A TM-BASED HARDWOOD-CONIFER MIXTURE INDEX FOR CLOSED CANOPY FORESTS IN THE OREGON COAST RANGE

by

Thomas K. Maiersperger

A RESEARCH PAPER

submitted to

THE GEOSCIENCES DEPARTMENT

in partial fulfillment of the
requirements for the
degree of

MASTER OF SCIENCE

GEOGRAPHY PROGRAM

May 1997

Directed by
Dr. A. Jon Kimerling

ACKNOWLEDGMENTS

I am particulary grateful to Dr. Warren B. Cohen for his guidance and insight, and to Dr. David Wallin for his enthusiastic support of this work. I would also like to thank Dr. A. Jon Kimerling and Dr. Julia Jones for their valuable input and instruction. The statistical assistance of Lisa Ganio is gratefully acknowledged. This research was funded by NASA Grant NAGW-3745 and the H.J. Andrews LTER.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

A TM-BASED HARDWOOD-CONIFER MIXTURE INDEX FOR CLOSED CANOPY FORESTS IN THE OREGON COAST RANGE

ABSTRACT

The purpose of this study was to develop, implement, and test methods for quantifying the relative proportion of hardwood and conifer cover from Thematic Mapper (TM) imagery. The research was focused on closed canopy forests in the Oregon Coast Range, where hardwood, conifer, and mixed stand conditions are prevalent. Based on an understanding of the patterns of spectral variation expressed by these forests in TM data space, it was hypothesized that a vegetation index could be developed to measure hardwood-conifer mixing proportions. An approach based on the Gramm-Schmidt orthogonalization process was used to derive three slightly different hardwood-conifer mixture indices (HCMIs). Using correlation and regression techniques, the effectiveness of these indices as a measure of closed canopy hardwood proportion was compared with three other groups of spectral variables: (1) the untransformed TM reflectance bands, (2) the tasseled cap indices of brightness, greenness, and wetness, and (3) the first three principal components of closed canopy forest pixels. Results indicate that the Gramm-Schmidt process was an effective method for deriving an index that was strongly correlated with closed canopy hardwood proportion ($r = 0.82$).

1. INTRODUCTION

Patterns of land cover in the Pacific Northwest have undergone significant changes due to natural and anthropogenic disturbances. An important forest management issue in the region is how these temporal alterations of landscape structure affect biological diversity. The work presented here supports regional scale modeling of the presence and abundance of several vertebrate species using remotely-sensed measures of forest structure and composition derived from Thematic Mapper (TM) data. Recent studies in western Oregon have established techniques for quantifying several structural attributes of closed canopy conifer forests from TM data (Cohen *et al.* 1995, Cohen and Spies 1992). However, the relative amount of hardwood and conifer cover occurring within the forest is a particularly important explanatory variable for biodiversity modeling, and methods for measuring this attribute in the region are less well developed. Based on an understanding of spectral variation in closed canopy forests, it was hypothesized that a vegetation index could be developed from TM reflectance data to measure continuous hardwood-conifer mixtures in a simple and effective manner.

Vegetation indices are produced by transforming the original multiband data into a lesser number of image variables that are strongly related to the physical phenomena of interest. Several kinds of continuous phenomena have been remotely sensed using vegetation indices. Examples of crop information measured by indices include density (Kauth and Thomas 1976), biomass, leaf water content, and chlorophyll content (Tucker 1979), and leaf area (Weigand and Richardson 1982). In forest systems, multispectral indices have been used to estimate attributes such as the basal area and biomass of conifer stands (Franklin 1986), the size, density, and age of conifer stands (Cohen and Spies 1992), and conifer mortality (Collins and Woodcock 1994). The goal of this study was to extend the development and application of vegetation indices to the problem of measuring hardwood-conifer compositional mixtures.

Spectral indices may be calculated in many ways, including band ratios, band differences, and linear band combinations. In this research, the Gramm-Schmidt orthogonalization process (Jackson 1983) was used to derive a hardwood-conifer mixture index (HCMI). The Gramm-Schmidt process is a mathematically simple method for calculating the coefficients of one or more linear combinations of multiband data using the spectral response of a few suitable reference points. The coefficients represent an axis of spectral variation between reference endpoints in the multiband data space. If the physical phenomenon of interest produces a continuum of spectral response between two distinct points in the data space, the coefficients should yield a useful index.

It was expected that the Gramm-Schmidt process would be a suitable approach for deriving an HCMI based on exploratory analyses of closed canopy forest response. Using graphical representations of the TM data space, two main directions of spectral variation in closed canopy conditions were observed. The primary direction occurred within pure conifer conditions, and corresponded to the increasing development of canopy structure with age (e.g., multiple layers, large number of gaps, tree size variability, and overall canopy roughness). The secondary direction of spectral variation occurred between conifer and pure hardwood conditions, and appeared to correspond to the relative mixing proportions between the two composition types. Pure hardwood conditions exhibited a relatively small amount of spectral variation. This observation is supported by previous studies that suggest that the simple structure of dense hardwood canopies do not show significant spectral change with age (Horler and Ahern 1986, Spanner *et al.* 1984).

Based on these observations, a two-step strategy emerged to separate compositional information from structural information using two independent indices. First, the spectral variation associated with closed canopy forest structure was addressed by establishing an initial axis through young conifer and old conifer reference points. This axis is termed the canopy structure index (CSI). Subsequently, a second axis was defined orthogonal to the first toward a pure hardwood reference point to produce coefficients of the hardwood-conifer mixture index (HCMI). Reference points for index formulation were selected using methods of spectral space visualization and analysis (Esbensen and Geladi 1989, Johnson *et al.* 1985).

The effectiveness of the HCMI as a measure of closed canopy hardwood proportion was examined relative to three other groups of spectral variables: (1) the untransformed TM reflectance bands, (2) the TM tasseled cap indices of brightness, greenness, and wetness (Crist *et al.* 1986), and (3) the first three principal components of closed canopy forest pixels. The association between hardwood proportion and each of the spectral variables was determined using correlation techniques. In addition, regression models were generated and assessed for each group of spectral variables to determine the relative predictive strength of the HCMI for quantifying hardwood proportion. Cosine of the solar incidence angle (COSI) (Smith *et al.* 1980) was also included in the analyses to provide information about the sensitivity of the spectral variables to topographically-induced variation in illumination conditions.

2. METHODS

2.1 Study Area

The study area for this project was an 800,000 hectare section of the Oregon Coast Range (figure 1). The region was defined by the intersection of TM scene 46/29 with the Willamette Valley margin on the east and the Pacific coast on the west. A maritime climate prevails in this area, with mild, wet winters and warm, dry summers (Franklin and Dyrness 1988). These seasonal conditions are largely responsible for the natural dominance of conifer species over deciduous hardwood species in the coastal forests (Franklin and Dyrness 1988). The study area is composed of two major vegetation zones: the *Picea sitchensis* Zone and the *Tsuga heterophylla* Zone (Franklin and Dyrness 1988). The *Picea sitchensis* Zone, a narrow strip adjacent to the ocean, is characterized by slightly wetter and milder conditions than the remainder of the study area that falls in the *Tsuga heterophylla* Zone (Franklin and Dyrness 1988).

The most important conifer species in the study area are Douglas fir and western hemlock, while red alder is the most common hardwood species (table 1). Conifer species are dominant in the region, but hardwoods proliferate in specialized habitats (e.g., riparian zones) and rapidly colonize disturbed sites (Franklin and Dyrness 1988). Because of extensive historical disturbances from fire and timber extraction, hardwood and mixed hardwood-conifer forests are significant features of the central Oregon Coast Range.

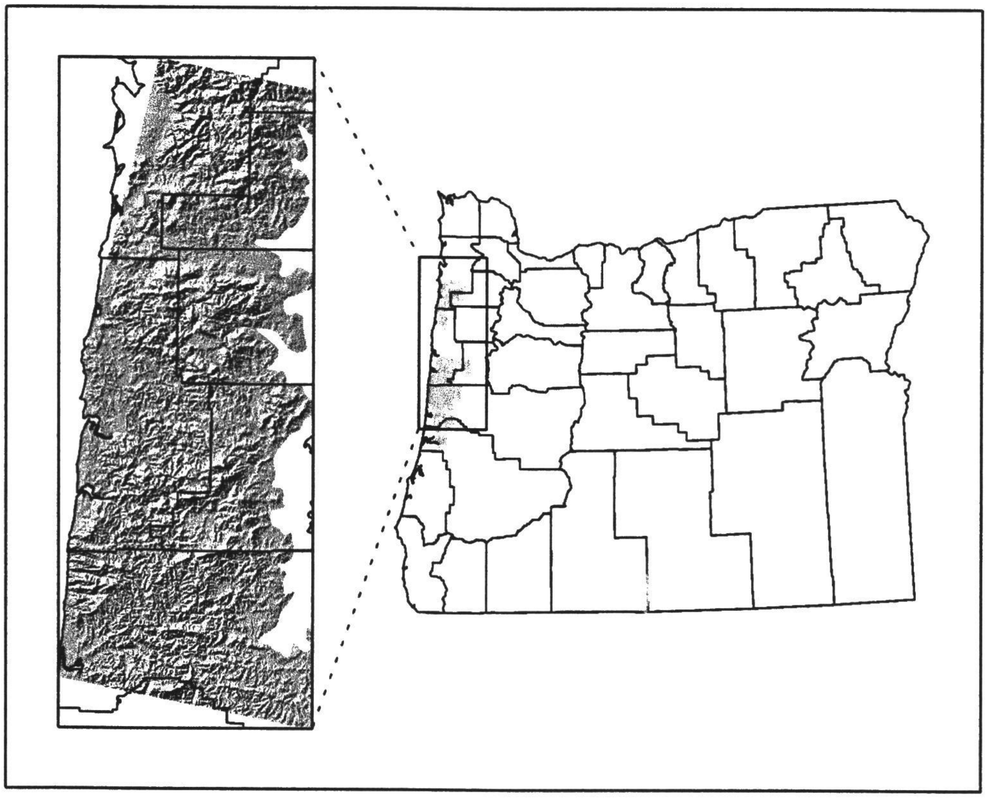

Figure 1. Location of the study area in Oregon. The gray area on the Oregon state map corresponds to the Siuslaw National Forest. The enlargement shows a shaded relief image bounded by the study area. County boundaries are also shown.

Table 1. Relative Importance of Hardwood and Conifer Species by Vegetation Zone (after Franklin and Dyrness 1988).

Species	*Common Name*	*Picea sitchensis Zone*	*Tsuga heterophylla Zone*
Abies amabilis	Pacific silver fir	minor	minor
Abies grandis	grand fir	minor	minor
Picea sitchensis	Sitka spruce	major	minor
Pinus contorta	lodgepole pine	minor	minor
Pseudotsuga menziesii	Douglas fir	major	major
Tsuga heterophylla	western hemlock	major	major
Alnus rubra	red alder	major	major
Acer macrophyllum	bigleaf maple	minor	minor

2.2 Image and DEM Data

A Landsat Thematic Mapper (TM) scene (Landsat 5, Path 46, Row 29) dating from August 29, 1988 was used in this study. Preprocessing yielded 25 meter resolution data georeferenced to the Universal Transverse Mercator (UTM) grid coordinate system. In addition, 1:250,000 scale digital elevation models (DEM) from the USGS were used. The DEM data were mosaicked, converted to UTM, and clipped to the study area. The resulting pixel resolution was 63.3 meters in the x direction and 92.5 meters in the y direction.

2.3 Reference Data

Reference data for the project were selected from a database of 913 photointerpreted polygons distributed throughout the Coast Range and registered to the satellite imagery. This database was compiled cooperatively by Oregon State University, the USDA Forest Service and the USDI Bureau of Land Management. A team of experienced photointerpreters used 1:12,000 aerial photos from the summer months of 1988 and 1989 to estimate the proportion of the following stand components for each polygon: conifer tree cover, hardwood tree cover, brush cover, and open (i.e., non-vegetated or dead vegetation). Additionally, many of the stands had supporting ground survey data associated with them. The ground data were important for ascertaining the age-related structure of certain stands during the analysis. A total of 330 closed canopy forest stands (100% tree cover with various proportions of hardwood and conifer cover) fell within the study area and were used in this study.

2.4 Data Analysis

Data analysis was comprised of three main objectives: (1) to generate the HCMI, tasseled cap, principal component, and cosine of the solar incidence angle variables for closed canopy forest conditions, (2) to conduct correlation and regression analyses, and (3) to assess the error of the regression models. These objectives were met in several steps using image processing and statistical software (ERDAS Imagine Version 8.2, SAS/STAT Version 6.10).

2.4.1 *Data Stratification by Closed Canopy Forest Conditions*

This study was concerned only with closed canopy forest (CCF) conditions, so it was necessary to exclude any image or reference data containing non-forest elements (e.g., soil, water, cloud) from the analysis. Both the image and reference data were stratified using a single approach based on the visual selection of relevant classes in spectral space (Cetin *et al.* 1993, Esbensen and Geladi 1989). The procedure had three phases: data reduction and enhancement, data visualization, and analyst-based classification.

The purpose of data reduction and enhancement was to transform the six band data into two image variables that provided the best visualization of spectral space for the CCF class. For this, the brightness (BRT) and greenness (GRN) indices were used because previous studies (Crist *et al.* 1986, Cohen *et al.* 1995) have shown that a distinct region corresponding to dense forest conditions occurs in the "Plane of Vegetation" (i.e., BRT-GRN space). Wetness (WET) was also calculated because all three tasseled cap indices were required for the subsequent correlation and regression analyses.

Data visualization was accomplished by plotting the locations of every reference data pixel on top of the distribution of every image pixel in BRT-GRN space. This technique allowed the relationship between the reference and image data distributions to be quickly interpreted. As expected, the reference data formed an elliptical cloud along a diagonal axis within BRT-GRN space (figure 2). Apart from the heavy concentration of reference data pixels in this closed canopy forest region, several pixels were scattered across other locations in BRT-GRN space. These pixels were generally much brighter than those falling within the main concentration of reference data. These outliers were noted, and the stands to which they belonged were examined in the imagery. Each of these stands exhibited properties consistent with small areas of non-green vegetation (e.g., road, soil, dead vegetation). Subsequently, sixty-four stands having such outliers were removed from the analysis, leaving a total of 266 reference stands.

The classification of closed canopy forest involved digitizing an elliptical boundary to enclose the dense cloud of remaining CCF reference pixels in BRT-GRN space (figure 2). The ellipse was used by the computer as a decision region to classify the imagery on a per-pixel basis. The resulting CCF class was used to mask the six band TM data, the tasseled cap indices, and the DEM data.

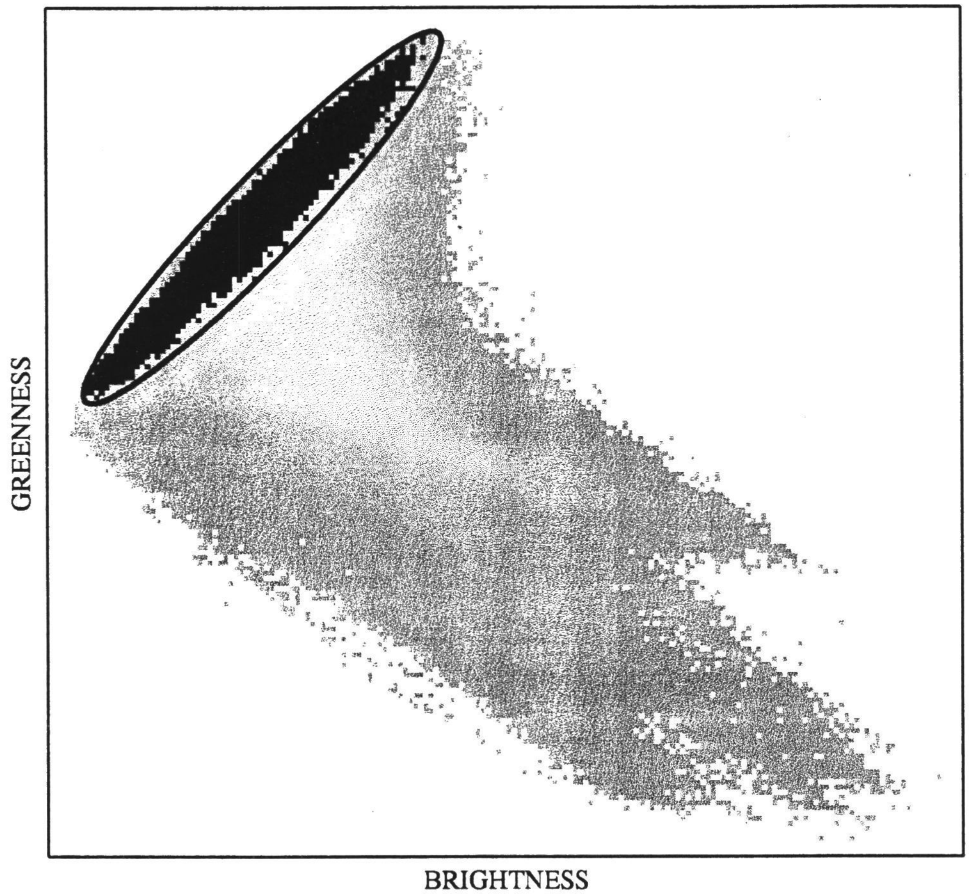

Figure 2. Spectral space classification. The frequency distribution of all the image pixels in BRT-GRN space is shown by the continuum between shades of darker gray (low frequency) and lighter gray (high frequency). The cloud of black points represents the distribution of all the closed canopy forest (CCF) reference pixels in BRT-GRN space after the removal of outliers. The ellipse was digitized to enclose the reference data and used as a decision region to produce a CCF mask.

2.4.2 *Cosine of the Solar Incidence Angle Calculation*

The Oregon Coast Range is a topographically complex landscape, and variable illumination conditions were identified as a potential confounding factor in the modeling of hardwood-conifer proportions. In an effort to address this problem, cosine of the solar incidence angle (COSI) (Smith *et al.* 1980) was calculated as an additional image variable. Slope and aspect were derived from the DEM data and used in the formula to calculate COSI along with the solar elevation and azimuth from the time of TM image acquisition.

2.4.3 *Principal Component Analysis*

After masking the six band TM image with the CCF class, a standardized principal component (PC) analysis was run to capture the major directions of spectral variation in CCF pixels. The PC analysis provided a purely statistical approach for seeking potentially useful variables for quantifying hardwood-conifer proportions from the imagery. The analysis also provided a technique for data reduction and enhancement for use in the HCMI development process. The coefficients of each component (eigenvectors) were applied to the six band CCF image to produce a three component CCF image.

2.4.4 *Development of Canopy Structure and Hardwood-Conifer Mixture Indices*

In this study, vegetation indices were derived using an algebraic formulation of the Gramm-Schmidt process (Jackson 1983). The spectral response of young conifer, old conifer, and pure hardwood reference points were required as inputs to the Gramm-Schmidt calculations. The young and old conifer points were needed to define a canopy structure index

(CSI) in the TM data space, while the hardwood point was needed to define a hardwood-conifer mixture index (HCMI) orthogonal to the CSI. These reference points were selected in principal component space using an extension of the methods described by Johnson *et al.* (1985) to identify the response of pure materials for a mixture modeling application.

The first and second principal components of the six band CCF image (as described in section 2.4.3) were used to construct a PC1-PC2 plot (figure 3). These two components accounted for 97.8% of the six band spectral variation in CCF pixels, and therefore provided a concise visualization of the TM data structure for that forest condition. Ten stands, including the youngest and oldest conifer samples, were chosen from the photointerpreted reference data to represent pure hardwood conditions and a range of pure conifer age conditions (table 2). The mean response of these stands (the crosses labeled one through ten in figure 3), and the response of each pixel in the stands (not shown in the interest of clarity) were plotted in PC1-PC2 space. As expected, old conifer, young conifer, and pure hardwood conditions occupied distinct regions near the edges of the data space. Consequently, pixels occupying these extreme regions were considered candidates for use as index reference points.

While the general location of candidate pixels for young conifer, old conifer, and hardwood reference points was obvious (figure 3), it was unclear which of many possible selections would produce the best HCMI. This problem was addressed by collecting three separate groups of pixels so that three slightly different sets of coefficients could be derived and compared. The first and second group (OC1, YC1, HD1 and OC2, YC2, HD2 in figure 3) corresponded to actual reference stand pixels that were near the old conifer, young conifer, and hardwood edges of PC1-PC2 space. The third set (OC3, YC3, HD3) corresponded to

Table 2. Characteristics of Reference Stands Plotted into PC1-PC2 Space. Age Data was not Available for All Stands.

Sample	*Cover Type*	*Age*
1	Conifer	400
2	Conifer	115
3	Conifer	110
4	Conifer	-
5	Conifer	80
6	Conifer	78
7	Conifer	18
8	Hardwood	-
9	Hardwood	-
10	Hardwood	-

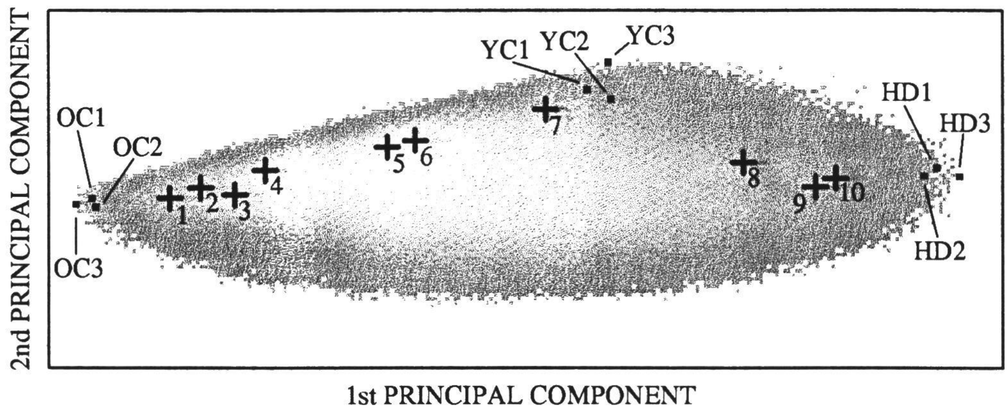

Figure 3. Selection of index reference points in PC1-PC2 space. The frequency distribution of all the CCF pixels in PC1-PC2 space is shown by the continuum between shades of darker gray (low frequency) and lighter gray (high frequency). The crosses labeled 1-10 indicate the mean PC1-PC2 response of the reference stands in table 2. Three sets of reference points representing old conifer, young conifer, and pure hardwood conditions (OC1/ YC1/ HD1, OC2/YC2/HD2, and OC3/ YC3/HD3) were selected visually from pixels in the PC1-PC2 plot. The six band TM spectral response was determined for each endpoint (table 3). The Gramm-Schmidt process was applied to the digital numbers in table 3 to calculate the coefficients for three slightly different canopy structure and hardwood-conifer mixture indices (CSI 1-3 and HCMI 1-3).

pixels that were at the very extremes of the image data envelope. The six band spectral response of the index endpoints (table 3) was found by locating the pixels of interest in the original TM imagery. The Gramm-Schmidt process was applied to the digital numbers in table 3 to produce the coefficients of three different canopy structure indices and three different hardwood-conifer mixture indices.

2.4.5 *Correlation and Regression Analyses*

The photointerpreted reference stands (n=266) were randomly partitioned into two equal groups, one for the correlation analysis and regression model building, and one for regression model validation. The correlation analysis was used to quantify the relationship between stand hardwood proportion (as interpreted from aerial photography) and the mean stand response of each image variable in the following groups: (1) the TM reflectance bands (TM1, TM2, TM3, TM4, TM5, TM7), (2) brightness, greenness, and wetness (BRT, GRN, WET), (3) principal components one, two, and three (PC1, PC2, PC3), and (4) the three hardwood-conifer mixture indices (HCMI1, HCMI2, HCMI3). In addition, the correlation between cosine of the solar incidence angle (COSI) and each of the aforementioned image variables was calculated. Scatter plots were examined, and the relationship between the image variables and hardwood proportion appeared linear. This observation was supported by the fact that no data transformations were found to improve the correlations (e.g., log, square root, square, arcsin-square root).

A simple regression of hardwood proportion on the best HCMI (i.e., the index most highly correlated with hardwood proportion) was performed. For each of the other image

Table 3. Spectral Response (Digital Number) of Pixels Selected as Index Reference Points.

Point	*Cover Type*	*TM1*	*TM2*	*TM3*	*TM4*	*TM5*	*TM7*
OC1	Old Conifer	60	17	14	20	9	2
YC1	Young Conifer	66	24	21	118	46	10
HD1	Hardwood	72	28	33	178	95	25
OC2	Old Conifer	59	17	16	21	7	2
YC2	Young Conifer	67	25	21	114	42	9
HD2	Hardwood	72	28	24	180	95	22
OC3	Old Conifer	59	18	14	14	6	1
YC3	Young Conifer	69	25	21	120	39	9
HD3	Hardwood	72	33	27	183	101	29

variable groups (i.e., original bands, tasseled cap indices, principal components), a forward stepwise multiple regression approach was used to generate predictive models. For any given stepwise model, the criteria for accepting or dropping an explanatory variable was significance at the 0.05 level, and variables entered the model in order of highest significance. Stepwise modeling was performed again after adding the COSI variable as a potential predictor to the HCMI and other variable groups. Inclusion of COSI allowed for the comparison of models with and without explicit information related to topographically-induced variations in spectral response.

2.4.6 *Error Assessment*

The error of each regression model was assessed using the independent group of validation stands (n=133) set aside earlier. For any given model, the predicted hardwood proportion for each stand was calculated by applying the parameter estimates to the mean response of the appropriate image variables. The coefficient of determination (R-square) was calculated for each model based on the fit of predicted (modeled) versus observed (photointerpreted) hardwood proportion values. The root-mean-square error (RMSE) was also calculated to provide a measure of average prediction error.

Since thematic maps are often needed by the end user of satellite data, a discrete accuracy assessment was conducted to determine the percentage of stands correctly classified by each model. Three different class structures were created so that the overall accuracies could be compared in terms of potentially useful sets of hardwood proportion classes. The different class structures were: five classes (0-20%, 21-40%, 41-60%, 61-80%, 81-100%), four classes (0-25%, 26-50%, 51-75%, 76-100%), and three classes (0-30%, 31-70%, 71-100%). The predicted and observed hardwood proportion were used to determine class membership for each validation stand for each set of classes. In any given instance, if the predicted value was less than zero, the stand was placed into the smallest proportion class. Conversely, if the predicted value was greater than 100%, the stand was assigned to the largest class. Subsequently, the overall percent correctly classified was calculated for each set of classes for each of the predictive models (Congalton 1991).

2.4.7 *Evaluation of CSI-HCMI Information Content*

While the primary emphasis of this research concerned the development and testing of a hardwood-conifer mixture index, the general characteristics of the transformed (CSI-HCMI) data space was also evaluated. From the full set of reference plots (n=266), a number of samples was selected at random from each of the six following categories: hardwood stands (0-25% conifer cover, n=7), hardwood-dominated mixed stands (26-50% conifer cover, n=7), conifer-dominated mixed stands (51-75% conifer cover, n=7), young conifer stands (76-100% conifer cover, <100 years old, n=3), medium aged conifer stands (76-100% conifer cover, 100-200 years old, n=3), and old conifer stands (76-100% conifer, >200 years old, n=3). The mean spectral response of each stand (total n=30) was plotted into CSI-HCMI space to visualize forest information trends in the transformed data structure.

3. RESULTS

3.1 Index Coefficients

The coefficients of the tasseled cap, principal component, CSI, and HCMI transformations are shown in table 4. Although the tasseled cap coefficients were not derived in this study, they are presented for comparative purposes. Examination of these coefficients provides insight into the relative importance of each TM band in the various transformations.

Brightness, the first tasseled cap index, has positive loadings in all reflectance bands, and corresponds to overall scene brightness (Crist and Cicone 1984). Greenness, like many other correlates of vegetation amount (e.g., NDVI) is a contrast between the visible bands (especially TM3) and the near-infrared (TM4). Wetness presents a contrast of the visible and near-IR bands (weak positive loadings) with the mid-IR bands (strong negative loadings).

The first three principal components contained 98.1% of the six band spectral variation in closed canopy forest pixels. PC1 (92%) appears to be a measure of greenness, with a strong contrast between TM3 and TM4. PC2 (5.8%) is a contrast of TM3 and TM4 with the mid-IR bands. The weights and loadings for PC3 (0.3%) are irregular and inconsistent across the visible, near-IR, and mid-IR bandwidth categories.

Using three slightly different sets of reference points, three canopy structure indices (CSI 1-3) and three hardwood-conifer mixture indices (HCMI 1-3) were derived. The coefficients of the CSIs are reported but not examined further. Not surprisingly, HCMIs 1-3 have similar TM band loadings. The most important band was TM5, with strong positive loadings in each index. Moderately strong loadings were also found for TM bands 4 (negative)

Table 4. Coefficients of the Tasseled Cap, Principal Component, Canopy Structure, and Hardwood-Conifer Mixture Indices.

Index	*TM1*	*TM2*	*TM3*	*TM4*	*TM5*	*TM7*
BRT	0.2909	0.2493	0.4806	0.5568	0.4438	0.1706
GRN	-0.2728	-0.2174	-0.5508	0.7221	0.0733	-0.1648
WET	0.1446	0.1761	0.3322	0.3396	-0.6210	-0.4186
PC1	0.0687	-0.0953	-0.8570	0.4569	-0.1846	-0.0940
PC2	0.0701	-0.0554	-0.3114	-0.3347	0.2383	0.8522
PC3	0.0699	-0.1124	-0.3621	-0.6641	0.3831	-0.5134
CSI1	0.0568	0.0662	0.0662	0.9272	0.3501	0.0757
CSI2	0.0797	0.0797	0.0498	0.9265	0.3487	0.0697
CSI3	0.0891	0.0624	0.0624	0.9448	0.2941	0.0713
HCMI1	0.0630	-0.0361	0.2563	-0.3596	0.8272	0.3399
HCMI2	-0.0536	-0.1260	-0.0380	-0.3381	0.8917	0.2652
HCMI3	-0.0962	0.0691	0.0229	-0.2982	0.8875	0.3301

and 7 (positive). The visible bands were generally less important than the infrared portion of the electromagnetic spectrum, with weak loadings of variable weights. The HCMIs appear similar to an inverse of the tasseled cap wetness feature. However, the visible bands are even less important than in wetness, and these indices may best be interpreted as a contrast between the near-IR and mid-IR wavelengths with particular emphasis on TM5.

3.2 Correlation and Regression Analyses

Table 5 shows the correlations between stand hardwood proportion, COSI, and the image variables. All correlations with hardwood proportion were positive except for wetness and the second principal component. TM bands 5 and 7 were both highly correlated with stand hardwood proportion ($r = 0.80$). The tasseled cap and principal component variables did not yield stronger correlations, with brightness ($r = 0.69$) and PC1 ($r = 0.68$) having the best association with hardwood proportion. HCMI3 produced the highest correlation coefficient of any image variable ($r = 0.82$). The other two HCMIs were not as highly correlated with hardwood proportion. Based on these results, HCMI3 was selected for the regression modeling phase.

All correlations with the COSI variable were positive. Overall, the transformed image variables exhibited weaker relationships with COSI than the original TM band data. The WET, PC2, and PC3 variables showed extremely low, non-significant correlations with COSI. The HCMI1, HCMI2, HCMI3, and TM1 variables exhibited moderately weak associations with COSI, with correlation coefficients ranging between 0.32 and 0.40. The remaining image variables produced higher correlations, ranging between 0.49 and 0.56.

Table 5. Correlation (r) of Image Variables with Hardwood Proportion and COSI for 133 Reference Stands.

Image Variable	*Hardwood Proportion*	*COSI*
TM1	0.21*	0.32**
TM2	0.50**	0.56**
TM3	0.52**	0.55**
TM4	0.63**	0.52**
TM5	0.80**	0.52**
TM7	0.80**	0.51**
BRT	0.69**	0.54**
GRN	0.60**	0.49**
WET	-0.46**	0.00
PC1	0.68**	0.53**
PC2	-0.19*	0.11
PC3	0.37**	0.01
HCMI1	0.73**	0.32**
HCMI2	0.79**	0.33**
HCM13	0.82**	0.40**

* p-value < 0.05, ** p-value < 0.01

The results of regressing stand hardwood proportion on HCMI3 and each group of image variables (both with and without COSI) appear in table 6. The COSI variable was a statistically significant predictor when added to each model, except when coupled with HCMI3. However, COSI did not improve the R-square value of any model by more than 2%. The untransformed band data explained the highest amount of variance in hardwood proportion (75%, 76% with COSI). The tasseled cap and principal component models explained similar, but slightly less amounts of the variance. Sixty-seven percent of the variance in stand hardwood proportion was explained by the simple regression on HCMI3 response.

3.3 Error Assessment

Due to the difficulty in collecting accurate reference data at the pixel scale, regression models are often generated and tested using stand level data (e.g., Cohen and Spies 1992, Ripple *et al.* 1991, Butera 1986). To test whether the stand level models were valid when applied at the pixel level, the correlations between stand and pixel level estimates of hardwood proportion from each regression model were calculated. First, the mean spectral response per image variable per stand was calculated and the appropriate model parameters were applied to derive an estimate of hardwood proportion. Then the regression models were applied to the imagery on a per-pixel basis, and the mean hardwood proportion was calculated for each stand for each model. The correlation between the stand and pixel level estimates were very high for all models, ranging between 0.988 and 0.996. This result supports the validity of using mean stand response to develop models applied at the pixel level for this problem.

Table 6. Regression Model Statistics.

Image Variable	*Parameter Estimate*	*Standard Error*	*Partial R-square*	*Model R-square*
Intercept	57.64			
TM7	-13.11	4.89	0.65	
TM2	-4.13	1.42	0.05	
TM4	-1.11	0.25	0.04	
TM5	7.37	1.49	0.01	0.75
Intercept	53.23			
TM7	-12.29	4.83	0.65	
TM2	-3.40	1.44	0.05	
TM4	-1.05	0.25	0.04	
COSI	-25.09	11.25	0.01	
TM5	7.15	1.47	0.01	0.76
Intercept	34.51			
BRT	0.34	0.34	0.48	
WET	-4.42	0.49	0.23	
GRN	1.04	0.48	0.01	0.72
Intercept	-24.78			
BRT	1.19	0.08	0.48	
WET	-3.74	0.35	0.23	
COSI	-36.04	11.30	0.02	0.73
Intercept	55.90			
PC1	0.91	0.06	0.47	
PC2	-2.67	0.29	0.21	
PC3	1.95	0.52	0.03	0.71
Intercept	54.19			
PC1	1.01	0.07	0.47	
PC2	-2.75	0.28	0.21	
PC3	1.72	0.51	0.03	
COSI	-33.40	11.48	0.02	0.73
Intercept	-10.87			
HCMI3	3.77	0.23	0.67	0.67

Results of the continuous error assessment are shown in table 7. The R-square values were moderately high for all models, ranging from 0.72 to 0.78. These values were similar to, and reflected the same trends as the model R-squares (table 6). The average prediction error (as measured by the RMSE) ranged between 14% and 16% hardwood cover. Including COSI in the models increased the R-square and decreased the RMSE values slightly.

Table 8 shows the results of the discrete (class) accuracy assessment. As expected, overall accuracy increased as the number of classes was reduced (Cohen *et al.* 1995). The overall accuracy ranged from 60% to 70% for five hardwood proportion classes (0-20%, 21-40%, 41-60%, 61-80%, 81-100%). Reducing the number of classes to four (0-25%, 26-50%, 51-75%, 76-100%) increased the overall accuracy by between 7% and 16%. The best overall accuracies were achieved with three proportion classes (0-30%, 31-70%, 71-100%). The models achieved accuracies differing by only 2% and 3% within the four and three class layouts respectively. While COSI was found to improve the model R-squares (table 6), and the predicted versus observed R-squares and RMSEs (table 7), this variable tended to degrade the overall classification accuracy slightly for the predictive models.

Table 7. Regression Model Error Assessment (Predicted vs. Observed Stand Hardwood Proportion, n=133).

Model	*R-square*	*RMSE*
TM 7245	0.77	14.42
TM 7245&COSI	0.78	14.01
BGW	0.76	14.84
BW&COSI	0.78	14.29
PC 123	0.75	15.19
PC 123&COSI	0.77	14.59
HCMI3	0.72	16.04

Table 8. Regression Model Error Assessment (Overall Percent Correctly Classified for Three Class Structures, n=133).

Model	*5 Classes*	*4 Classes*	*3 Classes*
TM 7245	70	77	84
TM 7245&COSI	69	76	83
BGW	64	78	83
BW&COSI	63	77	83
PC 123	63	78	83
PC 123&COSI	65	77	83
HCMI3	60	76	81

3.4 Evaluation of CSI-HCMI Information Content

Because HCMI3 was the strongest hardwood-conifer mixture index, this variable along with CSI3 (the canopy structure index associated with HCMI3) were selected for further evaluation of combined information content. The location of the six selected forest types in CSI-HCMI space is shown in figure 4. It is important to note that low values of the CSI correspond to more complex canopies, while higher CSI values coincide with simpler canopies. This counter-intuitive relationship is a function of the process used to define the CSI spectral axis (i.e., the index originates at an old conifer reference point and extends toward a young conifer reference point). The shape of the transformed data envelope is narrow at low values of CSI and HCMI (complex canopy structure, high proportion of conifer cover), and tapers out toward higher values of CSI and HCMI (simple canopy structure, high proportion of hardwood cover). The continuum between conifer and hardwood cover is captured by the HCMI axis, while separation between the conifer age classes is apparent along the CSI.

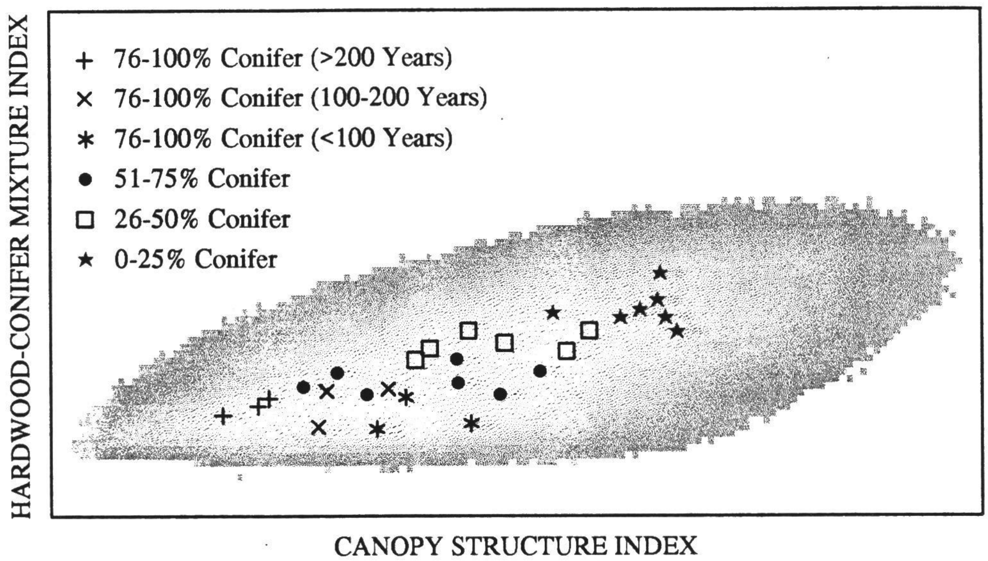

Figure 4. Evaluation of CSI-HCMI information content. The frequency distribution of all the CCF pixels in CSI-HCMI space is shown by the continuum between shades of darker gray (low frequency) and lighter gray (high frequency). Symbols indicate the mean response from samples of six categories of reference stands (total n=30).

4. DISCUSSION AND CONCLUSIONS

The results of this study support the hypothesis that a vegetation index can be developed from TM imagery to measure hardwood-conifer mixture proportions. This is evidenced by the strong correlation of HCMI3 with stand hardwood proportion ($r = 0.82$). The Gramm-Schmidt process was an effective method for deriving the index, and only a very few data points were required for the computation. The results of this study also show that moderately strong regression models can be developed to quantify hardwood-conifer proportions in the Oregon Coast Range from either the raw or transformed TM data.

The strength of the HCMI and regression models are most likely a function of the clear differences exhibited by conifer and hardwood cover in the infrared wavelengths (e.g., table 3). While a complex assortment of internal and external factors influence the reflectance of forest canopies, the divergence of hardwood and conifer response in the infrared is thought to be caused primarily by differences in leaf structure, leaf water content, and canopy geometry (Guyot *et al.* 1989, Knipling 1970). The importance of the infrared bands was evident in several parts of this research. The near-IR band (TM4) was moderately correlated with hardwood proportion ($r = 0.63$), and the mid-IR bands (TM5 and TM7) were strongly correlated ($r = 0.80$). The infrared bands produced the most significant HCMI coefficients (table 4), and were also important variables in the multiple regression model using untransformed TM data (table 6).

Although HCMI3 produced the strongest correlation with hardwood proportion of any single image variable (table 5), it did not produce the best regression model (table 6). The untransformed data produced the best fit, the tasseled cap and principal component transformations performed slightly worse, and the single HCMI3 variable produced the weakest model. The progressively weaker fit of the models is not surprising considering the potential for information loss as data dimensionality was reduced during successive transformations. However, the error assessments show that practical differences between the models were small (tables 7 and 8). These results indicate that much of the hardwood-conifer mixture information in TM data was captured with the single HCMI.

A detailed assessment of topographic effects was beyond the scope of this study, but inclusion of the COSI variable provided an indication of the sensitivity of the spectral variables to terrain-induced differences in illumination. The COSI variable was only weakly associated with HCMI3, and was not a statistically significant predictor when combined with HCMI3 in a stepwise regression model. This suggests that HCMI3 is less sensitive to topographic effects than the original image variables. However, the significance of these results is limited by the coarse resolution of the available DEM data. The generally small contribution of COSI to the regression analysis may be due in part to the cell size mismatch between DEM and satellite data.

The development of an HCMI is directly applicable to biodiversity modeling in the Pacific Northwest. However, an evaluation of information contained in CSI-HCMI space suggests the results of this research may have broader significance. The CSI-HCMI transformation provides a direct linkage between important physical properties and patterns of variation in multispectral space (figure 4). The distribution of reference stands in CSI-HCMI space suggests that significant information about forest canopy structure and composition is captured and separated by this transformation of TM data. The connection between physical space and the transformed data space is also apparent from the full distribution of closed canopy forest pixels in the CSI-HCMI plot (figure 4). There is an absence of CCF pixels in both the complex canopy, hardwood dominated region and simple canopy, conifer dominated region of CSI-HCMI space. This may be attributed to the smooth, simple structure of hardwood canopies relative to the rougher, more complex, and more variable canopy structures exhibited by conifer forests. Based on these preliminary observations, it appears that the CSI-HCMI transformation may provide a potentially useful compression and enhancement of TM data for applications in dense forest conditions. Further study is required to ascertain the value of the CSI-HCMI transformation for forests in the Pacific Northwest as well as for closed canopy forest systems in other geographic regions.

REFERENCES

Butera, M.K., 1986, A correlation and regression analysis of percent canopy closure versus TMS spectral response for selected forest sites in the San Juan National Forest, Colorado, *IEEE Transactions on Geoscience and Remote Sensing*, **24,** 122-129.

Cetin, H., Warner, T.A., and Levandowski, D.W., 1993, Data classification, visualization, and enhancement using n-dimensional probability density functions (nPDF): AVIRIS, TIMS, TM, and geophysical applications. *Photogrammetric Engineering and Remote Sensing*, **59,** 1755-1764.

Cohen, W.B., and Spies, T.A., 1992, Estimating structural attributes of Douglas-fir/western hemlock forest stands from Landsat and SPOT imagery. *Remote Sensing of Environment*, **34,** 167-178.

Cohen, W.B., Spies, T.A., and Fiorella, M., 1995, Estimating the age and structure of forests in a multi-ownership landscape of western Oregon, U.S.A. *International Journal of Remote Sensing*, **16,** 721-746.

Collins, J.B., and Woodcock, C.E., 1994, Change detection using the Gramm-Schmidt transformation applied to mapping forest mortality. *Remote Sensing of Environment*, **50,** 267-279.

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

Crist, E.P., and Cicone, R.C., 1984, A physically-based transformation of Thematic Mapper data-the TM tasseled cap. *IEEE Transactions on Geoscience and Remote Sensing*, **22,** 256-263.

Crist, E.P., Laurin, R., and Cicone, R.C., 1986, Vegetation and soils information contained in transformed Thematic Mapper data. In *Proceedings, IGARSS '86 Symposium, Zürich, Switzerland, 8-11 September 1986,* SP-254 (Paris: European Space Agency), pp. 1465-1470.

Esbensen, K., and Geladi, P., 1989, Strategy of multivariate image analysis (MIA). *Chemometrics and Intelligent Laboratory Systems*, 7, 67-86.

Franklin, J., 1986. Thematic mapper analysis of coniferous forest structure and composition. *International Journal of Remote Sensing*, **7,** 1287-1301.

Franklin, J.F., and Dyrness, C.T., 1988, *Natural Vegetation of Oregon and Washington* (Corvallis: Oregon State University Press).

Guyot, G., and Riom, D.G.J., 1989. Factors affecting the spectral response of forest canopies: A review. *Geocarto International*, **3,** 3-18.

Horler, D.N.H., and Ahern, F.J., 1986, Forestry information content of Thematic Mapper data. *International Journal of Remote Sensing*, 7, 405-428.

Jackson, R.D., 1983, Spectral indices in *n*-space. *Remote Sensing of Environment*, **13,** 409-421.

Johnson, P.E., Smith, M.O., and Adams, J.B., 1985, Quantitative analysis of planetary reflectance spectra with principal component analysis. *Journal of Geophysical Research*, **90**, C805-C810.

Kauth, R.J. and Thomas, G.S., 1976, The tasseled cap - a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. In *Proceedings on the Symposium on Machine Processing of Remotely Sensed Data*, 4b:41-51, 6 June-2 July 1976 (West Lafayette, Indiana: LARS, Purdue University).

Knipling, W.B., 1970, Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, **1,** 155-159.

Ripple, W.J., Wang, S., Isaacson, D.L., and Paine, D.P., 1991, A preliminary comparison of Landsat Thematic Mapper and SPOT-1 HRV multispectral data for estimating coniferous forest volume. *International Journal of Remote Sensing*, **12,** 1971-1977.

Smith, J.A., Lin, T.L., and Ranson, K.J., 1980, The lambertian assumption and Landsat data. *Photogrammetric Engineering and Remote Sensing*, **9**, 1183-1189.

Spanner, M.A., Brass, J.A., and Peterson, D.L., 1984, Feature selection and the information content of Thematic Mapper simulator data for forest structural assessment. *I.E.E.E. Transactions on Geoscience and Remote Sensing*, **22**, 482.

Weigand, C.L., and Richardson, A.J., 1982, Comparisons among a new soil index and other two- and four-dimensional vegetation indices. In *Technical Papers, ACSM-ASP Convention, Denver, Colorado, March 14-20 1982*. pp. 2