwaha, S.P.S., Roy, P.S., 2001. Evaluation of various image fusion techniques and imaging scales for forest features interpretation. *Curr. Sci. (Bangalore)* 81 (9), 1218–1224.
- Rogers, J.H., 1974. Soil survey of the Tahoe Basin area, California and Nevada. In: U.S.C. Service (Editor). Washington, DC.
- Sawaya, K.E., Olmanson, L.G., Heinert, N.J., Brezonik, P.L., Bauer, M.E., 2003. Extending satellite remote sensing to local scales: land and water resource monitoring using high-resolution imagery. *Remote Sens. Environ.* 88 (1–2), 144–156.
- Solberg, A., 1999. Contextual data fusion applied to forest map revision. *IEEE Trans. Geosci. Remote Sens.* 37 (3), 1234–1243.
- Strahler, A.H., 1980. The use of prior probabilities in maximum-likelihood classification of remotely sensed data. *Remote Sens. Environ.* 10 (2), 135–163.
- Strahler, A.H., Woodcock, C.E., Smith, J.A., 1986. On the nature of models in remote sensing. *Remote Sens. Environ.* 20, 121–139.
- Thenkabail, P.S., Enclona, E.A., Ashton, M.S., Legg, C., De Dieu, M.J., 2004. Hyperion, IKONOS, ALI, and ETM+ sensors in the study of African rainforests. *Remote Sens. Environ.* 90 (1), 23–43.
- Underwood, E., Ustin, S.L., DiPietro, D., 2003. Mapping nonnative plants using hyperspectral imagery. *Remote Sens. Environ.* 86, 150–161.
- Wald, L., 1999. Some terms of reference in data fusion. *IEEE Trans. Geosci. Remote Sens.* 37 (3), 1190–1193.
- Whittaker, R.H., 1951. A criticism of the plant association and climatic climax concepts. *Northwest Sci.* 25, 17–31.
- Wilfried, T., Araujo, M.B., Lavorel, S., 2003. Generalized models vs. classification tree analysis: predicting spatial distributions of plant species at different scales. *J. Veg. Sci.* 14, 669–680.
- Winthers, E., Fallon, D., Haglund, J., Demeo, T., Tart, D., Ferwerda, M., Robertson, G., Gallegos, A., Rorick, A., Cleland, D.T., Robbie, W., Shadis, D., 2003. *Terrestrial Ecological Unit Inventory Technical Guide*. In: U.F.S.E.M.C. Staff (Editor), 245 pp.
- Wulder, M., 1998. Optical remote sensing techniques for the assessment of forest inventory and biophysical parameters. *Prog. Phys. Geogr.* 22 (4), 449–476.
- Zarco-Tejada, P.J., Miller, J.R., 1999. Land cover mapping at BOREAS using red edge spectral parameters from CASI imagery. *J. Geophys. Res.* 104 (D22), 921–933.