INT. J. REMOTE SENSING, 2001, VOL. 22, NO. 6, 1053–1066

A TM-based hardwood-conifer mixture index for closed canopy forests in the Oregon Coast Range

T. K. MAIERSPERGER

Department of Forest Science, Oregon State University, Corvallis, OR, 97331, USA; e-mail: maier@fsl.orst.edu

W. B. COHEN

USDA, Forest Service, Pacific Northwest Research Station, Corvallis, OR, USA

and L. M. GANIO

Department of Forest Science, Oregon State University, Corvallis, OR, USA

(*Received 15 April 1997; in final form 15 July 1999*)

Abstract. The purpose of this study was to develop and evaluate a multi-spectral vegetation index for quantifying relative amounts of hardwood and conifer cover from Thematic Mapper (TM) imagery. We focused on closed canopy forests in the Oregon Coast Range, where hardwood, conifer, and mixed stand conditions are prevalent. An approach based on the Gramm-Schmidt orthogonalization process was used to derive three different hardwood-conifer mixture indices (HCMIs). Using correlation and regression analyses, the capacity of these indices to predict closed canopy hardwood percentage 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 reflectance. Results show that while similar amounts of information were explained by HCMI, TM band, Tasseled Cap, and principal component models, only predictions derived from the HCMI1 and HCMI2 variables were unbiased with respect to topographic effects.

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 modelling of the presence and abundance of several vertebrate species using remotely-sensed measures of forest structure and composition derived from Thematic Mapper (TM) data. In earlier studies (Cohen and Spies 1992, Cohen *et al.* 1995), regression models based on the response of the TM wetness index have proven useful in predicting several structural attributes of closed canopy conifer forests in western Oregon. In this research, we turned our attention to the problem of quantifying hardwood-conifer mixtures, and focused on closed canopy forests in the Oregon Coast Range where a full compositional gradient is well represented.

ISSN 0143-1161 print/ISSN 1366-5901 online
http://www.tandf.co.uk/journals

Based on preliminary examination of forest response in TM data space, we hypothesized that a hardwood-conifer mixture index could be developed to effectively measure the relative amount of these cover types. By plotting reference data distributions within a series of raw TM and Tasseled Cap spectral space plots (e.g., TM4–TM5 space, brightness-greenness space), two major axes of closed canopy forest response were visualized. The first spectral axis occurred between young and old conifer stands, and we associated this variation with the increasing development of canopy structure with age (as characterized by multiple layers, abundance of gaps, tree size variability, and overall canopy roughness). The second spectral axis appeared to span the space between pure conifer and hardwood stands. In contrast to the large amount of within-conifer variation observed, the pure hardwood samples exhibited a relatively small amount of spectral variability. This observation is supported by previous studies which suggested that the simple structure of dense hardwood canopies do not show significant spectral change with age (Spanner *et al.* 1984, Horler and Ahern 1986).

In light of these preliminary findings, we developed a two-step strategy for separating structural and compositional variation into two independent indices. First, the spectral variation associated with closed canopy forest structure was addressed by defining an initial axis through young conifer and old conifer reference spectra. We term this axis the canopy structure index (CSI). Subsequently, a second axis was defined orthogonal to the first using the spectra of a pure hardwood reference point, producing the coefficients of a hardwood-conifer mixture index (HCMI). Reference points for index formulation were selected using methods of spectral space visualization and analysis (Johnson *et al.* 1985, Esbensen and Geladi 1989), and the Gramm-Schmidt process (Jackson 1983) was used to calculate the coefficients.

The effectiveness of the HCMI as a predictor of stand hardwood percentage was assessed using correlation and regression techniques. Because the study was limited to closed canopy forest conditions, the selection of hardwood percentage as the independent variable provided sufficient information for the estimation of hardwood-conifer compositional mixtures (i.e. conifer percentage is the compliment of hardwood percentage). For comparison, three other groups of spectral variables were analysed: (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 spectra. Recognizing that topographic effects can have a major confounding influence on TM information extraction, we also included a cosine of the solar incidence angle image (Smith *et al.* 1980) in the analysis. This variable was used to determine the effect of topography on the spectral variables and related model predictions.

2. Materials and methods

2.1. *Study area and reference data*

The study area for this project was an 800 000 ha 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. The study area is composed of two major vegetation zones: the *Picea*

Figure 1. Location of the study area in Oregon. The shaded relief enlargement depicts the topographic complexity of the region (illumination from the north-west, twice vertical exaggeration).

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.

The most common conifer species in the study area are *Pseudotsuga menziesii* (Douglas fir) and *Tsuga heterophylla* (western hemlock), with less important species including *Abies amabilis* (Pacific silver fir), *Abies grandis* (grand fir), *Picea sitchensis* (Sitka spruce), and *Pinus contorta* (lodgepole pine). *Alnus rubra* (red alder) is the most prevalent hardwood species, with *Acer macrophyllum* (bigleaf maple) present in less significant amounts. Although conifer species dominate the landscape in this region, hardwoods are abundant 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.

Reference data for the project were selected from a database of 913 photo-interpreted 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.

2.2. *Imagery, DEM data, and preprocessing*

A Landsat Thematic Mapper scene (Landsat 5, Path 46, Row 29) dating from 29 August, 1988 was used in this study. The imagery was acquired with UTM georeferencing, and was obtained with a spatial resolution of 25 m. The standard Tasseled Cap coefficients (Crist *et al.* 1986) were applied to the image to derive the brightness, greenness, and wetness indices. In addition, 30 m resolution DEM data were acquired from the US Geological Survey. The DEM quads were mosaiced, converted to UTM, and clipped to the study area boundary. Slope and aspect were derived from the DEM data, and were used along with the solar elevation and azimuth from the time of TM image acquisition to produce a cosine of the solar incidence angle (COSI) image (Smith *et al.* 1980).

2.3. *Data stratification by closed canopy 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. Although statistical techniques for data stratification were considered (e.g., supervised classification), we selected a data visualization approach (Esbensen and Geladi 1989, Cetin *et al.* 1993) because it provided a simple and efficient method for simultaneously classifying closed canopy forest and screening the reference data for outliers.

An important first step for visualizing data distributions is to choose two variables which produce an appropriate viewing perspective of the data structure for the class of interest. For this, we selected the brightness (BRT) and greenness (GRN) indices 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). Using the feature space analysis tools provided by the image processing software (ERDAS Imagine Version 8.2), we plotted the locations of every reference data pixel on top of the distribution of every image pixel in BRT-GRN space (figure 2). This technique allowed us to quickly interpret the relationship between the reference and image data distributions. As expected, the reference data formed an elliptical cloud along a diagonal axis within BRT-GRN space. 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 crowns). Subsequently, sixty-four stands having such outliers were removed from the analysis, leaving a total of 266 reference stands.

The feature space analysis tools provide a direct linkage between spectral space and scene space (i.e. the spectral space location of any image pixel can be found, and the image location of pixels occupying any position in spectral space can be found). To perform the classification, an elliptical boundary was digitized on the computer screen (using the shape drawing tools provided by the software) to enclose

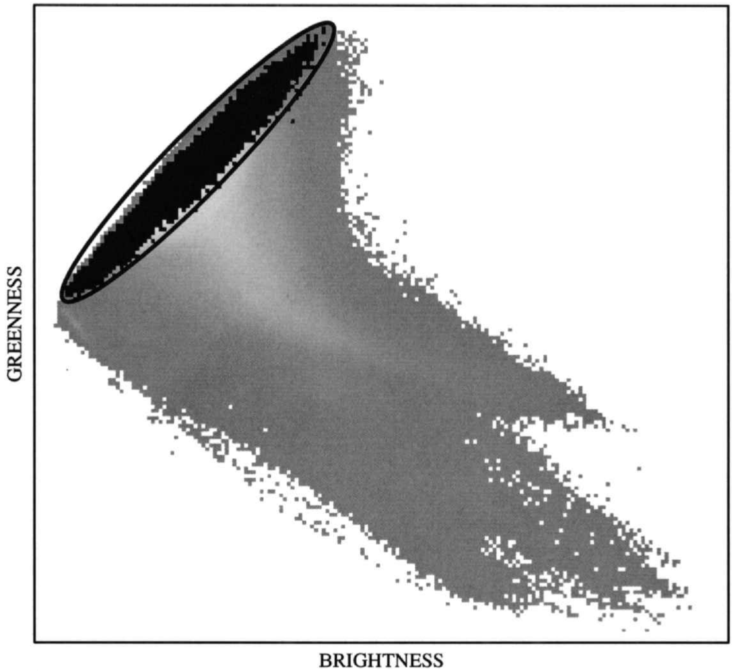

Figure 2. Spectral space classification. The frequency distribution of all the image pixels in BRT-GRN space is shown by the continuum between shades of darker grey (low frequency) and lighter grey (high frequency). The cloud of black points represents the distribution of all the closed canopy forest (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.

the dense cloud of remaining CCF reference pixels in BRT-GRN space (figure 2). The ellipse was used by the software as a decision region to label the corresponding pixels in BRT-GRN scene space. The resulting CCF class was examined relative to the raw bands and Tasseled Cap indices, and it was confirmed that unwanted elements such as clear cuts, urban zones, cloud, and agriculture were adequately eliminated. The CCF class was then used as a mask to eliminate non-forest variation from the six band TM data, the Tasseled Cap indices, and the COSI image. The range of CCF variation remaining in the TM data was large enough to expect a sufficient amount of spectral information relative to noise for further analysis (92, 36, 36, 185, 108, and 57 DN for bands 1–5 and 7 respectively).

2.4. *Principal Component Analysis*

A standardized principal component (PC) analysis (Imagine 8.2) was run on the six band CCF image to capture the major directions of spectral variation in closed canopy forest pixels. The PC analysis provided a purely statistical approach for seeking potentially useful variables for quantifying hardwood percentage from the imagery. As described in the next section, the analysis also furnished variables for data visualization in our process of HCMI development. The eigenvectors of the first three principal components were calculated, and the coefficients were applied to produce a three component CCF image.

2.5. *Development of canopy structure and hardwood-conifer mixture indices*

We used an algebraic formulation of the Gramm-Schmidt process (Jackson 1983) to derive our canopy structure and hardwood-conifer mixture indices. To derive any two orthogonal indices, the calculations simply require the raw TM spectra of three reference endpoints. In this case, we sought to determine the spectral response of young and old conifer points to define the canopy structure index (CSI), while a single hardwood point was needed to establish an orthogonal hardwood-conifer mixture index (HCMI). The selection of reference endpoint spectra may be accomplished by any number of methods depending upon the objectives at hand. For example, reference data could be drawn from several geographically dispersed scenes to provide a bulk transformation with general applicability (Crist and Cicone 1984), or from a single scene as presented here to provide a more finely tuned index. The reference spectra should not only be representative of the scene elements of interest, but should also characterize the reflectance extrema between elements so that maximum discrimination is achieved (Jackson 1983). To meet these objectives, we used an extension of the methods described by Johnson *et al.* (1985) in which the extreme (i.e. endmember) spectra of pure materials were selected for a mixture modelling application with the aid of principal component plots.

The first and second principal components of the six band CCF image (as described in §2.4) 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 the closed canopy 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 1). The mean response of these stands (the crosses labelled 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,

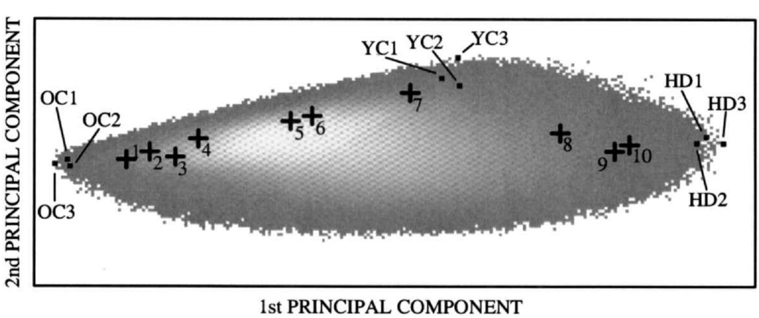

Figure 3. Selection of index reference points in PC1–PC2 space. The frequency distribution of all the CCF pixels in PC1–PC2 space is shown by the continuum between shades of darker grey (low frequency) and lighter grey (high frequency). The crosses labelled 1–10 indicate the mean PC1–PC2 response of the reference stands in table 1. 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 2). The Gramm-Schmidt process was applied to the digital numbers in table 2 to calculate the coefficients for three slightly different canopy structure and hardwood-conifer mixture indices (CSI 1–3 and HCMI 1–3).

Table 1. 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	—

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 pixels that were not contained in the reference plots, but were located at the very extremes of the image data envelope. Using the feature space analysis tools provided by the software, the spatial location of each of the index endpoints was found in the six band TM imagery. The six band response was then recorded for each point (table 2). The Gramm-Schmidt process was applied to the digital numbers in table 2 to produce the coefficients of three different canopy structure indices and the three associated hardwood-conifer mixture indices.

2.6. *Correlation and regression analysis*

A total of 266 reference stands (each with photo-interpreted hardwood percentage) were used as the basis for statistical analysis. The mean response for each stand was extracted from the following image variables: COSI, TM1, TM2, TM3, TM4, TM5,

Table 2. 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

TM7, BRT, GRN, WET, PC1, PC2, PC3, HCMI1, HCMI2, and HCMI3. Scatterplots of hardwood percentage versus the spectral variables revealed consistently linear relationships, so no transformations of the data were performed. Correlation coefficients were then calculated to determine the linear association between the variables of interest.

Regression modelling was executed in two phases. First, simple regressions of hardwood percentage on each spectral variable were carried out. Second, within each group of spectral variable (i.e. original bands, Tasseled Cap indices, principal components), multiple regression models were generated. To avoid problems associated with multicolinearity, initial variable sets were chosen so that no two variables had correlation coefficients greater than 0.8. A forward stepwise approach was then used to generate models in which the criteria for accepting or dropping an explanatory variable was significance at the 0.05 level. To examine the effect of illumination condition on the distribution of prediction errors, the residual (observed minus predicted) values were plotted and correlated against mean COSI response.

3. Results and discussion

3.1. *Transformation coefficients*

The coefficients of the Tasseled Cap, principal component, CSI, and HCMI transformations are shown in table 3. Although the Tasseled Cap coefficients were not derived in this study, they are presented here for comparison with our results. Examination of these coefficients provides insight into the relative importance of each TM band in the various transformations.

The first Tasseled Cap index (BRT), 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 (red) and TM4 (near-IR). The second

Table 3. 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 third principal components accounted for only 5.8 and 0.3% of the spectral variation, and it is difficult to attribute physical meaning to these components by examination of the coefficients.

Using three slightly different sets of reference points (table 2), three canopy structure indices (CSI 1–3) and three hardwood-conifer mixture indices (HCMI 1–3) were derived using the Gramm-Schmidt process. The CSIs, contrived to capture the variation between old and young conifer points, are completely dominated by TM4, with a contribution from TM5 at about one-third of the TM4 weighting. The coefficients reflect, and are indeed nearly proportional to, the large DN differences exhibited by the conifer reference spectra in TM4 and TM5 (table 2). HCMIs 1–3 were formulated to be orthogonal to their corresponding CSI toward a pure hardwood reference point. Not surprisingly, the three HCMIs also have quite similar coefficients. The dominance of TM5 as well as the strength of the TM7 and TM4 loadings are attributable to the important hardwood-conifer differences exhibited by these bands in the reference data (table 2). The HCMIs appear similar to an inverse of the Tasseled Cap wetness feature, although TM5 is more important while the visible bands are nearly insignificant. The substantial dynamic range of the HCMIs (between 79 and 85 DN) indicated the potential for significant information content.

3.2. *Correlation and regression analysis*

The correlation matrix for stand hardwood percentage, COSI, and the spectral variables for the 266 reference plots appears in table 4. All correlations with hardwood percentage were positive except for wetness and the second principal component. HCMI3 ($r=0.83$) and HCMI2 ($r=0.80$) were strongly correlated with hardwood percentage, while TM5 ($r=0.79$), TM7 ($r=0.79$), and HCMI1 ($r=0.77$) showed slightly weaker relationships. Brightness ($r=0.66$) and PC1 ($r=0.65$) were the most significant correlates among the Tasseled Cap and principal component variables.

While positive correlations between the spectral variables and COSI were expected, the relationship between hardwood percentage and COSI (albeit weak), was an unanticipated result. From an ecological standpoint, we would expect the more highly illuminated south and south-eastern slopes to be warmer and drier, thus providing a less amenable environment for hardwood establishment. Intensive

Table 4. Correlation matrix for the study variables ($n=266$).

	Hard	Cosi	TM1	TM2	TM3	TM4	TM5	TM7	BRT	GRN	WET	PC1	PC2	PC3	H1	H2	H3
Hard	1.00	0.27	0.21	0.50	0.51	0.59	0.79	0.79	0.66	0.56	−0.49	0.65	−0.26	0.35	0.77	0.80	0.83
Cosi		1.00	0.33	0.60	0.58	0.60	0.58	0.57	0.61	0.58	0.03	0.61	0.17	0.05	0.34	0.35	0.43
TM1			1.00	0.78	0.78	0.46	0.42	0.43	0.51	0.38	0.23	0.47	0.13	−0.75	0.25	0.19	0.27
TM2				1.00	0.98	0.81	0.80	0.79	0.86	0.75	0.09	0.83	0.18	−0.29	0.50	0.49	0.60
TM3					1.00	0.77	0.80	0.80	0.83	0.69	0.01	0.80	0.08	−0.30	0.56	0.54	0.64
TM4						1.00	0.91	0.84	0.97	0.99	0.17	0.99	0.44	0.17	0.39	0.46	0.60
TM5							1.00	0.99	0.95	0.87	−0.26	0.95	0.02	0.27	0.75	0.79	0.88
TM7								1.00	0.91	0.80	−0.36	0.90	−0.10	0.25	0.82	0.85	0.93
BRT									1.00	0.96	0.04	0.99	0.31	0.14	0.52	0.57	0.70
GRN										1.00	0.22	0.98	0.51	0.24	0.32	0.40	0.54
WET											1.00	0.06	0.92	−0.41	−0.80	−0.79	−0.68
PC1												1.00	0.34	0.19	0.49	0.55	0.68
PC2													1.00	−0.09	−0.65	−0.59	−0.45
PC3														1.00	0.24	0.35	0.33
H1															1.00	0.99	0.97
H2																1.00	0.99
H3																	1.00

harvest activity in the Coast Range has certainly contributed to a broader distribution of hardwood species in drier upland forest sites, but a more rigorous examination of this finding was beyond the scope of this research.

The effect of illumination conditions on TM spectral response is clearly indicated by the moderately strong correlations between COSI and the TM band data. The significance of the topographic brightness gradient is also evident in the primary Tasseled Cap and principal component features (the highly inter-correlated BRT, GRN, and PC1 transforms). Because the major portion of illumination-related spectral variance is captured by these first variables, the secondary or tertiary features (i.e. WET, PC2, and PC3) are left to vary independently of topographic effects. This finding supports earlier studies in which the effects of topography were reduced substantially in the second and third principal components of TM data (Conese *et al.* 1988) and wetness (Cohen and Spies 1992, and Cohen *et al.* 1995). Similarly, the moderately weak association between the HCMIs and COSI likely arises from the establishment of these indices as secondary directions of variation in closed canopy forest conditions. It is also interesting to note that the HCMIs, whose coefficients resemble an inverse of wetness, indeed show substantial negative correlation with this index.

A subregion of the study area characterized by high relief, riparian hardwoods, and upland conifer forest was selected to further illustrate the impact of topography on selected image variables (figure 4). Brightness (figure 4(*a*)) relates nearly as well to COSI ($r=0.61$) as to hardwood percentage ($r=0.66$). The topographic effect is clearly visible as south and south-eastern slopes appear in highly illuminated contrast to surrounding areas. Patterns of hardwood versus conifer cover are strengthened in TM7, while the appearance of topographic relief is reduced but still readily apparent (figure 4(*b*)). There is also prominent separation between hardwood and conifer in each of the three HCMIs (figures 4(*c*, *d*, and *e*)), but topographic variation is noticeably reduced. Wetness (figure 4(*f*)) appears virtually flat, with the negative relationship to hardwood percentage manifested as a pattern of dark riparian corridors running through the lighter conifer matrix.

For comparison with HCMI-based regression models, the best single and multivariate models were selected from among the TM band, Tasseled Cap, and principal component variable sets (table 5). The R-square value indicates the proportion of variation in hardwood percentage explained by each model, while the RMSE provides a measure of average prediction error. The residual correlation with COSI is a useful indicator of prediction bias arising from the topographic effect. These results are generally consistent with the trends seen in the correlation matrix. While certain TM band, Tasseled Cap, and principal component models can match or slightly outperform the HCMI models in terms of R-square and RMSE, only the HCMI1 and HCMI2 model predictions are unbiased with respect to topography. Although TM7 and HCMI2 contained similar information content (i.e. association with hardwood percentage) and were strongly correlated with one another ($r=0.85$), the differing sensitivity to topographic effects is readily observed in figure 5. The negative slope of the TM7 model residuals indicates that estimates of hardwood percentage are biased toward under-prediction in areas of low illumination and over-prediction in brightly illuminated terrain.

3.3. *Practical considerations for using the Gramm-Schmidt process*

The Gramm-Schmidt process can be an effective method for deriving physically significant indices from multi-spectral data, and only a few reference points are

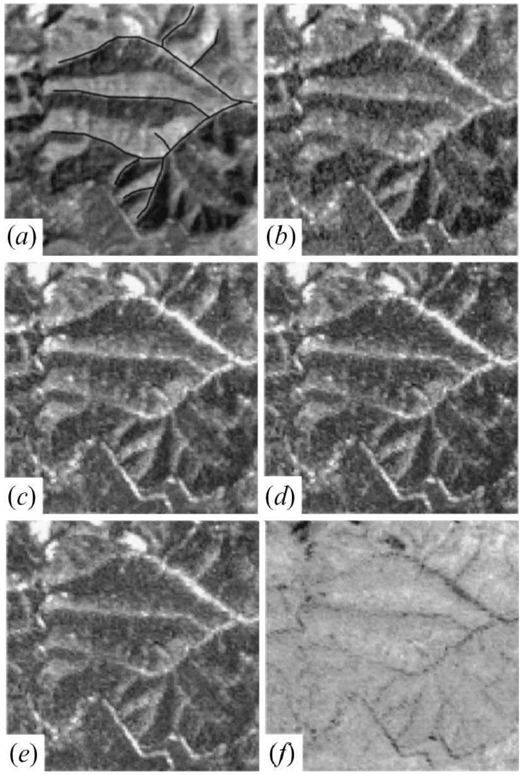

Figure 4. The visual expression of topography in selected image layers for a sub-region of the study area. A two standard deviation linear stretch was applied to each image, and hardwood-dominated riparian zones were delineated from an aerial photograph and shown for reference in (*a*). The topographic effect is most evident in brightness (*a*), and decreases progressively in TM7 (*b*), HCMI3 (*c*), HCMI2 (*d*), HCMI1 (*e*), and wetness (*f*).

Table 5. Regression modelling results ($n=266$).

Regression model	Model R-square	RMSE (%)	Residual correlation with COSI (*p*-value)
− 59.91 + 7.93 (TM7)	0.63	17.96	− 0.29*
54.08 + 10.54 (TM7) − 5.90 (TM2)	0.67	16.94	− 0.23*
− 84.74 + 0.99 (BRT)	0.44	22.14	− 0.18*
− 27.62 + 1.03 (BRT) − 4.01 (WET)	0.71	15.86	− 0.25*
− 49.77 + 0.75 (PC1)	0.43	22.38	− 0.17*
39.11 + 0.92 (PC1) − 2.95 (PC2) + 1.70 (PC3)	0.71	15.98	− 0.25*
− 37.62 + 4.20 (HCMI1)	0.59	19.03	0.02 (0.72)
11.16 + 4.48 (HCMI2)	0.65	17.56	− 0.02 (0.79)
− 11.23 + 3.85 (HCMI3)	0.69	16.46	− 0.16*

*p-value <0.01.

required for the simple calculations. However, the information axis must occupy a significant linear span within the multi-spectral data structure to expect a useful amount of information to be captured . In this research, we relied upon exploratory

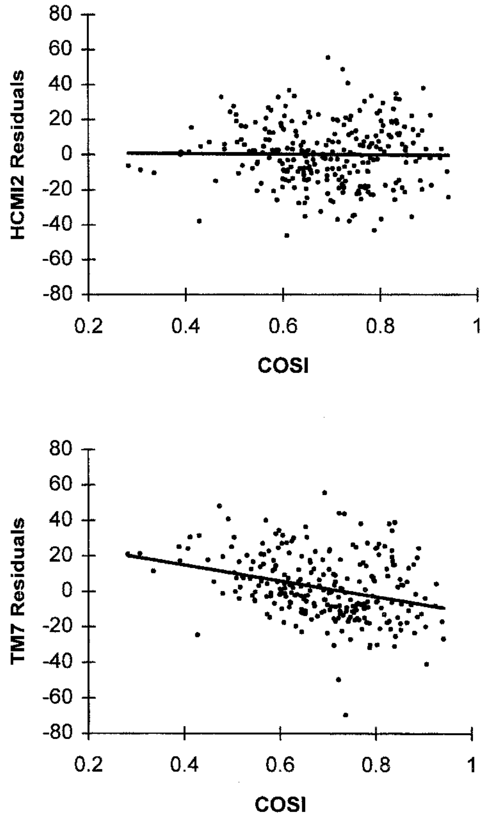

Figure 5. HCMI2 and TM7 model residuals plotted against COSI. HCMI2 predictions are unbiased with respect to topography, while the TM7 predictions are significantly influenced by illumination condition.

data visualization to help us understand patterns of spectral variation, and this led directly to an appropriate strategy for index formulation. Also of critical importance is the selection of suitable index endpoints, since slight differences in these reference spectra can have considerable effect on the meaningfulness of the final index (table 5). Again we turned to graphical representations of multi-spectral space, this time to find extreme regions in the data structure from which to cull index endpoints. Uncertainty about optimal reference point selection was accounted for by choosing three different sets of spectra. This approach is feasible since the computational cost of deriving and testing several similar indices is not prohibitive.

4. Conclusions

The results of this study support our hypothesis that a vegetation index can be developed from TM imagery to measure hardwood-conifer mixing proportions in the Oregon Coast Range. Because the HCMI was derived explicitly within closed canopy forest space, the hardwood-conifer relationships captured by the index cannot be expected to hold in areas where soil or other background materials are prevalent. Nevertheless, the hardwood-conifer trends described in this work should be present in the closed canopy response of similar forested systems, and we would expect that comparable HCMIs could be contrived to capture this information.

Historically, multi-spectral indices have often been used as correlates of vegetation amount, for example density (Kauth and Thomas 1976), leaf area (Weigand and Richardson 1982), and biomass (Franklin 1986). In this research, we have successfully extended the vegetation index approach to the problem of hardwood-conifer mixtures. The use of indices to measure variables such as conifer age (Cohen *et al.* 1995) and tree mortality (Collins and Woodcock 1994) also demonstrates that a broad array of phenomena may be addressed by this strategy. Furthermore, the use of secondary or tertiary axes of spectral variation when appropriate (e.g. HCMI, wetness) can provide information from which the effects of topography are greatly reduced or eliminated. Given the inconsistent results arising from the application of various topographic correction algorithms (Gu and Gillespie 1998), we suggest that the use of appropriate multi-spectral transformations may significantly improve image analysis in areas of high relief.

Acknowledgments

We wish to thank David Wallin and Julia Jones for their enthusiastic support of this work and their constructive reviews of the original manuscript. This research was supported by NASA Grant NAGW-3745, by the H.J. Andrews LTER, and by the Coastal Landscape Analysis and Modelling Study (CLAMS).

References

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, USA. *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.

Conese, C., Marracchi, F., Miglietta, F., Maselli, F., and Sacco, V. M., 1988, Forest classification by principal component analysis of TM data. *International Journal of Remote Sensing*, **9**, 1597–1612.

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. *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).

Gu, D., and Gillespie, A., 1998, Topographic normalization of Landsat TM images of forest based on subpixel sun-canopy-sensor geometry. *Remote Sensing of Environment*, **64**, 166–175.

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. *Proceedings on the Symposium on Machine Processing of Remotely Sensed Data*, 6 June–2 July 1976 (West Lafayette, Indiana: LARS, Purdue University), pp. 41–51.

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. *IEEE Transactions on Geoscience and Remote Sensing*, **22**, 482–489.

Weigand, C. L., and Richardson, A. J., 1982, Comparisons among a new soil index and other two-and four-dimensional vegetation indices. *Technical Papers, ACSM-ASP Convention, Denver, Colorado, March 14–20 1982* (Bethesda, MD: American Congress on Surveying and Mapping), pp. 210–227.