Available online at www.sciencedirect.com

SCIENCE DIRECT®

Remote Sensing of Environment 95 (2005) 517–531

Remote Sensing of Environment

www.elsevier.com/locate/rse

Patterns of covariance between forest stand and canopy structure in the Pacific Northwest

Michael A. Lefsky[a,*], Andrew T. Hudak[b], Warren B. Cohen[c], S.A. Acker[d]

[a]Colorado State University, Department of Forest Sciences, 131 Forestry Building, Fort Collins, CO 80523-1472, USA
[b]USDA Forest Service, Forest Sciences Laboratory, Rocky Mountain Research Station, 1221 South Main Street, Moscow, ID 83843, USA
[c]USDA Forest Service, Forestry Sciences Laboratory, Pacific Northwest Research Station, 3200 SW Jefferson Way, Corvallis, OR 97331, USA
[d]Olympic National Park, 600 East Park Avenue, Port Angeles, WA 98362-6798, USA

Received 2 May 2004; received in revised form 9 January 2005; accepted 26 January 2005

Abstract

In the past decade, lidar (light detection and ranging) has emerged as a powerful tool for remotely sensing forest canopy and stand structure, including the estimation of aboveground biomass and carbon storage. Numerous papers have documented the use of lidar measurements to predict important aspects of forest stand structure, including aboveground biomass. Other papers have documented the ability to transform lidar measurements to approximate common field measures, such as cover, stand height, and vertical distributions of foliage density and light transmittance. However, only a small number of existing works have thoroughly examined relationships between comprehensive assemblages of forest canopy and forest stand structure indices. In this work, canonical correlation analysis of coincident lidar and field datasets in western Oregon and Washington is used to define seven statistically significant pairs of canonical variables, each defining an axis of variation that stand and canopy structure have in common. The first major axis relates mean stand height, and related variables, to aboveground biomass. The second relates canopy cover and volume to leaf area index and stem density. The third relates canopy height variability to mean stem diameter and the basal area of deciduous species. Of the four remaining axes, three are related to contrasts between mature and old-growth stands. Canonical correlation analysis provides a method for ranking the importance of these effects, and for placing both canopy and stand structure indices within the overall covariance structure of the two datasets. In this sense, and for the study area involved, the first three factors (mean height, cover or leaf index area, height variability) represent the same kind of enhancement of lidar data that the tasseled cap indices [Crist, C.P., R.C. Cicone, 1984. A physically-based transformation of thematic mapper data—the TM tasseled cap. IEEE Transactions on Geoscience and Remote Sensing 22, 256–263.] represent for optical remote sensing.

Keywords: Lidar; Laser; Forest; Canopy; Stand; Regional; Canonical correlation analysis

1. Introduction

1.1. Background

It is now recognized that Lidar has an unsurpassed capability for making remote measurements of forest canopy structure, which can then be used to predict forest stand structure. Lidar instruments directly measure the vertical structure of forests by measuring the distance between the sensor and (in this context) a land surface target through the precise measurement of the time elapsed between the emission of a pulse of laser light from the sensor and the detection of the reflection of that light pulse from a target (in this case, a forest). In addition, waveform—recording lidar systems, such as the SLICER (Scanning Lidar Imager of Canopies by Echo Recovery, Blair et al., 1994; Harding et al., 1994, 2001) sensor used in this work and the Laser Vegetation Imaging System (LVIS, Blair & Hofton, 1999), measure the time-resolved quantity

* Corresponding author.
E-mail addresses: lefsky@cnr.colostate.edu (M.A. Lefsky), ahudak@fs.fed.us (A.T. Hudak), warren.cohen@oregonstate.edu. (W.B. Cohen), steve_acker@nps.gov (S.A. Acker).

0034-4257/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.rse.2005.01.004

of laser energy reflected from (in this case) the geometrically complex arrangement of canopy and ground surfaces. The distribution of return energy, the lidar waveform, when reflected from a forest, records the vertical distribution of illuminated vegetation and soil surfaces from the top of the canopy to the ground.

One application for lidar measurements of canopy height and structure information is the prediction of aboveground biomass and carbon storage. Accurate estimates of terrestrial carbon storage are required to determine its role in the global carbon cycle, to estimate the degree that anthropogenic disturbance (i.e., land use/land cover change) is altering that cycle, and to monitor mitigation efforts that rely on carbon sequestration through reforestation. Remote sensing has been a key technology in existing efforts to monitor carbon storage and fluxes (Cohen et al., 1996; Running et al., 1999) and has been identified as an essential tool for monitoring compliance with treaties such as the Kyoto protocol (Ahern et al., 1998).

Numerous papers have documented the use of lidar-measured canopy structure indices as independent variables to predict important aspects of stand structure, including aboveground biomass, basal area, mean DBH, stem density, etc. (Lefsky et al., 1999a, 1999b, 2002; Maclean & Krabill, 1986; Means et al., 1999; Nelson et al., 1984). Other papers have documented the ability to transform lidar measurements to approximate common field measures, such as cover, stand height, and vertical distributions of foliage density and light transmittance (Lefsky et al., 1999a, 1999b, 2002; Means et al., 1999). However, few existing studies have attempted to thoroughly examine the relationships between comprehensive assemblages of forest canopy and stand structure indices. In this work canonical correlation analysis is used to define pairs of canonical variables, each defining an axis of variation common to the canopy and stand structure datasets. In this way, the ranking of various effects can be understood as they relate to the explanation of variance in each dataset, and axes of variation that connect forest canopy and stand structure can be rigorously defined.

Lefsky et al. (1999b) attempted to relate canopy and stand structure in Douglas-fir/western hemlock forests through an analytical framework focused on the delineation of canopy volume classes. This study advances the analysis of canopy and stand structure, as measured by lidar and in the field (respectively) through the use of a rigorous statistical framework, and across a range of coniferous forest types in the Pacific Northwest. In a companion study (Lefsky et al., 2005), the lidar-derived components of each pair of canonical variables are used to predict forest stand structure. The goals of this work are a better understanding of what each variable represents in the context of a suite of variables that define an axis of variation, and an understanding of what each axis of variation represents within the context of numerous axes of variation. In this way, we hope to avoid placing too much importance on any particular variable, or by extension, any method of canopy structure analysis.

1.2. Objectives

The goal of this study is to document the statistical relationships between two multivariate datasets containing coincident lidar measurements of canopy structure and field measurements of stand structure. We expect that with such a quantification of these relationships, these two alternative perspectives on forest structure can be reconciled, and the main effects ranked.

2. Methods

To document the relationships between canopy and stand structure (Fig. 1) SLICER waveforms (A) were transformed into four classes of canopy structure measurement (B). Field measurements of stem diameter and height (C) were then used to create several classes of derived stand structure indices (D). Canonical Correlation Analysis (E) was then used to create pairs of correlated axes from the canopy and stand structure (F). Finally, each pair of axes is interpreted using their correlations with the original stand and canopy indices (G).

2.1. Study regions

Field data were collected in five locations (Fig. 2), selected to sample the maximum practicable range of environment conditions and forest composition in the Pacific-Northwest region of the United States. Table 1 describes the environmental conditions at each location. Within the forested areas of western Washington and

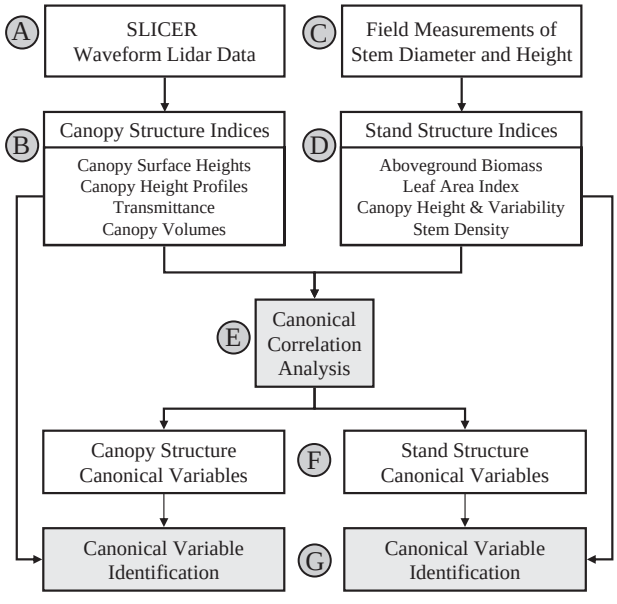


Fig. 1. Flowchart of study analyses.

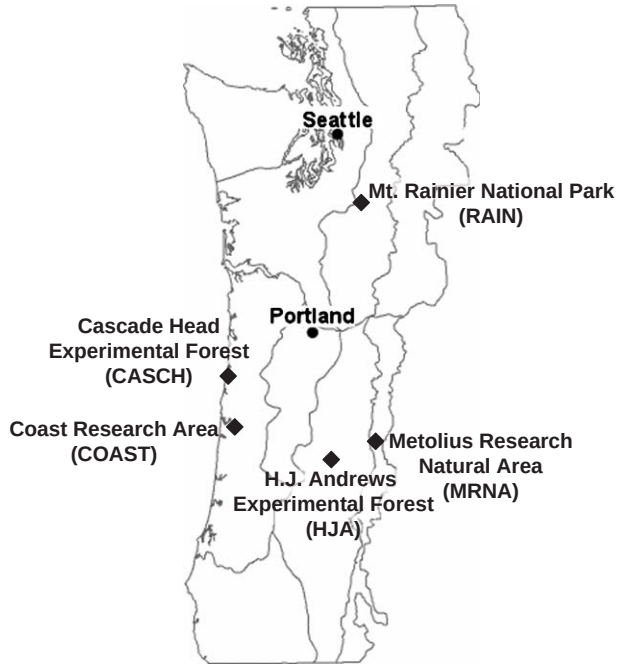


Fig. 2. Map of five study locations in western Washington and Oregon.

Oregon, 12.1% of that area has lower precipitation than observed in the plots used in this study, and 16.5% has higher precipitation, indicating that the range of our sites covered 71.4% of the variation in this variable (Daly et al., 1997). For mean annual temperature, 11.4% of this area has lower temperatures than observed in these same plots, and 10% has higher temperatures, indicating that we sampled 78.6% of the variation in this variable.

Tree composition at these locations (Table 2) reflects climate and edaphic variability, potential vegetation type (PVT), and past and present management practices in Pacific Northwest forests (Franklin & Dyrness, 1988). Cascade Head (CASCH), the most productive site, is dominated by *Picea sitchensi* (Sitka spruce) and *Tsuga heterophylla* (western hemlock). Both the Coast Range (COAST) forest and H.J. Andrews (HJA) sites are predominately *Pseudotsuga menziesii* (Douglas-fir), with significant *T. heterophylla* (western hemlock) at HJA, and abundant *Alnus rubra* (red alder) in the understory of the coastal forest. The plots at Mt. Rainier (RAIN) are all above 1300 m elevation and their composition is largely made up of a variety of "true" firs: *Abies amabilis* (Pacific silver fir), *Abies lasiocarpa* (sub-alpine fir), and *Abies procera* (noble fir) as well a number of other species, including *Chamaecyparis nootkatensis* (Alaskan cedar), *T. heterophylla*, and *T. mertensiana* (mountain hemlock). The Metolius Research Natural Area (MRNA) on the east side of the Cascade Range near Sisters, Oregon, is dominated by *Pinus ponderosa* (Ponderosa Pine), which accounts for 88% of basal area.

At each location, every effort was made to select a series of stands spanning the full range of stand structure (Table 3A). While stands reflecting appropriate maximum heights were found in all five study locations, finding shorter stands in some locations was made difficult by the lack of recent human or natural disturbance. This was especially true at Cascade Head, where mean canopy height was 15.6 m and the shortest plot, an observation which reflects both the low disturbance frequency and the high productivity of this site. The tallest of the shortest plots among the other four sites had a mean canopy height of 6.4. Mean tree heights in the study locations ranged from a maximum of 40 m at Cascade Head to a minimum of 17 m at Metolius, and generally reflect declining productivity (Cascade Head>Coastal, Forest>H.J. Andrews>Mt. Rainier>Metolius, Table 3B).

2.2. Field data collection

Field sampling was carried out in 1996 for H.J. Andrews, 1998 for Metolius, 1999 for Cascade Head and Coast Range, and 2000 for Mt. Rainier. In total, eighty-six 0.25 ha field plots were established under SLICER transects flown in 1995; most plots were associated with a five-by-five array of SLICER footprints. Only forested sites were sampled—sites dominated by herbaceous and shrub vegetation were not. At each plot a 50-by-50 meter sampling area was oriented with the bearing of the SLICER transect, and laid out with dimensions corrected for slope. The intensity of field sampling was a function of stand structure. On old-growth plots all trees greater than 1.37 m tall were measured. On young and mature plots where tree densities were higher, all trees greater than 1.37 m tall were measured on subplots.Tree diameters were initially measured on 3 or 5 subplots.Then the number of additional subplots (5, 9, or 13) needed to sample at least 30 total dominant and codominant trees was estimated and regularly spaced to cover the full extent of the plot. In each subplot, species, diameter at breast height (DBH), and crown ratio (the proportion of the bole with live crown) of all trees greater than breast height (1.37 m) was recorded.

Total aboveground biomass was estimated from DBH and tree height using allometric equations generated from a dataset of tree volumes collected in 18 different protected areas and experimental forests throughout the Pacific Northwest and Colorado (Table 4, Harmon & Franklin, 2002). Site productivity has a significant effect on the allometry relating tree height and DBH, and as a consequence, aboveground biomass and DBH. The Schumacher

Table 1
Mean environmental conditions at the five study locations

	Annual precipitation (mm)	Annual temperature °C	Elevation (m)
Cascade Head, OR	857	11	172
Coast Range, OR	879	11	274
H.J. Andrews, OR	799	9	774
Mt. Rainier, WA	888	5	1443
Metolius, OR	199	7	1015

Table 2
Basal area (m^2 ha^{-1}) of important tree species at the five study locations

	Abies amabilis	*Abies lasiocarpa*	*Abies procera*	*Alnus rubra*	*Chamaecyparis nootkatensis*	*Pinus ponderosa*	*Picea sitchensis*	*Pseudotsuga menziesii*	*Tsuga heterophylla*	*Tsuga mertensiana*
Cascade Head, OR				4.4			40.1	5.2	36.3	
Coast Range, OR				14.5				36.7		
H.J. Andrews, OR								40.6	10.6	
Mt. Rainier, WA	20.5	9.0	12.9		9.0				11.9	14.3
Metolius, OR						24.4				

Species comprising, on average, less than 1 m^2 ha^{-1} were removed from consideration in this analysis.

equation (Schumacher & Hall, 1933) was adopted to reduce the impact of site productivity on estimates of aboveground biomass at each site. The Schumacher equation uses both the height and diameter of trees to predict stem volume, or when wood density is considered, bole biomass. Because trees on high productivity sites are generally taller for a given diameter, they will also have higher volume and biomass than trees of the same diameter on lower productivity sites. Therefore equations based on DBH alone may be biased when applied at sites of varying productivity. Wood and bark densities were taken from the USDA Forest Products Laboratory's Wood Handbook (1999). Following M.E. Harmon (personal communication), an additional 10% was added to the bole biomass to account for branch biomass, which along with bole biomass provides woody biomass.

To use these equations, estimates of height and DBH are required for every tree. Measuring the height of each individual tree was not feasible for all 11,280 sampled in this study— the heights of 1096 trees were measured. If a simple regression approach had been used to generate heights for the unmeasured trees, the benefits of using the Schumacher equation would have been lost, as the relationship between tree height and DBH would be insensitive to productivity. A regression approach that factored site productivity as an explicit variable might have overcome this drawback, but as with any ordinary least squares regression technique, would have further reduced the variability in predicted heights.

An alternative to using regression to predict unmeasured tree heights from DBH measurements was imputation, a method in which missing data are replaced by plausible values, selected from a pool of measured values. One advantage of this approach is that the variability in measured values is preserved in the distribution of predicted values. Another advantage of imputation over regression is that the multivariate relationships of the data are preserved (Moeur & Stage, 1995). Imputation selects a stand-in data value (in this case, tree height) using a similarity function, which relates the imputed variable to other, more frequently measured, variables (in this case, DBH).

Table 3
Lidar and field measurements of canopy and stand structure

A. Mean stand structure variables measured at the five study locations

	Number of plots	Basal area (m^2 ha^{-1})	Deciduous basal area (m^2 ha^{-1})	Leaf area index (m^2 m^{-2})	Density (ha^{-1})	Lorey's height (m)	Aboveground biomass (Mg ha^{-1})	Number of stems > 100 cm (ha^{-1})
Cascade Head, OR	13	86.3	4.4	8.5	1119.2	39.2	667.4	24.6
Coastal Forest, OR	25	53.2	15.5	5.2	460.5	34.8	469.5	13.9
H.J. Andrews, OR	26	53.9	0.6	7.1	1257.0	28.8	445.4	12.0
Mt. Rainier, OR	10	78.7	0.0	9.8	2501.5	27.9	498.2	12.0
Metolius, OR	12	26.4	0.1	2.4	903.2	25.1	149.4	0.8

B. Lidar-measured canopy height indices at the five study locations

	Maximum canopy height (m)	Mean canopy height (m)	Standard deviation of canopy height (m)	Mean number of stems taller than 55 m
Cascade Head, OR	63	40	7	4.1
Coast Range, OR	64	33	7	3.0
H.J. Andrews, OR	64	30	6	2.5
Mt. Rainier, WA	62	27	8	1.3
Metolius, OR	42	17	11	0.0

Table 4
Coefficients, regression statistics, and related statistics for allometric equations of aboveground biomass (ABGM) predicted from height and diameter at breast height using the Schumacher (Schumacher & Hall, 1933) equation ($ABGM = B_0 * DBH^{B_1} * Height^{B_2}$)

Species	*N*	Maximum diameter (cm)	Maximum height (m)	B_0	B_1	B_2	R^2
Abies amabilis	68	80.0	58.9	3.77E-05	1.9884	0.7588	0.99
Abies concolor	56	158.4	74.1	3.80E-05	1.8052	0.9675	0.99
Abies lasiocarpa	15	46.9	30.2	2.01E-05	2.0785	0.7651	0.97
Abies magnifica	31	143.2	89.0	1.09E-04	2.1897	0.2504	0.96
Abies procera	310	235.5	89.8	1.00E-04	1.6688	0.888	0.98
Calocedrus decurrens	25	143.9	39.4	3.69E-07	2.2027	1.6633	0.98
Pinus contorta	30	48.5	30.6	4.11E-05	1.7858	0.9509	0.95
Picea engelmannii	18	66.8	33.5	2.82E-05	1.9132	0.8573	0.98
Pinus jefferyi	21	133.1	52.4	5.12E-07	3.4607	0.0463	0.98
Pinus lambertiana	60	179.6	56.8	2.21E-06	1.8815	1.6172	0.97
Pinus ponderosa	49	117.7	50.0	2.23E-05	2.0914	0.7539	0.98
Picea sitchensis	83	283.0	92.7	5.33E-05	1.6606	0.9845	0.98
Pseudotsuga menziesii	171	206.5	94.0	3.85E-05	1.8698	0.903	0.96
Thuja plicata	35	123.7	55.2	2.16E-05	1.6368	1.1853	0.95
Tsuga heterophylla	272	172.3	78.4	1.10E-04	1.8403	0.6389	0.96
Tsuga mertensiana	404	125.7	49.9	1.91E-05	1.9274	1.0074	0.98

Trees were sampled in the following locations: Arapaho National Forest, Cascade Head Experimental Forest, Deschutes National Forest, Gifford-Pinchot National Forest, H.J. Andrews Experimental Forest, Hood River National Forest, Metolius Research National Area, Mt. Baker-Snoqualmie National Forest, Mt. Hood National Forest, Mt. Rainier National Park, Neskowin Crest Research National Area, Quinault Research National Area, Rogue River National Forest, Sequoia National Park, Siuslaw National Forest, Torrey-Carlton Research National Area, Umpqua National Forest, and Willamette National Forest.

A database of over 300,000 trees was created from the Current Vegetation Survey and Forest Inventory Analysis dataset, all with measured DBH, height and approximate UTM coordinates, and was used to select the most similar neighbor for trees without measured heights. For every imputed tree height, the algorithm first limited the search to only those of the same species. A similarity function was then used to select a subset of trees from the reference database:

$$mnf_{(i,r)} = \frac{kd_{(i,r)}}{\bar{d}_{(i,r)}} \mid \mathrm{DBH_i} - \mathrm{DBH_r} \parallel \mathrm{CC_i} - \mathrm{CC_r} \mid \qquad (1)$$

where *mnf* is the distance function to be minimized, i is the target tree being imputed, *r* is the list of trees with known heights, $d_{(i,r)}$ is the geographic distance between the target tree (i) and each of the reference trees (*r*), *k* is a scaling constant (in this case 40, which was empirically derived), $|\mathrm{DBH_i} - \mathrm{DBH_r}|$ is the absolute difference between the target and reference trees' DBH, and $|\mathrm{CC_i} - \mathrm{CC_r}|$ is the absolute difference between the canopy class of the target and reference trees. Canopy classes are defined as emergent (coded as 1), dominant (2), sub-dominant (3), intermediate (4), and suppressed (5). The tree with the lowest *mnf* is picked as the most similar tree. If, as is often the case, multiple trees have values within 5% of the minimum value, then a tree is picked at random from among these trees.

This method was tested by splitting the database of source trees into a model dataset (75% of the data) and a target dataset (25% of the data), all with known heights. Heights for the target dataset were then imputed and compared to the measured height (Table 5). Of the 31 species that had more than 100 trees in the target dataset, reduced major axis regressions (Curran & Hay, 1986) between imputed and observed heights resulted in equations which explained 73% of variance on average (range was 52% to 87%), with a mean slope of 1.0 (standard deviation was 0.02), and a mean intercept of −0.12 m (standard deviation was 0.28 m). Bootstrap analysis was applied to each equation to test if slope and intercepts differed significantly ($\alpha = 0.05$) from one and zero, respectively. In four cases, slopes were found to differ from one, with a mean difference of 1.5% from the ideal value. In three cases, intercepts were found to differ significantly, with values between −0.48 and 0.33 m. These results, although statistically significant in all seven cases, were not considered to effect the validity of the overall analysis, because the slopes were not biased by more than 2%, the intercepts were not biased by more than 50 cm, and the species in question tended to account for only a minor fraction of aboveground biomass on the plots.

Plot-level estimates of leaf area index were calculated from all-sided leaf area estimates for individual trees. Tree leaf area was estimated using allometric equations of leaf area on sapwood area or diameter at breast height, depending on species. Species–specific equations were used, expect for rarer species, where equations for species of similar form were used. LAI regressions based on sapwood area were based on estimates of sapwood area developed using regressions between new field measurements of DBH for all trees and optical measurements of sapwood area for a subset of trees. *R*-square values for regression of sapwood area on DBH for important species were 0.88 for *Tsuga heterophylla*, 0.92 for *Pseudotsuga menziesii*, 0.91 for *Picea sitchensis*, 0.99 for *Abies grandis*, and 0.95 for *Pinus ponderosa*. Leaf area to sapwood area ratios for each species

Table 5
Test of imputation for generating tree heights: reduced major axis regression of observed vs. imputed heights for 69,717 trees

Species	Number of trees	r2	Bias (m)	Variance ratio	Slope (m)	Intercept (b)	Test of Slope ≠1	Intercept ≠0
Abies amabilis	3784	0.83	0.08	1.00	1.00	−0.14		
Abies concolor	2121	0.82	0.16	1.00	0.99	−0.02		
Abies grandis	1268	0.84	0.24	1.01	0.98	0.15	True	
Abies lasiocarpa	452	0.71	0.57	1.01	0.98	−0.27		
Abies magnifica	473	0.87	0.05	1.01	0.99	0.22		
Abies procera	794	0.85	0.33	1.00	1.01	−0.48		True
Acer macrophyllum	1237	0.53	0.44	1.00	1.00	−0.52		
Alnus rubra	2907	0.64	0.28	1.01	0.98	0.07		
Arbutus menziesii	1286	0.55	−0.25	0.98	1.03	−0.17		
Castanopsis chrysophylla	618	0.74	−0.03	1.00	0.99	0.07		
Calocedrus decurrens	1238	0.87	−0.05	0.99	1.02	−0.16		
Chamaecyparis lawsoniana	249	0.76	−0.01	0.99	1.02	−0.31		
Chamaecyparis nootkatensis	214	0.71	0.53	1.01	0.98	−0.20		
Juniperus occidentalis	247	0.52	−0.06	0.97	1.07	−0.45		
Larix occidentalis	210	0.75	0.88	1.00	1.00	−0.91		
Lithocarpus densiflorus	929	0.7	−0.16	0.99	1.01	0.04		
Pinus contorta	2444	0.74	0.04	1.00	1.00	−0.08		
Picea engelmannii	244	0.73	−0.19	1.00	1.01	0.08		
Pinus lambertiana	539	0.85	0.21	1.00	1.00	−0.30		
Pinus monticola	433	0.77	−0.10	0.99	1.02	−0.29		
Pinus ponderosa	4325	0.82	0.04	1.01	0.99	0.17	True	True
Picea sitchensis	362	0.68	0.41	1.01	0.99	−0.05		
Populus trichocarpa	162	0.76	0.24	0.99	1.01	−0.52		
Prunus emarginata	106	0.56	0.04	0.99	1.03	−0.35		
Pseudotsuga menziesii	25,953	0.85	0.39	1.01	0.98	0.08	True	
Quercus chrysolepis	226	0.56	0.17	0.99	1.01	−0.24		
Quercus garryana	373	0.87	−0.20	1.01	0.99	0.33		True
Quercus kelloggii	287	0.54	0.11	1.02	0.97	0.19		
Thuja plicata	3138	0.78	0.27	1.01	0.98	0.19	True	
Tsuga heterophylla	10,704	0.81	0.18	1.00	1.00	−0.12		
Tsuga mertensiana	1983	0.75	−0.08	1.00	1.00	0.14		
Mean	2236	0.73	0.15	1.00	1.00	−0.12		
Standard Deviation		0.11	0.26	0.01	0.02	0.28		

(Waring, 1982) served in the calculation of tree leaf area. Leaf area of *Juniperus occidentalis* was estimated using an equation from Gholz et al. (1979). For *Thuja plicata*, *Acer macrophyllum*, *Alnus rubra*, and minor deciduous species, leaf area was estimated with allometric equations of total leaf biomass on bole diameter at breast height (Gholz et al., 1979, Koerper's equation, reported in Means et al., 1994). Biomass was then multiplied by a specific leaf area coefficient (Burton et al., 1991; Gholz et al., 1976) to obtain tree leaf area. Additional corrections were made to tree leaf area values. For needle-leafed trees, estimates of leaf surface area were corrected for bias from planar area resulting from the three-dimensional form of the needles (Gholz et al., 1976). For deciduous species, petiole mass was subtracted from leaf biomass estimates (Gholz et al., 1976). Summary statistics for field-measured attributes are given in Table 3A.

2.3. SLICER data collections

Lidar waveforms were collected by the SLICER (Scanning Lidar Imager of Canopies by Echo Recovery) instrument in September 1995. SLICER is a modified scanning version of a profiling laser altimeter developed at Goddard Space Flight Center (Blair et al., 1994). The SLICER system digitizes the entire height-varying return laser power signal, or waveform, from the upper-most canopy surface to the ground. The waveform records the vertical distribution of light reflected from multiple canopy elements (foliage and woody structure) over a circular footprint (5–25 m diameter) at the wavelength (1064 nm) of the transmitted pulse. The lidar waveforms used in this work had a nominal footprint diameter of 10 m and were collected in a 50 m swath of 5 contiguous footprints. Georeferencing of lidar footprints is accomplished by combining laser ranging data with aircraft position, obtained via kinematic GPS methods, and laser pointing, obtained with a laser-ring gyro Inertial Navigation System mounted on the SLICER instrument (Blair et al., 1994). Georeferencing of the SLICER data was done at Goddard Space Flight Center using software developed by J. Bryan Blair (personal communication). For the measurements in this study, the vertical resolution of the SLICER waveforms was set at 11 cm, which when combined with the 600 sample-wide waveform, limited the waveform to a maximum height of 66 m. Due to this and additional constraints in the waveform processing software, all waveforms that would have been greater than 63 m were truncated

to 63 m. Comparison of the lidar and field data suggested that the truncation problem affected about 3% of the waveforms used in these analyses. Ground returns on some footprints in old-growth plots with trees greater than 63m tall had to be set by hand due to loss of the ground return as a consequence of the truncation error.Ground return positions were set based on the characteristics of adjacent footprints and independent estimates of topography (Means et al., 1999).

2.4. SLICER data analysis

Four approaches were employed for the description of canopy structure, each implemented using data from the SLICER instrument. The most basic method of canopy description, canopy surface height measurements, only uses the instrument's height measuring capability. A second set of measurements was made by transforming the raw waveform data into an estimate of the vertical distribution of the canopy— the canopy height profile (CHP). A third set of measurements described the transmittance of light in the canopy. A fourth was derived from a system for the measurement of canopy structure, the canopy volume method (CVM), which summarizes the total volume and spatial organization of filled and empty space within the canopy. Details of these methods can be found in Lefsky et al. (1999b).

On 65 of the plots (76% of all plots), 25 waveforms (collected as a five-by-five array) were used to generate forest canopy statistics. Ten plots with less than 25 waveforms were located in the Metolius location, where the relatively open structure of the ponderosa pine stands meant that a 50×50 m plot would have encompassed heterogeneous stand conditions, and so between 5 and 23 waveforms (12 on average) defined a plot. Three plots in the Mt. Rainier location had fewer than 25 waveforms, while an additional three had more that 25 waveforms. These plots are all found where varying aircraft speed led to the distance between waveforms being stretched or compressed in the direction of flight, changing the number of waveform footprints that fit within the standard 50×50 m sampling plots. For one plot in the Coast Range location and four plots in the H.J. Andrews location, fewer than 25 waveforms were used in order to sample conditions that were less than 50×50 m in size, or due to individual waveforms being unusable.

2.4.1. Canopy surface height

These are the simplest class of measurements, which use only the height measurement capability of the sensor. The height of each waveform in the five-by-five array of waveforms associated with each field plot was measured as the vertical distance between the elevation of the first return energy and the average elevation of the ground return. The elevation of the first return energy is the point at which the power of the reflected light exceeds a threshold value; passing this threshold triggers the sensor's waveform recording process. The position of the mean ground return is calculated as the height at which the peak of the ground return is found using IMH (Interactive MacArthur–Horn) waveform processing software (Harding et al., 2001).

Nine indices of canopy surface height were calculated directly from the heights in the five-by-five (25) array of waveforms (See Appendix A for a full list of canopy and stand structure index abbreviations), including maximum height (the maximum of the 25 waveform heights), mean and median height (mean and median of the 25 waveform heights), the standard deviation of canopy surface heights and the number of canopy heights whose height exceeded 55 m.

2.4.2. Canopy height profile

The second approach to canopy structure description was based on the CHP which is a modification of the foliage height profile or FHP (MacArthur & Horn, 1969). The FHP quantifies the distribution of foliage surface area as a function of height. Because SLICER cannot distinguish woody surface area from foliage surface area, the CHP is defined as the distribution of both foliar and woody surface area as a function of height. The CHP can be calculated as relative (with the total vector scaled to sum to one) or absolute (with the total vector scaled to sum to the total leaf or plant area index of the canopy). In this work relative canopy height profiles were used exclusively. A review of these methods and a validation of the SLICER estimates of the CHP are presented in Harding et al. (2001).

For this study, two basic measurements of the average height of the CHPs were calculated, mean canopy height and quadratic mean canopy height (Lefsky et al., 1999b). Mean canopy height was calculated as the mean of the canopy height profile weighted by the height of each element. Quadratic mean canopy height was calculated as the mean of the canopy height profile weighted by the squared height of each element, and has been shown to be more valuable in the prediction of stand characteristics in an eastern deciduous forest (Lefsky et al., 1999a). These two variables differ from those calculated using the canopy surface height method, because they reflect the average height of all canopy surfaces, foliar and woody, not just the total height of the canopy.

Aerial cover in each field plot was calculated as:

$$\text{Cover} = 1 - \frac{K^{*}\text{GroundReturn}}{\text{CanopyReturn} + K^{*}\text{GroundReturn}} \qquad (2)$$

where the ground and canopy returns are the total power reflected from the ground and canopy, respectively. The ground return power of the waveform was multiplied by K to account for differences in the albedo of ground and foliage (about a two-fold difference) so K was set to 2.0. Values of these indices for the 25 canopy height profiles associated with each plot were averaged to obtain each plot-

level estimate. A full list of variables associated with this method is given in Appendix A.

2.4.3. Canopy transmittance

The calculation of transmittance from the SLICER waveforms is described in Parker et al. (2001) and is similar to the calculation of canopy height profiles (e.g. Harding et al., 2001 and Lefsky et al., 1999a), but it omits the adjustment for the shielding of far surfaces by near ones (the MacArthur–Horn transformation). First, background noise was removed from the waveforms (Harding et al., 2001). Second, the ground returns were delineated and removed from the waveform. Next, the power of the canopy return was accumulated downward from the top of the canopy, and normalized by the total power in the waveform (canopy plus ground).Such normalized cumulative power distributions (NCPDs) are equivalent to the closure distribution of Harding et al. (2001) and can be averaged, using the ground as the reference elevation. In averaging these distributions, the cumulative power above the topmost canopy height was set to zero. Transmittance was then estimated from the averaged NCPD as follows:

$$T_{\mathrm{SLICER}}(\mathrm{h}) = 1 - \mathrm{NCPD}(\mathrm{h}+1), \tag{3}$$

where $T_{SLICER}(h)$ is the SLICER estimate of transmittance at height h and NCPD(h+1) is the normalized cumulative power distribution at h+1. The estimate of transmittance profiles from reflected energy does not explicitly account for canopy absorption of laser light; Parker et al. (2001) demonstrates why absorption should be small in the portion of the near-infrared (1064 nm) used by SLICER.

We defined several aspects of a transmittance profile with potential functional significance.The height at which transmittance was 0.98 (h98) was considered indicative of canopy's "radiation-effective height." Slopes of the transmittance profiles were calculated by calculating the bin-to-bin difference in mean transmittance at a one meter resolution, which was then smoothed using a five-unit boxcar window, in order to summarize the local average slope. The height at which transmittance was 0.50 and the transmittance weighted mean height were considered measures of canopy light penetration.

2.4.4. Canopy volume

We used the canopy volume method (CVM) to describe the three-dimensional geometry of forest canopies (Lefsky et al., 1999b). This method is explicitly volumetric as it uses a five-by-five grid of contiguous lidar waveforms to characterize the forest canopy as a three dimensional array. The elements of the array are 10 m in diameter and 1 m tall; corresponding to a 1 m vertical bin within a single waveform. First, each element of the waveform was classified into either "filled" or "empty" volume, depending on the presence or absence of returned energy in the waveform. A second step classified the filled elements of the array into a "euphotic" zone, containing all filled elements of the profile that are within the uppermost 65% of total energy returned from the canopy, and an "oligophotic" zone, consisting of the balance of the filled elements of the profile.

The first two classifications (filled vs. empty, euphotic vs. oligophotic) are then combined to form three canopy structure classes: empty volume within the canopy (i.e., closed gap space), filled volume within the euphotic zone, and filled volume within the oligophotic zone. These classes were then computed for each of the SLICER waveforms associated with a plot. The waveforms were then compared, and a fourth class is added: "open" gap volume, defined as the empty space between the top of each of the waveforms and the maximum height in the array. At this point, the total volume of each of the four canopy classes can be tabulated for the five-by-five array of waveforms associated with each plot. Filled canopy volume is equal to the total volume of euphotic and oligophotic zones and represents the total volume of "filled" canopy. Finally, the average number of each of the four canopy structure classes (open and closed gaps, oligophotic and euphotic zones) occurring at each height was calculated for each plot, to measure the degree of the classes' vertical interspersion. A more detailed description of this method can be found in Lefsky et al. (1999b), and a full list of indices associated with this method can be found in Appendix A.

2.5. Canonical correlation analysis

Ordinary least square (OLS) regression methods have both simple (single X) and multiple (several X) forms (Steel & Torrie, 1980). The use of OLS regression in its single Y on multiple X form is familiar to most remote sensing analysts conducting regression modeling. Although much less familiar, there are also multiple regression methods for relating datasets with multiple X and Y variables (Brown, 1979). One form, Canonical Correlation Analysis (CCA, SAS Institute, 1990), is a generalized form of multiple regression that permits the examination of interrelationships between two sets of variables (multiple X's and multiple Y's) (Tabachnick & Fidell, 1989); its applicability in remote sensing is demonstrated and described in detail by Cohen et al. (2003). CCA maximizes the correlation between a composite of variables from one set with a composite of variables from another set. The advantage of CCA is that it quantifies the redundancy in each set of variables. This, in turn, allows us to group both X and Y variables in terms of their relationships to other variables within their own dataset and to variables in the other dataset.

3. Results

Due to the complexity of this multi-layered analysis, initial interpretation of the results (e.g., the axes defined by

Table 6
Canonical correlation analysis: canonical variable summary

Canonical correlation pair	Canonical correlation	Approximate standard error	Squared canonical correlation	Eigenvalue	Percent of variance (%)	P>F
1	0.99	0.00	0.97	37.6528	61	<.0001
2	0.95	0.01	0.90	8.9713	15	<.0001
3	0.91	0.02	0.83	4.7546	8	<.0001
4	0.89	0.02	0.80	3.9206	6	<.0001
5	0.83	0.03	0.70	2.2863	4	<.0001
6	0.82	0.04	0.67	1.9934	3	0.0008
7	0.79	0.04	0.63	1.6857	3	0.0175

Multivariate statistics and F approximations

Statistic	Value	*F* value	Num *df*	Den *df*	P>F
Wilks' Lambda	0.00	2.84	486	688.77	<.0001
Pillai's Trace	8.57	1.95	486	1044	<.0001
Hotelling–Lawley Trace	66.54	5.37	486	320.49	<.0001
Roy's Greatest Root	37.65	80.88	27	58	<.0001

the canonical correlation analysis) will be presented along with the results themselves. Higher-level analysis of the results (e.g. the ecological significance of canonical variable pairs) will be left for the Discussion section.

3.1. Canonical correlation analysis

There were seven statistically significant pairs of canonical variables from the dataset of lidar canopy structure estimates and the corresponding dataset of forest stand structure (Table 6); the interpretations of these variables are summarized in Table 7. Canonical correlation coefficients (the correlation between the pairs of canonical variables for the two datasets) ranged from 0.99 to 0.79 (between 98% and 63% of variance in common). For the seven canonical variables discussed, a test of the hypotheses that these and all remaining canonical correlations were equal to zero was rejected ($P<0.0001$). Four multivariate tests and F test approximations all rejected the null hypothesis that the canonical correlations were zero ($P<0.0001$).

3.1.1. Canonical variables

Since CCA produces pairs of correlated canonical variables (e.g. one for each dataset), each pair will be discussed in turn.

3.1.1.1. Canonical variable pair 1

Canopy structure. Correlations between lidar indices of canopy structure and lidar canonical variables identified which indices were most closely related to each canonical variable (Table 8). The first lidar canonical variable (LI-1) accounted for 61% of total variance explained (Table 6) and was highly correlated with most of the measures derived from the canopy height profile, including mean, median and maximum waveform height (CHP_H_X, CHP_H_M, CHP_H_MAX) and their squared values (CHP_H_X^2, CHP_H_M^2, CHP_H_MAX^2). It is also highly correlated with height of the 98th percentile of the canopy light transmittance in the canopy and the volume of closed canopy space.

Stand structure. As expected from earlier analyses (Lefsky et al., 1999b), the first canonical variable from the stand structure dataset was highly correlated with stand height variables (Table 9), including maximum field measured height (HTMAXM), Lorey's height (LOREY), and the height of dominant and co-dominant stems (HTDCD). Also highly correlated were aboveground biomass (BIOMASS) and the standard deviation of diameter at breast height (DBHSTD), which is known to increase with stand height and total aboveground biomass in many Pacific Northwest forests (Lefsky et al., 1999b). Six of the 12 other

Table 7
Summary of canonical variable pairs

Canonical variable	Description of ecological significance	Variable with highest correlation
1	Total stand height, and related variables, such as aboveground biomass	CHP_H_M2
2	Cover, euphotic and total canopy volume, leaf area index	COVER_X
3	Canopy variability, deciduous basal area	CHP_H_MIN
4	Canopy vertical distribution, separates young and mature stands	FILLED
5	Canopy variability, increased minimum height, coniferous/ deciduous balance	CHP_H_SD
6	Cover, mean DBH of all stems, stand density; separates mature and old-growth	CHP_Q_SD
7	Cover, oligophotic canopy volume, correlates with mature stands	HGT55

Table 8
Correlations between lidar indices of canopy structure and their canonical variables

Canopy structure indices		LI-1	LI-2	LI-3	LI-4	LI-5	LI-6	LI-7	Tally
Canopy surface height indices									
1	CHP_H_X	**0.93**	0.18	−0.07	0.22	0.07	−0.04	0.01	1
2	CHP_H_X2	**0.94**	−0.01	0.00	0.23	0.13	−0.03	0.07	1
3	CHP_H_M	**0.94**	0.18	−0.03	0.21	0.05	−0.03	0.00	1
4	CHP_H_M2	**0.94**	−0.02	0.06	0.22	0.10	−0.02	0.06	1
5	CHP_H_SD	0.57	0.21	**0.31**	−0.34	**−0.27**	0.00	−0.01	2
6	CHP_H_MAX	**0.93**	**0.23**	−0.05	0.08	0.01	−0.04	0.03	2
7	CHP_H_MAX2	**0.93**	0.10	0.03	0.11	0.05	−0.06	0.11	1
8	CHP_H_MIN	0.66	0.05	**−0.40**	0.33	**0.20**	−0.02	0.05	2
9	HGT55	0.71	**−0.27**	0.11	0.16	0.14	**0.13**	**0.30**	3
Canopy height profile indices									
10	COVER_X	0.44	**0.48**	**−0.24**	−0.09	−0.08	**0.22**	**−0.20**	4
11	CHP_MN_X	0.85	0.08	−0.21	**0.37**	0.09	−0.03	0.00	1
12	CHP_MN_SD	0.79	0.12	0.11	0.08	−0.09	−0.09	**0.14**	1
13	CHP_Q_X	0.84	0.09	−0.18	**0.39**	0.12	−0.06	0.03	1
14	CHP_Q_X2	0.86	−0.09	−0.12	0.33	**0.15**	0.08	0.05	1
15	CHP_Q_SD	0.80	0.16	**0.24**	0.16	−0.05	**−0.23**	0.02	2
16	MNH_COV	0.85	0.10	**−0.27**	**0.37**	0.14	−0.06	−0.05	2
17	QMCH_COV	0.84	0.10	−0.23	**0.39**	**0.17**	−0.08	−0.02	2
Canopy transmittance indices									
18	TRANS_MN_X	0.86	**0.30**	0.07	0.23	0.08	**−0.13**	**0.14**	3
19	TRANS_MN_SD	0.85	0.10	0.17	−0.14	−0.13	−0.01	**0.16**	1
20	TRANS_P50_X	0.89	0.09	**−0.27**	0.19	**0.15**	0.02	**−0.12**	3
21	TRANS_P50_SD	0.89	0.14	0.18	0.00	0.00	0.10	0.04	0
22	TRANS_P98_X	**0.94**	0.19	−0.08	0.20	0.08	−0.02	0.01	1
23	TRANS_P98_SD	0.60	0.21	**0.32**	**−0.38**	**−0.25**	−0.03	−0.03	3
Canopy volume indices									
24	OPEN	0.33	**0.22**	0.04	**−0.36**	**−0.16**	−0.02	0.07	3
25	CLOSED	**0.92**	−0.04	0.05	−0.09	−0.03	**0.12**	0.03	2
26	EUPHOTIC	0.61	**0.35**	0.06	**0.42**	0.09	−0.10	**0.17**	3
27	OLIGO	0.74	0.21	**−0.35**	0.34	**0.17**	**−0.21**	**−0.19**	4
28	FILLED	0.74	**0.32**	−0.15	**0.43**	0.14	**−0.17**	0.00	3
29	LCOMP	0.72	0.18	0.10	0.18	0.05	**−0.17**	−0.03	1

Bold numbers indicate the top 25th percentile of correlations of canopy structure indices with each canonical variable. Ties for eighth place were ignored. Tally indicates the number of canonical variables for which each variable was important. See Appendix A for definition of canopy structure indices.

Table 9
Correlations between stand structure indices and their canonical variables

Stand structure indices	SS-1	SS-2	SS-3	SS-4	SS-5	SS-6	SS-7
LAI	0.43	**0.53**	−0.05	**0.22**	**0.41**	**−0.25**	0.01
BASAL	0.79	0.25	−0.08	0.09	**0.35**	**−0.29**	−0.01
DENSITY	−0.27	0.34	0.07	−0.10	**0.45**	−0.18	0.03
LNDENSITY	−0.09	**0.53**	−0.07	−0.13	0.03	−0.03	**−0.14**
NT100CM	0.83	−0.09	0.26	0.19	0.05	−0.20	**0.19**
DBHMAX	0.88	0.19	0.07	−0.13	0.00	**−0.28**	−0.01
DBHX	0.67	−0.04	**−0.30**	0.17	0.04	**0.37**	**0.13**
DBHU	0.89	0.16	0.22	−0.16	−0.09	−0.05	−0.01
DBHSTD	0.89	0.11	**0.28**	0.00	−0.06	−0.12	−0.05
DECID_BA	−0.02	−0.02	**−0.78**	0.19	−0.32	−0.19	0.00
CONIF_BA	0.75	0.25	0.19	0.03	**0.44**	−0.20	−0.01
BSC	**0.91**	0.08	0.02	**0.22**	0.18	−0.19	0.04
HTMAX	0.89	0.34	0.04	0.07	−0.10	−0.08	0.00
HTMAXM	**0.94**	0.20	−0.02	0.00	−0.06	−0.03	0.08
HTDCD	**0.92**	0.19	0.10	0.03	−0.05	0.04	−0.08
LOREY	**0.96**	0.16	0.04	0.09	−0.10	−0.01	−0.04
COVER	0.40	**0.49**	**−0.32**	**0.35**	−0.01	−0.17	0.02
Tally	4	3	4	3	4	4	3

Bold numbers indicate the top 25th percentile of correlations of canopy structure indices with each canonical variable. Ties for fourth place were ignored. See Appendix A for definition of stand structure indices.

stand structure variables have correlations with the first stand structure canonical variable that exceed 0.75, indicating the high covariance of stand height with other stand structure variables.

3.1.1.2. Canonical variable pair 2

Canopy structure. The second lidar canonical variable accounted for 15% of total variance explained (Table 6) and is related to increases in cover (COVER_X, Table 8) and indices of canopy structure variability. As cover increases, more foliage is likely to be in shadow and dependent on diffuse light, therefore the increase in cover is directly related to the increase in euphotic (dimly lit) space (EUPHOTIC) and total filled space (FILLED). It is also related to increasing variability of canopy structure, as indicated by the positive correlations between this canonical variable, maximum canopy height (CHP_H_MAX), the volume of open space above the local canopy (OPEN) and the decrease in the number of waveforms taller than 55 m (HGT55). This last variable peaks during the mature phase of development in these forests, and declines as the heterogeneity of the forest canopy decreases.

Stand structure. Consistent with this analysis of the second lidar canonical variable, the second stand structure canonical variable had high correlations (Table 9) with field measured cover (COVER), leaf area index (LAI), tree density (DENSITY) and its natural log (LNDENSITY). Canonical variables are constrained so that they are not correlated with any prior canonical variables, so all subsequent correlations should be interpreted in that context. Consequently, the second pair of canonical variables indicated greater or less than average cover than would be expected given the cover values for the first pair of canonical variables. Those plots that had higher cover than average also had greater density and LAI.

3.1.1.3. Canonical variable pair 3

Canopy structure. The third lidar canonical variable accounted for 8% of the variance explained (Table 6), and was directly related to canopy variability. It was positively correlated (Table 8) with the standard deviation of canopy height (CHP_H_SD), the standard deviation of quadratic mean canopy height (CHP_Q_SD) and the standard deviation of 98th percentile of the transmittance curve (TRANS_P98_SD). It was negatively correlated with minimum waveform height (CHP_H_MIN). Increases in minimum canopy height will, as long as other aspects of canopy structure remain constant, tend to reduce the range of canopy variability.In contrast to the second canonical variable, increases in variability were negatively correlated with cover-related indices, such as cover (COVER_X and MNH_COV) and the oligophotic volume (OLIGO).

Stand structure. The third stand structure canonical variable had a high negative correlation (Table 9) with basal area of deciduous trees (DECID_BA). A positive correlation between canopy height variability (CHP_H_SD) and the corresponding canopy structure canonical variable indicates that as deciduous species, which tend to hold their leaves higher in the canopy than the conical conifer crowns, increase in importance, minimum height increases and as a consequence height variability decreases. The increase in cover may be a result of that colonization.

3.1.1.4. Canonical variable pair 4

Canopy structure. The fourth canonical variable accounted for 6% of total variance explained (Table 6) and was related to a shift in foliage distribution to higher levels in the canopy, as indicated by positive correlations (Table 8) with mean height and quadratic mean height of the canopy height profile (CHP_MN_X, CHP_Q_X). This is also related to an increase in the total volume occupied by the canopy (FILLED) and a decrease in open space (OPEN) above the local canopy. This suite of correlations suggests that this variable is related to canopy structure associated with mature stand structure.

Stand structure. The fourth stand structure canonical variable was also indicative of mature stand structure. Increases in the stand structure canonical variable were correlated (Table 9) with aboveground biomass (BIOMASS), LAI, and COVER, all indicative of a canopy dominated by relatively even-sized stems, whose high cover and LAI are associated with simplified stand structure. This suite of correlations indicates a contrast between young and mature stands.

3.1.1.5. Canonical variable pair 5

Canopy structure. The fifth canonical variable accounted for 4% of the variance explained (Table 6), and was again related to reduced canopy variability, as expressed by low standard deviations of waveform height (CHP_H_SD, Table 8) and canopy transmittance (TRANS_P98_SD). This is correlated with increases in the minimum waveform height (CHP_H_MIN) and the volume of oligophotic canopy structure (OLIGO), and a decrease in open space (OPEN) above the local canopy.

Stand structure. The stand structure indices correlating positively (Table 9) with the corresponding stand structure canonical variable include LAI, basal area (BASAL), density (DENSITY) and coniferous basal area (CONIF_BA). Examination of site averages for this canonical variable indicated that it was roughly indicative of the coniferous/deciduous ratio of each area.

3.1.1.6. Canonical variable pair 6

Canopy structure. The sixth canonical variable explained 3% of variance (Table 6) and exhibited a positive correlation (Table 8) with cover and negative correlations with both the standard deviation of quadratic mean height (CHP_Q_SD) and the volume of oligophotic foliage (OLIGO). These conditions also discriminate the typical mature condition (with a single canopy layer, high cover, low height variability and low volume of oligophotic

foliage) from old-growth conditions (multiple canopy layers, lower cover, greater height variability and high volume of oligophotic foliage) as opposed to canonical variable 4 which separates young and mature stands.

Stand structure. The sixth stand structure canonical variable was positively correlated (Table 9) with the mean DBH of all stems (DBHX) and is negatively correlated with LAI, basal area (BASAL), and the maximum DBH (DBHMAX). Taken together, the canopy and stand variables indicate mature stands with high mean DBH, low oligophotic volume, and low height variability, which will in turn result in lower indices of total stand structure (such as LAI and basal area).

3.1.1.7. Canonical variable pair 7

Canopy structure. The seventh and last significant canonical variable explained 3% of the total variance (Table 6) and was most closely correlated (Table 8) with the number of waveform heights above 55 m. As with the fourth canonical variable, this correlation, and the negative correlation with cover and oligophotic canopy volume (OLIGO) were indicative of mature forests. The interpretation of canonical variables 4, 6 and 7 as indicative of contrasts between mature and old-growth plots does not contradict the rule that individual canonical variables are uncorrelated, because there is little overlap between the specific variables involved, indicating that each of these three canonical variables are related to different distinguishing traits of mature and old-growth stands (e.g. Franklin & Spies, 1991).

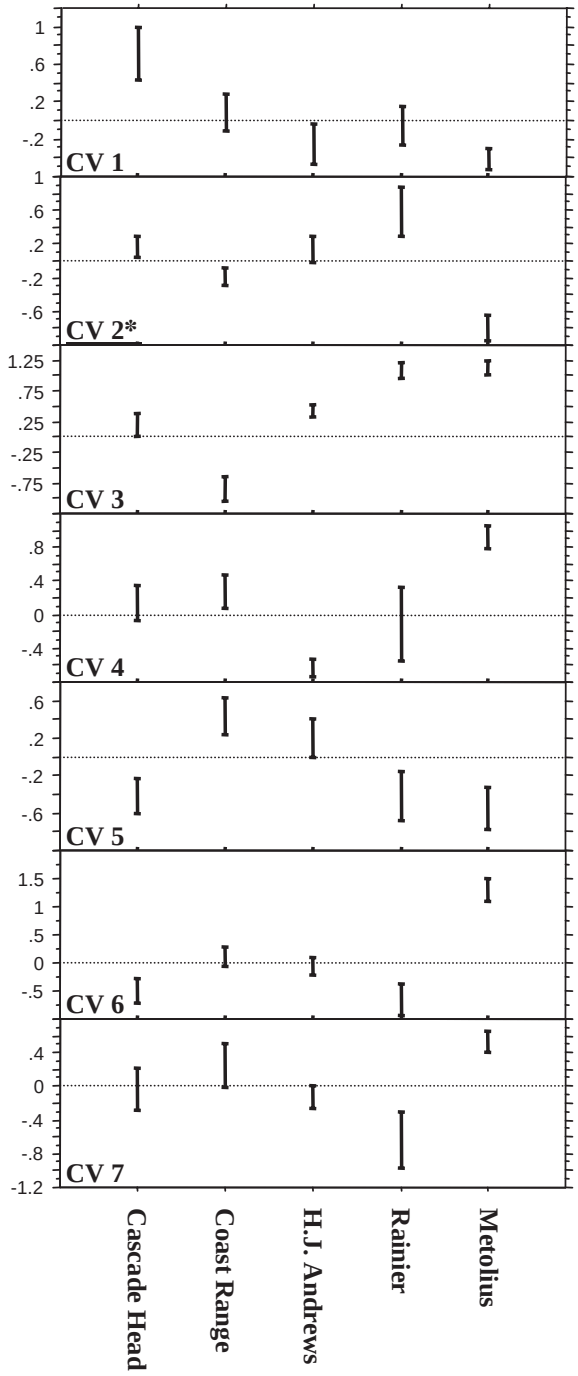
Stand structure. Correlations (Table 9) between the stand structure canonical variable and the stand structure variables were generally modest. The positive correlation with the number of stems greater than 100 cm, and the negative correlation with log-transformed density are consistent with the interpretation of this variable as related to another contrast between mature and old-growth forests.

4. Discussion

4.1. Site analysis

The first canonical variable reflected the time since stand replacement disturbance of each site, as indicated by increasing field and lidar measured height, and in aboveground biomass and standard deviation of DBH. Fig. 3 indicates sites were ordered according to productivity, with the highest score going to the Cascade Head, and the lowest to Metolius. Metolius received the lowest score because of lower tree density, which gave it the lowest average aboveground biomass and mean canopy height (due to the wider spacing between dominant trees).

The second canonical variable was most closely related to LAI, foliage cover, an increased volume of dimly lit space, and variability of canopy surface height. A consequence of this increase was increased density and leaf area index. Although the first and second canonical variables lack a significant linear correlation (a constraint of the CCA process), there is a significant third order relationship between these two canonical variables (Fig. 4) that approaches the "horseshoe" response found in factor analysis of many ecological datasets. To properly analyze



* Detrended by CV1

Fig. 3. Means and standard errors for the seven lidar-estimated canonical variables for each of the five study locations.

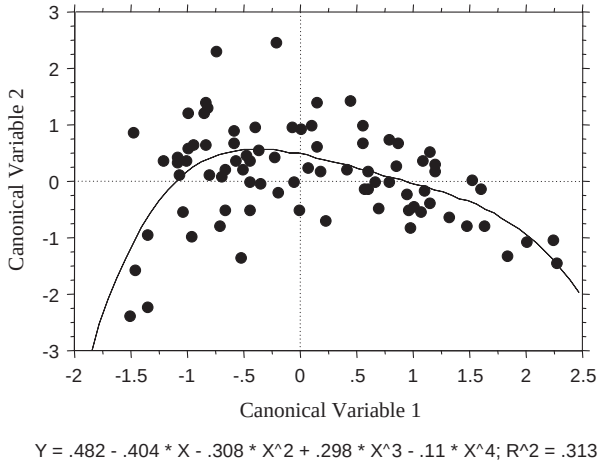


Fig. 4. Relationship between canonical variables 1 and 2, illustrating the "horseshoe" shaped relationship between the two variables.

site scores for this variable, the residual of the second canonical variable (with respect to the first) was calculated and used in Fig. 3. Adjusted in this way, the Metolius site score decreased while the Rainier site score increased for Factor 2. This is also true of LAI for these sites (although not for cover), which is consistent with the interpretation of canonical variable 2 as related to LAI. This also indicates that this score is sensitive to LAI differences across the study region, and not simply as a response to the first order effect of leaf area index increasing with increased height.

The third pair of canonical variables is directly related to horizontal spatial variability in vertical canopy structure, as described by the numerous statistics describing the standard deviation of various height indices, and by increased minimum heights (which decrease variability). Canopy height variability has been linked to the overall successional state (with older forests having canopies of more variable height), the balance of shade-tolerant and shade-intolerant species, and the standard deviation of DBH in Douglas-fir forests (Lefsky et al., 1999b). In this analysis, canopy variability separates coniferous and deciduous dominated plots, although almost all of these plots occur at the Coast Range location (Fig. 3), so further investigation of the generality of this observation is necessary.

These three main effects, which account for 84% of the variance explained by the analysis, represent a set of axes to which observed canopy variation in the Pacific Northwest can be compared. Lefsky et al. (1999b) observed two main axes of variation in canopy structure, related to mean height and height variability, in stands of Douglas-fir/western hemlock at the H.J. Andrews Experimental Forest. With the inclusion of a wider variety of stands, leaf area index (LAI) becomes more important than stand height variability as a source of variance in the dataset. On this second axis of variability, Metolius and Rainier have (respectively) lower and higher values of LAI, with respect to their height.

Canonical variables 4 through 7 are statistically significant but represent smaller fractions of the total variance in common between the canopy and stand structure datasets. In a related study (Lefsky et al., 2005), the inclusion of canonical variables 4–7 in a stepwise multiple regression of 17 stand structure variables added an average of 8% of variance explained, in comparison to regression using canonical variables 1–3 only. However, these variables are less stable, and their interpretation is less certain than the first three (Table 7).

4.2. Classes of canopy structure variables

Lefsky et al. (1999b) reported that one class of variables, those defined by the canopy volume method, were most important in stepwise multiple regressions of stand structure indices. In these analyses, important variables were classified as those variables which fall within the top 25th percentile of correlation magnitude. To estimate the relative importance of each class of variables, the ratio of the number of times that a variable was important for any axis to the number of times it could have been important (i.e. the number of variables in the class multiplied by the number of canonical axes) was calculated. Six of the 29 independent variables fell in the canopy volume method class, and they could each have been "important" in predicting any of the 7 canonical variables. Of these 42 opportunities, they were actually important in 16 cases, for an importance rate of 38%. In comparison, the canopy surface height rate was 22%, the canopy height profile rate was 27% and the canopy transmittance rate was 24%. When the hypothesis that the difference in rates was significant was tested using ANOVA and the Scheffe post-hoc test, the 38% value was not significantly different than the values for the other classes. Nevertheless, the marginally non-significant p-value ($P=0.07$) suggests that canopy volume indices may describe canopy structure more accurately than traditional measurements.

4.3. Approaches to describing canopy structure

One key finding of this work was that most canonical variables had strong correlations with canopy structure indices developed using all four methods of describing canopy structure, despite the differing analytical frameworks supporting each method. There are two exceptions: none of the strongest correlations with the first canonical variable were with canopy height profile related variables, and none of the strongest correlations with the fourth canonical variable were with the canopy surface height variables. Nevertheless, it is clear that a wide range of variables are, for statistical purposes, redundant.

5. Conclusion

Existing work relating lidar measurements to both canopy and stand structure has (understandably) been

focused on what can be achieved, and after two decades of work, it is clear that lidar is an exceptional tool for forest remote sensing. Numerous different methods for interpreting lidar data have emerged, both as a consequence of the steady improvement in the quality and spatial density of the data itself, and as a function of the tools and backgrounds of individual researchers themselves. While the focus on empirical results has established the utility of lidar, the focus on improving our predictive ability has overshadowed the similarity (for predictive purposes) between indices that have very different conceptual bases. Ultimately, the utility of these indices for predicting stand structure lies in their ability to accurately summarize important aspects of stand structure. The goal for this paper was the grouping of redundant canopy and stand indices, and a ranking of the numerous aspects of canopy and stand structure that significantly covary. In this sense, and for the study area involved, the first three factors (mean height, cover or leaf index area, height variability) represent the same kind of enhancement of lidar data that the tasseled cap indices (Crist & Cicone, 1984) represent for optical remote sensing. Whether for lidar or optical data, each method summarizes a large number of potential variables into a small number of indices that can be used to quickly assess the information in the total dataset. Nevertheless, verification in other forest types, and adaptation to other sensors, will be needed before this result can be widely accepted.

Acknowledgments

Thanks are due to four anonymous reviewers who took the time to review this and a related manuscript. This work was supported by a grant from the Terrestrial Ecology Program of NASA to Drs. Cohen and Lefsky. Development of the SLICER instrument was supported by NASA's Solid Earth Science Program and the Goddard Director's Discretionary Fund. SLICER data sets available for public distribution are documented at http://core2.gsfc.nasa.gov. Acquisition of the SLICER data used here was supported by a Terrestrial Ecology Program grant to Dr. Harding.

Appendix A. Symbols and Abbreviations

Lidar canopy structure variable abbreviations	
Canopy surface height indices	
CHP_H_X	Mean of maximum canopy heights
CHP_H_2	Mean of maximum canopy heights, squared
CHP_H_M	Median of maximum canopy heights
CHP_H_M2	Median of maximum canopy heights, squared
CHP_H_SD	Standard deviation of maximum canopy heights
CHP_H_MAX	Maximum of maximum canopy heights
CHP_H_MAX2	Maximum of maximum canopy heights, squared
CHP_H_MIN	Minimum of maximum canopy heights
HGT55	Number of canopy heights above 55 m
Canopy height profile indices	
COVER_X	Mean cover
CHP_MN_X	Mean of mean canopy heights
CHP_MN_SD	Standard deviation of mean canopy heights
CHP_Q_X	Mean of quadratic mean canopy heights
CHP_Q_X2	Mean of quadratic mean canopy heights, squared
CHP_Q_SD	Standard deviation of quadratic mean canopy heights
MNH_COV	Product of CHP_MN_X and COVER_X
QMCH_COV	Product of CHP_Q_X and COVER_X
Canopy transmittance indices	
TRANS_MN_X	Mean of the mean transmittances height in the canopy
TRANS_MN_SD	Standard deviation of the mean transmittance heights in the canopy
TRANS_P50_X	Mean of the height of the 50th percentile of transmittances in the canopy
TRANS_P50_SD	Standard deviation of the 50th percentile height of transmittances in the canopy
TRANS_P98_X	Mean of the height of the 98th percentile of transmittances in the canopy
TRANS_P98_SD	Standard deviation of the 98th percentile height of transmittances in the canopy
Canopy volume indices	
OPEN	Volume of space with no canopy present, above local canopy height but below maximum plot height[a]
CLOSED	Volume of space with no canopy present, below other foliage[a]
EUPHOTIC	Volume of space with canopy present, brightly lit[a]
OLIGO	Volume of space with canopy present, dimly lit[a]
FILLED	EUPHOTIC+OLIGO[a]
LCOMP	Linear complexity score[a]
Stand Structure Variable Abbreviations	
LAI	Leaf area index (m^2m^{-2})
Basal	Basal area (m^2ha^{-1})
Density	Stem density (ha^{-1})
lnDensity	Natural log transformed stem density
NT100CM	Density of stems greater than 100 cm (ha^{-1})
DBHMAX	Maximum DBH (cm)
DBHX	Mean DBH of all stems (cm)
DBHU	Mean DBH of dominant and co-dominant stems (cm)
DBHSTD	Standard deviation of DBH (cm)
DECID_BA	Basal area of deciduous species (m^2ha^{-1})
CONIF_BA	Basal area of coniferous species (m^2ha^{-1})
BIOMASS	Aboveground Biomass ($Mgha^{-1}$)
HTMAX	Maximum tree height of both measured and allometrically predicated trees (m)
HTMAXM	Maximum tree height of just measured trees (m)
HTDCD	Mean height of dominant and co-dominant trees (m)
LOREY	Lorey's height: $h_Z=\sum g*h/\sum g$ where g=basal area of individual trees, h=height of individuals trees
COVER	Aerial cover of foliage and woody material

[a] See text and Lefsky et al., 1999b for detailed explanations.

References

Ahern, F. J., Janetos, A. C., & Langham, E. (1998). Global observation of forest cover: One component of CEOS' Integrated Global Observing Strategy. *27th International Symposium on Remote Sensing of Environment,* Tromsø, Norway.

Blair, J. B., Coyle, D. B., Bufton, J. L., & Harding, D. J. (1994). Optimization of an airborne laser altimeter for remote sensing of vegetation and tree canopies. *Proceedings of the International Geosciences Remote Sensing Symposium* (pp. 939–941). Pasadena, CA: California Institute of Technology.

Blair, J. B., & Hofton, M. A. (1999). Modeling laser altimeter return waveforms over complex vegetation using high-resolution elevation data. *Geophysical Research Letters*, *26*, 2509–2512.

Brown, G. (1979). An optimization criterion for linear inverse estimation. *Technometrics*, *2*, 575–579.

Burton, A. J., Pregitzer, K. S., & Reed, D. D. (1991). Leaf area and foliar biomass relationships in northern hardwood forests located along an 800 km acid deposition gradient. *Forest Science*, *37*, 1041–1059.

Cohen, W. B., Harmon, M. E., Wallin, D. O., & Fiorella, M. (1996). Two decades of carbon flux from forests of the Pacific Northwest. *Bioscience*, *46*, 836–844.

Cohen, W. B., Maiersperger, T. K., Gower, S. T., & Turner, D. P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+. *Remote Sensing of Environment*, *84*, 561–571.

Crist, E. P., & 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.

Curran, P. J., & Hay, A. (1986). The importance of measurement error for certain procedures in remote sensing at optical wavelengths. *Photogrammetric Engineering and Remote Sensing*, *52*, 229–241.

Daly, C., Taylor, G., & Gibson, W. (1997). The PRISM approach to mapping precipitation and temperature, 10th Conf. on Applied Climatology, Reno, NV. *Amer. Meteor Soc.*, 10–12.

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

Franklin, J. F., & Spies, T. A. (1991). Composition, function, and structure of old-growth Douglas-fir forests. In L. F. Ruggiero, K. B. Aubry, A. B. Carey, M. H. Huff (Eds.), *Wildlife and Vegetation of Unmanaged Douglas-Fir Forests, General Technical Report PNW-GTR-285* (pp. 71–82). Portland, Oregon: USDA Forest Service Pacific Northwest Research Station.

Gholz, H. L., Fitz, F. K., & Waring, R. H. (1976). Leaf area differences associated with old-growth forest communities in the Western Oregon Cascades. *Canadian Journal of Forest Research*, *6*, 49–57.

Gholz, H. L., Grier, C. C., Campbell, A. G., & Brown, A. T. (1979). *Equations for estimating biomass and leaf area of plants in the Pacific Northwest research paper 41* (p. 37). Corvallis, OR: Forest Research Lab., School of Forestry, Oregon State University.

Harding, D. J., Blair, J. B., Garvin, J. B., & Lawrence, W. T. (1994). Laser altimetry waveform measurement of vegetation canopy structure. *Proceedings of the International Remote Sensing Symposium* (pp. 1251–1253). Pasadena, CA: California Institute of Technology.

Harding, D. J., Lefsky, M. A., & Parker, G. G. (2001). Lidar altimeter measurements of canopy structure: Methods and validation for closed-canopy broadleaf forest. *Remote Sensing of the Environment*, *76*, 283–297.

Harmon, M. E., & Franklin, J. F. (2002). *Dendrometer Studies for Stand Volume and Height Measurements: Long-Term Ecological Research.* Corvallis, OR: Forest Science Data Bank: TV009. [Database]. http://www.fsl.orst.edu/lter/data/abstract.cfm?dbcode=TV009 (29 August 2003).

Lefsky, M. A., Harding, D., Cohen, W. B., & Parker, G. G. (1999a). Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. *Remote Sensing of the Environment*, *67*, 83–98.

Lefsky, M. A., Cohen, W. B., Acker, S. A., Spies, T. A., Parker, G. G., & Harding, D. (1999b). Lidar remote sensing of biophysical properties and canopy structure of forest of Douglas-fir and western hemlock. *Remote Sensing of Environment*, *70*, 339–361.

Lefsky, M. A., Cohen, W. B., Harding, D. J., Parker, G. G., Acker, S. A., & Gower, S. T. (2002). Lidar remote sensing of aboveground biomass in three biomes. *Global Ecology and Biogeography*, *11*(5), 393–400.

Lefsky, M.A., Hudak, A.T., Cohen, W.B., & Acker, S.A. (2005). Geographic variability in lidar predictions of forest stand structure in the Pacific Northwest. *Remote Sensing of Environment*, *95*, 532–548.

MacArthur, R. H., & Horn, H. S. (1969). Foliage profile by vertical measurements. *Ecology*, *50*, 802–804.

Maclean, G. A., & Krabill, W. B. (1986). Gross-merchantable timber volume estimation using an airborne LIDAR system. *Canadian Journal of Remote Sensing*, *12*, 7–18.

Means, J. E., Acker, S. A., Harding, D. J., Blair, J. B., Lefsky, M. A., Cohen, W. B., et al. (1999). Use of large-footprint scanning airborne lidar to estimate forest stand characteristics in the western Cascades of Oregon. *Remote Sensing of the Environment*, *67*, 298–308.

Means, J. E., Hansen, H. A., Koerper, G. J., Alaback, P. B., & Klopsch, M. W. (1994). Software for computing plant biomass—BIOPAK users guide. *PNW Research Station*. Portland, Oregon: USDA F.S.

Moeur, M. A., & Stage, R. S. (1995). Most similar neighbor: An improved sampling inference procedure for natural resource planning. *Forest Science*, *41*(2), 337–359.

Nelson, R. F., Krabill, W. B., & Maclean, G. A. (1984). Determining forest canopy characteristics using airborne laser data. *Remote Sensing of Environment*, *15*, 201–212.

Parker, G. G., Lefsky, M. A., & Harding, D. J. (2001). PAR transmittance in forest canopies determined from airborne lidar altimetry and from in-canopy quantum measurements. *Remote Sensing of the Environment*, *76*, 298–309.

Running, S. W., Baldocchi, D. D., Turner, D. P., Gower, S. T., Bakwin, P. S., & Hibbard, K. A. (1999). A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. *Remote Sensing of Environment*, *70*, 108–127.

SAS Institute. (1990). *SAS/STATR user's guide, version 6, vol. 1–2* (4th ed.). North Carolina, USA: Cary, 943 pp. and 846 pp.

Schumacher, F. X., & Hall, F. D. S. (1933). Logarithmic expression of timber-tree volume. *Journal of Agricultural Research*, *47*, 719–734.

Steel, R., & Torrie, J. (1980). *Principles and Procedures of Statistics—a Biometrical Approach* (2nd ed.). New York: McGraw-Hill.

Tabachnick, B., & Fidell, L. (1989). *Using Multivariate Statistics* (2nd ed.). United Kingdom: Harper Collins Publishers.

USDA Forest Products Laboratory. (1999). *Wood Handbook: Wood as an Engineering Material. Gen. Tech. Rep. FPL-GTR-113.* Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory. 463 p.

Waring, R. M. (1982). Estimating forest growth and efficiency in relation to canopy leaf area. *Advances in Ecological Research*, *13*, 327–354.