Available online at www.sciencedirect.com

SCIENCE DIRECT

Remote Sensing of Environment 95 (2005) 532–548

Remote Sensing of Environment

www.elsevier.com/locate/rse

Geographic variability in lidar predictions of forest stand 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 80523-1470, United States*
[b]*USDA Forest Service, Forest Sciences Laboratory, Rocky Mountain Research Station, 1221 South Main Street, Moscow, ID 83843,United States*
[c]*USDA Forest Service, Forestry Sciences Laboratory, Pacific Northwest Research Station, 3200 SW Jefferson Way, Corvallis, OR 97331, United States*
[d]*Olympic National Park, 600 East Park Avenue, Port Angeles, WA 98362-6798, United States*

Received 6 May 2004; received in revised form 3 January 2005; accepted 26 January 2005

Abstract

Estimation of the amount of carbon stored in forests is a key challenge for understanding the global carbon cycle, one which remote sensing is expected to help address. However, carbon storage in moderate to high biomass forests is difficult to estimate with conventional optical or radar sensors. Lidar (*li*ght *d*etection *a*nd *r*anging) instruments measure the vertical structure of forests and thus hold great promise for remotely sensing the quantity and spatial organization of forest biomass. In this study, we compare the relationships between lidar-measured canopy structure and coincident field measurements of forest stand structure at five locations in the Pacific Northwest of the U.S.A. with contrasting composition. Coefficient of determination values (r^2) ranged between 41% and 96%. Correlations for two important variables, LAI (81%) and aboveground biomass (92%), were noteworthy, as was the fact that neither variable showed an asymptotic response.

Of the 17 stand structure variables considered in this study, we were able to develop eight equations that were valid for all sites, including equations for two variables generally considered to be highly important (aboveground biomass and leaf area index). The other six equations that were valid for all sites were either related to height (which is most directly measured by lidar) or diameter at breast height (which should be closely related to height). Four additional equations (a total of 12) were applicable to all sites where either Douglas-fir (*Pseudotsuga menziesii*), western hemlock (*Tsuga heterophylla*) or Sitka spruce (*Picea sitchensi)* were dominant. Stand structure variables in sites dominated by true firs (*Abies sp.*) or ponderosa pine (Pinus ponderosa) had biases when predicted by these four additional equations. Productivity-related variables describing the edaphic, climatic and topographic environment of the sites where available for every regression, but only two of the 17 equations (maximum diameter at breast height, stem density) incorporated them. Given the wide range of these environmental conditions sampled, we conclude that the prediction of stand structure is largely independent of environmental conditions in this study area.

Most studies of lidar remote sensing for predicting stand structure have depended on intensive data collections within a relatively small study area. This study indicates that the relationships between many stand structure indices and lidar measured canopy structure have generality at the regional scale. This finding, if replicated in other regions, would suggest that mapping of stand structure using lidar may be accomplished by distributing field sites extensively over a region, thus reducing the overall inventory effort required.

Keywords: Lidar; Laser; Biomass; Forest; Regional; Inventory

1. Introduction

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

* 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 Inc. All rights reserved.
doi:10.1016/j.rse.2005.01.010

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).

However, direct estimation of carbon storage in moderate to high biomass forests remains a difficult task for remote sensing. While remote sensing has had considerable success in measuring the biophysical characteristics of vegetation in areas where plant canopy cover is relatively sparse, quantification of vegetation structure where leaf area index (LAI) exceeds three has been less successful (Carlson & Ripley, 1997; Turner et al., 1999; Waring et al., 1995). High LAI forests, which generally have high aboveground biomass, occur in boreal, temperate and tropical regions. These forests cover less than 35% of the Earth's terrestrial surface, yet account for 67% of terrestrial net primary productivity (NPP) and 89% of terrestrial biomass (Waring & Schlesinger, 1985). Given their prominent role in global biogeochemistry and the likelihood that these high productivity areas will be prime areas for carbon sequestration efforts, better estimates of carbon storage in high biomass forests is desirable.

One promising remote sensing technique is lidar. Lidar instruments directly measure the vertical structure of vegetation by determining the distance between the sensor and a 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 that light pulse reflected from the target. Waveform-recording lidar systems, such as the SLICER (Scanning Lidar Imager of Canopies by Echo Recovery) instrument used in this work (Blair et al., 1994; Harding et al., 1994; Harding et al., 2001) and the Laser Vegetation Imaging System (LVIS, Blair & Hofton, 1999) measure the time-resolved amount of laser energy reflected from the many surfaces of a geometrically complex target. The distribution of return energy reflected from a vegetation surface, the lidar waveform, records the vertical distribution of illuminated vegetation and soil surfaces from the top of the canopy to the ground. For forests, a primary research goal has been relating these waveforms to conventional, primarily non-spatial, measurements of forest structure, such as aboveground biomass and stand basal area (Drake et al., 2002; Lefsky et al., 1999a,b; Lefsky et al., 2002; Means et al., 1999). In this study, we compare the relationships between lidar-measured canopy structure and coincident field measurements of aboveground biomass at five locations in the Pacific Northwest of the U.S.A., each with contrasting environmental conditions, productivity and species composition.

The goal of this work is to test the potential for regionally applicable relationships between lidar estimates of canopy structure and field estimates of stand structure. Five methods are evaluated for their ability to create unbiased regression equations that apply to all sites. Knowledge of the generality of these equations will help determine the effort and expense required to develop global forest structure estimates, including aboveground biomass, from lidar data.

In this study we:

1. Describe the intersite variability of relationships predicting forest stand structure from lidar estimated canopy structure.
2. Test the ability of environmental data to account for intersite biases.
3. Determine optimal methods for regression of multiple stand structure variables against multiple canopy structure indices.

2. Methods

2.1. Study areas

Field data were collected in five locations, selected to sample the maximum practicable range of environment conditions and forest composition in the Pacific-Northwest region of the United States. Considering only the forested areas of Washington and Oregon, our sites covered 71.4% of the variation in precipitation and 77.6% of the variation in mean annual temperature. Tree composition at these sites 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 Metolious 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. Further description of the study areas are available in (Lefsky et al. in review A).

2.2. Field measurements

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. Eighty-four 0.25 ha field plots

were established beneath SLICER transects flown in 1995; most plots were associated with a five-by-five array of SLICER footprints. Only forested sites were sampled, using a nested plot design that recorded species, diameter at breast height (DBH), and crown ratio (the proportion of the bole with live crown) for all trees, and tree height for a subset of trees. Plot level estimates of leaf area index were predicted using species-specific equations from sapwood area or diameter at breast height, depending on species (Lefsky et al. in Review).

Total aboveground biomass was estimated from DBH and 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 (Lefsky et al. in review A). The Schumacher equation (Schumacher & Hall, 1933), which uses both the height and diameter of trees to predict stem volume, was adopted to avoid minimizing the effects of site productivity on estimates of aboveground biomass at each site. To generate heights for trees which did not have tree height measurements, we used an imputation procedure (Moeur & Stage, 1995) to select most similar trees from a database of over 300,000 trees combined from the Current Vegetation Survey and Forest Inventory Analysis data bases, which all had measured heights. More details on the field measurements are available in (Lefsky et al. in review A).

2.3. SLICER measurements and data analysis

Lidar waveforms were collected by the SLICER 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. 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 used 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 (Parker et al., 2001). 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. (1999a) and Lefsky et al. (in review A).

2.4. Statistical analysis

2.4.1. 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), and 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 the other. In addition, when there is only one Y (e.g. LAI), CCA provides a set of coefficients for the X's that aligns them with the variation in the Y variable. However, CCA does not scale the resulting variable according to the units of the dependent variable, a step that in this analysis was performed using Reduced Major Axis regression.

2.4.2. Reduced major axis regression

Reduced major axis regression (RMA) is one of a class of similar models variously known as orthogonal regression, total least squares regression, or errors-in-variables modeling (Van Huffel, 1997). Orthogonal regression minimizes the sum of squared orthogonal distances from measurement points to the model function. Besides making no assumptions about errors in X and Y, RMA likewise makes no assumptions about dependency. Conrad and Gutmann (1996) refer to RMA as geometric mean regression, in that the slope is defined as the ratio of sample standard deviation for Y over the sample standard deviation for X, thus preserving in the model the relative variance structure of the sample dataset. The effect is to minimize or eliminate any attenuation or amplification of predictions. Mathematical similarities in the formulations of OLS and RMA regression models mean that the model intercepts are all equivalent, as are the coefficients of determination. What differ among these models are the root mean square errors (RMSEs) and the slopes of the relationships.

Bootstrapping was used to provide robust estimates of three parameters in this analysis: R^2 and the 95% confidence intervals of the slope and intercept parameters. Random subsets of the same size as the full dataset were drawn, with replacement, from the full dataset of plots. For each subset, 10,000 iterations were used to estimate the three parameters. From these estimates, mean values were determined, and the 95% confidence intervals were identified as the 500th and 9500th values in the sorted sequence of the slope and intercept arrays. These intervals were subsequently used to determine if the slope or

intercept was significantly different from an identity relationship (ie. slope=1 and intercept=0).

In comparing the R^2 estimates from RMA to results from stepwise regression alone, it was noted that RMA consistently produced higher values. However, it was established that the comparison was faulty—the stepwise regression results were (appropriately) *adjusted* R^2 values, and thus were corrected for potential model overfit due to the large number of independent variables. This potential model overfit would exist in any case when a large number of independent variables are being used to model a single dependent variable, even when the independent variables are summarized as a single index. Therefore, R^2 values from the CCA were adjusted following Healy (1984),

$$\mathrm{adj}R^2 = 1 - \frac{n-1}{n-m-1}\left(1-R^2\right) \quad (1)$$

where n is the number of observations, m is the number of independent variables, and R is the raw multiple correlation coefficient.

2.4.3. Roadmap for statistical analyses

Three sets of statistical analyses were performed. The first set of analyses compared three methods for relating lidar-measured canopy structure and field-measured stand structure. The second set of analyses tested the ability of environmental (topographic, climate and edaphic) indices to explain the residuals from the first set of regression analyses. Finally, variables derived from a canonical correlation analysis of environmental variables were added to the original regression datasets, and regressions were recalculated.

The first statistical analysis had three steps. First a CCA was performed to document patterns of variance and covariance in the lidar and stand structure datasets (this analysis is detailed in Lefsky et al. in review). Second, the three regression methods were compared to pick one that was most appropriate for estimating the multivariate relationships between lidar estimates of canopy structure and field measurements of stand structure. The regression methods used were 1) direct stepwise multiple regression with canopy structure variables (e.g. direct stepwise) which was used as a reference, 2) direct CCA with canopy structure variables (e.g. direct CCA) and 3) stepwise multiple regression with canonical *v*ariables (referred to here as SCV). The difference between direct CCA and SCV is that in the former, the CCA is performed with a single dependant variable (the stand structure index of interest) as in Cohen et al. (2003), whereas, in SCV, the multiple canonical variables derived from the canopy structure dataset are combined using stepwise multiple regression to predict the dependant variable. Third, as in Cohen et al. (2003), RMA was used after each analysis to scale the resulting canonical variable to the units of the variable in question, thus avoiding the biases associated with OLS regression (Cohen et al., 2003). For the stepwise analyses in this paper, scaling was not required, but RMA removes the biases introduced by OLS regression.

The second analysis involved a second round of CCA to relate residuals from each of the three regression analyses to topography, climate (Daly et al., 1997) and soils (USDA, 1994). The use of CCA in this context avoided the inflation of variance explained by the environmental variables that would have occurred if all the environmental variables had been included in the first set of regressions. Moreover, the subsequent CCA allowed us to define important environmental factors that influence the stand structure variables of interest. Finally, the environmental canonical variables and the lidar estimates of canopy structure were then used together to estimate stand structure.

3. Results and interpretation

Due to the complexity of this multi-layered analysis, initial interpretation of the results (e.g., the axes defined by the canonical correlation analysis) will be presented along with the results themselves. Higher-level analysis of the results (e.g. the ecological significance of the particular pattern of extracted axes) will be left for the Discussion.

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 1). Canonical correlation coefficients (the correlation between the pairs of canonical variables for the two datasets) ranged from 0.99–0.79 (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 ($P<0.0001$).

Details of this analysis are presented in Lefsky et al. A (in review), and will be summarized here (Table 2). The first canonical variable, as expected, reflected the management and disturbance history of each site, as reflected in positive correlations with both field and lidar measured height, and with aboveground biomass. The second canonical variable was correlated with LAI, foliage cover, and the volume of dimly lit space. The third pair of canonical variables was correlated with the horizontal spatial variability in canopy vertical structure, as indicated by positive correlations with the statistics describing the standard deviation of various height indices, and by a negative correlation with minimum heights (because higher minimum heights decrease variability). The third pair of canonical variables was also highly correlated with the basal area of deciduous trees. Canonical variables 1–3 explained 84% of variance in the analysis.

Table 1
Canonical correlation analysis: canonical variable summary

Canonical correlation pair	Canonical correlation	Approximate standard error	Squared canonical correlation	Eigen value	Percent of variance	Pr>F
1	0.99	0.00	0.97	37.6528	61%	<0.0001
2	0.95	0.01	0.90	8.9713	15%	<0.0001
3	0.91	0.02	0.83	4.7546	8%	<0.0001
4	0.89	0.02	0.80	3.9206	6%	<0.0001
5	0.83	0.03	0.70	2.2863	4%	<0.0001
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

Statistics	Value	F Value	Num DF	Den DF	Pr>F
Wilk's Lambda	0.00	2.84	486	688.77	<0.0001
Pillai's Trace	8.57	1.95	486	1044	<0.0001
Hotelling–Lawley Trace	66.54	5.37	486	320.49	<0.0001
Roy's Greatest Root	37.65	80.88	27	58	<0.0001

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. Canonical variables 4, 6 and 7 are related to various contrasts in the structural conditions associated with young, mature and old-growth stands, such as the number of waveforms greater than 55 m, the volume of shadowed canopy, and the mean DBH of all stems. Canonical variable 5 is related to the proportion of deciduous and coniferous basal area.

3.2. Regression analysis

3.2.1. Stepwise multiple regression using canopy structure variables (direct stepwise)

Abridged results from 17 stepwise multiple regressions (one for each dependent variable) are given in Table 3, which contains the correlation coefficients between each of the dependent and independent variables. The results of the stepwise multiple regressions are indicated by highlighting the variables selected by the stepwise analysis (and therefore each column can be thought of as a summary of the resulting equation). Adjusted R^2 values from each equation are indicated in the bottom row, and ranged from 0.0 for natural log transformed density (ln Density) to 0.92 for Aboveground Biomass (BIOMASS), with a median value of 0.76. Explanatory variables were evenly spread among the four methods for the description of forest canopies: canopy surface height measurements (each variable involved in an average of 1.89 equations), canopy height profile measurements (1.62), canopy transmittance indices (2.5), and canopy volume indices (1.83).

3.2.2. Direct CCA using lidar estimated canopy structure

Abridged results from direct CCA (using CCA to predict each of the stand structure variables directly from the lidar indices of canopy structure) are presented in Table 4. Correlations between the canopy structure variables and the canonical variable resulting from each analysis are given in the table. Correlations were higher than those in Table 3, because (in this analysis) they were between the lidar indices and the canonical variables, which were a function of the original canopy structure indices. Grey values indicated those variables which were considered statistically significant in a stepwise multiple regression of each canonical variable using the independent canopy variables. Again, explanatory variables were more or less evenly spread among the four methods for the description of forest canopies: canopy surface height measurements (each variable involved in an average of 7.1 equations), canopy height profile measurements (10.7), canopy transmittance indices (10.3), and canopy volume indices (7.7).

Table 2
Summary of canonical pairs

Canonical variable	Description of ecological significance	Lidar index 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; seperates mature and old-growth	CHP_Q_SD
7	Cover oligophotic canopy volume, correlates with mature stands	HGT55

Table 3
Pearson correlations between field measured stand structure and lidar measured canopy structure variables

	BASAL	BIOMASS	CONIF_BA	COVER	DBHMAX	DBHSTD	DBHU	DBHX	DECID_BA	DENSITY	HTDCD	HTMAX	HTMAXM	LAI	LN_DENSITY	LOREY	NT100CM	Equations using this variable
Canopy Surface Height Indices																		
CHP_H_X	0.82	0.91	0.75	0.55	0.84	0.84	0.80	0.67	0.06	–0.20	0.89	0.89	0.90	0.55	–0.04	0.93	0.78	0
CHP_H_X2	0.79	0.92	0.74	0.45	0.81	0.84	0.80	0.68	–0.01	–0.22	0.86	0.83	0.87	0.48	–0.13	0.90	0.83	1
CHP_H_M	0.81	0.90	0.75	0.53	0.84	0.85	0.82	0.66	0.04	–0.20	0.89	0.89	0.91	0.54	–0.03	0.93	0.78	3
CHP_H_M2	0.77	0.91	0.74	0.42	0.80	0.85	0.81	0.66	–0.03	–0.23	0.87	0.83	0.87	0.46	–0.13	0.90	0.85	3
CHP_H_SD	0.33	0.41	0.37	0.03	0.59	0.61	0.67	0.20	–0.16	–0.14	0.59	0.57	0.59	0.14	0.01	0.58	0.46	2
CHP_H_MAX	0.79	0.87	0.73	0.49	0.86	0.85	0.83	0.64	0.05	–0.20	0.88	0.90	0.91	0.54	–0.02	0.92	0.74	2
CHP_H_MAX2	0.78	0.89	0.73	0.43	0.85	0.86	0.82	0.64	0.00	–0.24	0.86	0.86	0.90	0.50	–0.11	0.91	0.79	2
CHP_H_MIN	0.64	0.69	0.53	0.55	0.56	0.48	0.44	0.62	0.22	–0.18	0.57	0.59	0.62	0.46	–0.06	0.63	0.48	2
HGT55	0.51	0.66	0.52	0.21	0.52	0.62	0.59	0.58	–0.13	–0.21	0.61	0.52	0.61	0.21	–0.20	0.64	0.74	2
Canopy Height Profile Indices																		
COVER_X	0.37	0.35	0.32	0.38	0.43	0.33	0.43	0.30	0.10	–0.10	0.49	0.48	0.49	0.32	0.20	0.48	0.18	2
CHP_MN_X	0.75	0.85	0.64	0.55	0.71	0.73	0.68	0.72	0.21	–0.19	0.80	0.80	0.80	0.49	–0.10	0.85	0.72	2
CHP_MN_SD	0.65	0.74	0.61	0.35	0.71	0.79	0.71	0.50	0.00	–0.20	0.72	0.77	0.77	0.40	–0.09	0.78	0.70	1
CHP_Q_X	0.76	0.86	0.65	0.57	0.71	0.74	0.67	0.70	0.19	–0.18	0.79	0.80	0.80	0.52	–0.09	0.84	0.73	1
CHP_Q_X2	0.70	0.83	0.62	0.42	0.66	0.72	0.68	0.73	0.11	–0.21	0.78	0.73	0.77	0.39	–0.17	0.81	0.75	3
CHP_Q_SD	0.72	0.80	0.70	0.40	0.76	0.85	0.77	0.45	–0.06	–0.16	0.77	0.80	0.77	0.49	–0.03	0.81	0.78	2
MNH_COV	0.78	0.86	0.66	0.58	0.71	0.70	0.66	0.69	0.22	–0.18	0.79	0.79	0.80	0.53	–0.07	0.83	0.70	0
QMCH_COV	0.79	0.87	0.68	0.59	0.72	0.72	0.65	0.69	0.20	–0.18	0.78	0.79	0.80	0.56	–0.08	0.83	0.70	1
Canopy Transmittance Indices																		
TRANS_MN_X	0.81	0.88	0.77	0.57	0.83	0.83	0.79	0.59	–0.03	–0.11	0.84	0.88	0.87	0.61	0.02	0.88	0.77	6
TRANS_MN_SD	0.60	0.72	0.59	0.25	0.78	0.81	0.85	0.49	–0.07	–0.19	0.82	0.79	0.85	0.30	–0.07	0.83	0.73	3
TRANS_P50_X	0.78	0.85	0.67	0.50	0.75	0.72	0.71	0.68	0.18	–0.21	0.82	0.80	0.83	0.50	–0.07	0.85	0.66	1
TRANS_P50_SD	0.66	0.77	0.67	0.32	0.79	0.86	0.85	0.60	–0.12	–0.19	0.86	0.82	0.84	0.36	–0.09	0.87	0.76	1
TRANS_P98_X	0.81	0.90	0.74	0.54	0.84	0.84	0.81	0.67	0.06	–0.19	0.89	0.89	0.91	0.54	–0.04	0.93	0.77	0
TRANS_P98_SD	0.37	0.43	0.41	0.06	0.63	0.65	0.70	0.21		–0.14	0.62	0.60	0.62	0.17	0.01	0.61	0.48	4
Canopy Volume Indices																		
OPEN	0.23	0.23	0.22	0.03	0.38	0.34	0.37	0.15	–0.02	–0.10	0.30	0.34	0.35	0.16	0.05	0.31	0.18	0
CLOSED	0.65	0.78	0.62	0.24	0.76	0.80	0.84	0.60	–0.05	–0.25	0.85	0.78	0.84	0.28	–0.13	0.86	0.75	3
EUPHOTIC	0.63	0.69	0.59	0.59	0.61	0.64	0.53	0.51	0.00	–0.04	0.62	0.71	0.65	0.56	0.05	0.68	0.58	2
OLIGO	0.77	0.79	0.64	0.61	0.69	0.60	0.55	0.54	0.25	–0.15	0.68	0.71	0.71	0.61	0.02	0.73	0.55	3
FILLED	0.78	0.82	0.69	0.67	0.72	0.69	0.60	0.58	0.13	–0.11	0.72	0.79	0.76	0.65	0.04	0.78	0.63	3
LCOMP	0.66	0.73	0.63	0.38	0.68	0.71	0.65	0.45	–0.02	–0.17	0.69	0.71	0.68	0.48	–0.07	0.72	0.66	0
Adjusted R^2	0.76	0.92	0.67	0.57	0.82	0.84	0.82	0.52	0.61	0.05	0.84	0.85	0.87	0.5	0	0.9	0.76	

Shaded boxes indicate variables selected in a stepwise regression of stand structure indices from canopy structure indices. See Appendix I, Lefsky et al. (2005) for definition of canopy structure indices.

3.2.3. Stepwise multiple regression using the canonical variables (SCV)

Abridged results from the stepwise multiple regressions of stand structure variables on the seven lidar-derived canopy structure canonical variables are presented in Table 5. Grey values indicate those independent canonical variables which were considered statistically significant in the stepwise multiple regression prediction of each dependent canopy variable. Canonical variables 1 and 2 were used in 15 and 14 (respectively) of the equations predicting 17 stand structure indices. Plotting the correlation between the stand structure variables and each of the first two canonical variables (Fig. 1) created an ordination diagram indicating which variables were more closely related to either stand height or cover.

Examination of this diagram indicated that there were four clusters of variables. The first cluster consisted of those variables that were correlated with canonical variable 1 (height), but not with canonical variable 2 (cover). These included the mean DBH of all stems and the number of stems greater that 100 cm in diameter. The second cluster had high correlation with the first canonical variable and moderate correlations with the second canonical variable. This cluster included a number of DBH-related indices (mean DBH, mean dominant and co-dominant DBH, standard deviation of DBH, max DBH, and the number of stems greater than 100 cm), as well as aboveground biomass, total and coniferous basal area, and Lorey's and maximum height.

Table 4
Pearson correlations between canopy structure variables and each stand structure canonical variable (CV)

	BASAL_CV	BIOMASS_CV	CONIF_BA_CV	COVER_CV	DBHMAX_CV	DBHSTD_CV	DBHU_CV	DBHX_CV	DECID_BA_CV	DENSITY_CV	HTDCD_CV	HTMAX_CV	HTMAXM_CV	LAI_CV	LN_DENSITY_CV	LOREY_CV	NT100CM_CV	Equations using this variable
Canopy Surface Height Indices																		
CHP_H_X	0.89	0.95	0.82	0.66	0.89	0.88	0.85	0.78	0.07	–0.28	0.93	0.94	0.94	0.66	–0.06	0.95	0.85	4
CHP_H_X2	0.86	0.96	0.81	0.54	0.86	0.88	0.84	0.79	–0.01	–0.32	0.91	0.88	0.91	0.57	–0.19	0.93	0.91	9
CHP_H_M	0.88	0.94	0.82	0.64	0.89	0.90	0.87	0.77	0.05	–0.28	0.94	0.94	0.94	0.64	–0.05	0.96	0.85	8
CHP_H_M2	0.84	0.95	0.81	0.51	0.85	0.90	0.86	0.77	–0.03	–0.33	0.92	0.88	0.91	0.54	–0.19	0.93	0.93	7
CHP_H_SD	0.36	0.43	0.41	0.04	0.62	0.65	0.71	0.23	–0.19	–0.20	0.62	0.60	0.62	0.17	0.01	0.59	0.50	8
CHP_H_MAX	0.86	0.91	0.80	0.60	0.91	0.89	0.87	0.75	0.06	–0.29	0.92	0.94	0.95	0.64	–0.03	0.94	0.81	5
CHP_H_MAX2	0.84	0.93	0.80	0.52	0.90	0.90	0.87	0.75	0.00	–0.34	0.91	0.91	0.93	0.59	–0.16	0.93	0.87	8
CHP_H_MIN	0.70	0.72	0.58	0.67	0.59	0.51	0.46	0.72	0.26	–0.25	0.60	0.62	0.65	0.55	–0.09	0.65	0.52	8
HGT55	0.55	0.69	0.57	0.26	0.56	0.65	0.62	0.69	–0.15	–0.30	0.64	0.55	0.64	0.25	–0.29	0.66	0.80	9
Canopy Height Profile Indices																		
COVER_X	0.41	0.37	0.35	0.47	0.46	0.35	0.45	0.35	0.12	–0.15	0.52	0.51	0.51	0.38	0.28	0.49	0.19	11
CHP_MN_X	0.82	0.89	0.70	0.67	0.76	0.77	0.72	0.84	0.25	–0.27	0.84	0.84	0.84	0.59	–0.13	0.87	0.79	10
CHP_MN_SD	0.71	0.77	0.67	0.43	0.75	0.83	0.75	0.59	0.00	–0.29	0.76	0.81	0.80	0.48	–0.12	0.80	0.77	14
CHP_Q_X	0.83	0.90	0.72	0.69	0.76	0.78	0.71	0.82	0.22	–0.26	0.83	0.84	0.83	0.62	–0.13	0.86	0.79	7
CHP_Q_X2	0.76	0.87	0.68	0.51	0.70	0.76	0.72	0.86	0.13	–0.30	0.82	0.77	0.80	0.46	–0.25	0.84	0.82	11
CHP_Q_SD	0.78	0.84	0.76	0.49	0.80	0.89	0.81	0.52	–0.08	–0.22	0.81	0.84	0.81	0.59	–0.04	0.84	0.86	13
MNH_COV	0.85	0.90	0.72	0.70	0.76	0.74	0.70	0.81	0.27	–0.26	0.83	0.83	0.83	0.64	–0.09	0.86	0.76	10
QMCH_COV	0.86	0.91	0.74	0.71	0.76	0.76	0.69	0.80	0.24	–0.26	0.82	0.83	0.83	0.67	–0.11	0.86	0.77	11
Canopy Transmittance Indices																		
TRANS_MN_X	0.88	0.92	0.84	0.69	0.88	0.88	0.83	0.69	–0.03	–0.15	0.89	0.93	0.90	0.73	0.03	0.91	0.85	17
TRANS_MN_SD	0.65	0.75	0.65	0.30	0.83	0.85	0.90	0.57	–0.08	–0.27	0.87	0.83	0.88	0.36	–0.09	0.85	0.80	12
TRANS_P50_X	0.85	0.89	0.74	0.60	0.80	0.76	0.75	0.79	0.21	–0.31	0.86	0.84	0.87	0.59	–0.09	0.87	0.72	8
TRANS_P50_SD	0.72	0.81	0.73	0.38	0.84	0.90	0.90	0.71	–0.14	–0.28	0.91	0.87	0.87	0.43	–0.13	0.90	0.83	10
TRANS_P98_X	0.88	0.94	0.81	0.65	0.89	0.88	0.86	0.78	0.07	–0.28	0.94	0.94	0.95	0.65	–0.06	0.95	0.84	7
TRANS_P98_SD	0.40	0.45	0.45	0.07	0.67	0.69	0.74	0.25	–0.21	–0.20	0.65	0.63	0.65	0.20	0.02	0.62	0.52	8
Canopy Volume Indices																		
OPEN	0.25	0.24	0.25	0.03	0.41	0.36	0.39	0.18	–0.02	–0.14	0.32	0.36	0.36	0.19	0.08	0.32	0.19	7
CLOSED	0.70	0.81	0.68	0.29	0.81	0.84	0.89	0.70	–0.05	–0.36	0.90	0.82	0.88	0.34	–0.18	0.88	0.82	9
EUPHOTIC	0.69	0.72	0.65	0.72	0.65	0.67	0.57	0.60	0.00	–0.06	0.66	0.75	0.68	0.67	0.07	0.70	0.64	9
OLIGO	0.84	0.83	0.70	0.74	0.73	0.63	0.58	0.63	0.30	–0.22	0.71	0.75	0.74	0.73	0.03	0.75	0.60	6
FILLED	0.85	0.86	0.75	0.81	0.76	0.73	0.64	0.68	0.16	–0.15	0.76	0.83	0.79	0.78	0.05	0.81	0.69	6
LCOMP	0.72	0.76	0.69	0.46	0.72	0.75	0.69	0.53	–0.02	–0.24	0.72	0.75	0.71	0.58	–0.10	0.74	0.72	9
Adjusted R^2	0.81	0.90	0.80	0.61	0.85	0.87	0.86	0.65	0.64	0.32	0.87	0.88	0.90	0.64	0.34	0.93	0.79	

Shaded boxes indicate variables selected in a stepwise regression of stand structure indices from canopy structure indices. See Appendix I, Lefsky et al. (2005) for definition of canopy structure indices.

The third cluster encompassed density, cover and LAI, which were most highly correlated with canonical variable 2, which indicates high values of lidar-measured variables related to cover. Of these variables, LAI has the higher correlation with the first canonical variable (related to height), which indicates that LAI is dependent on both cover and canopy height, as indicated in Lefsky et al. (1999a). The fourth and final cluster consists of one variable, the basal area of deciduous species, which had very low correlation to the first two canonical variables, but has a high correlation to the third canonical variable, as explained earlier and in Lefsky et al. (in Review).

3.2.4. Addition of environmental variables to regressions

Environmental factors were analyzed by creating two datasets of residuals; one from each of the estimates of stand structure (i.e. one from the direct CCA and a second from the SCV). CCA was then performed on each data set with the residuals as the dependent values and environmental data as the independent variables. In both cases, two pairs of canonical variables were found to be significant. Table 6a indicates which environmental variables contributed significantly to each of the environmental canonical variables. For the SCV, the first environmental variable was strongly related to the continentality of precipitation, the annual, minimum, summer, and winter temperature variables, the

Table 5
Correlations between the stand structure and the canonical variables of the lidar indices. Grey boxes indicate variables picked in stepwise regression of canopy structure predicted from lidar canonical variables

	LI1	LI2	LI3	LI4	LI5	LI6	LI7	Adjusted R^2
BASAL	0.78	0.24	–0.08	0.08	0.29	–0.24	–0.01	0.88
BSC	0.90	0.07	0.02	0.20	0.15	–0.16	0.03	0.92
CONIF_BA	0.74	0.24	0.17	0.02	0.37	–0.17	–0.01	0.88
DECID_BA	–0.02	–0.02	–0.71	0.17	–0.27	–0.16	0.00	0.5
DBHMAX	0.87	0.18	0.06	–0.12	0.00	–0.23	–0.01	0.81
DBHSTD	0.88	0.11	0.25	0.00	–0.05	–0.10	–0.04	0.88
DBHU	0.88	0.15	0.20	–0.15	–0.07	–0.04	–0.01	0.84
DBHX	0.66	–0.04	–0.27	0.15	0.03	0.30	0.10	0.65
DENSITY	–0.27	0.33	0.06	–0.09	0.38	–0.14	0.03	0.46
LNDENSITY	–0.09	0.50	–0.06	–0.12	0.03	–0.03	–0.11	0.41
LAI	0.43	0.50	–0.05	0.19	0.34	–0.21	0.01	0.81
COVER	0.39	0.47	–0.29	0.32	0.00	–0.14	0.01	0.67
HTDCD	0.91	0.18	0.09	0.02	–0.04	0.03	–0.07	0.89
HTMAX	0.88	0.32	0.04	0.07	–0.09	–0.06	0.00	0.89
HTMAXM	0.92	0.19	–0.02	0.00	–0.05	–0.02	0.06	0.91
LOREY	0.95	0.16	0.04	0.08	–0.08	–0.01	–0.03	0.96
NT100CM	0.82	–0.08	0.24	0.17	0.04	–0.16	0.15	0.74
Tally	15	14	7	9	8	13	9	

variability of precipitation and temperature, and the depth of soil, while the second environmental variable was weakly related to increasing elevation and related effects. For the direct CCA analysis, the first two environmental variables partitioned the variance explained by the SCV's first environmental variable. In this case, the first canonical variable was related to continentality and variability of precipitation, while the second canonical variable was related to temperature and elevation.

A number of stand structure indices had high correlations with the environmental variables (Table 6b). However, running the SCV regression analysis with the environmental variables showed (Table 7) that they made a meaningful improvement in just 1 equation: stem density (CV-2). In the direct CCA analysis, environmental variables were added to just two equations (stem density and LAI) and were a substantial improvement over direct CCA with canopy structure variables alone.

3.2.5. Regression method comparison: All sites

Table 7 compares the estimates produced by each of five regression methodologies: direct stepwise multiple regression (direct stepwise), direct canonical correlation analysis (direct CCA, with and without environmental variables), and stepwise multiple regression using canonical variables (SCV, with and without environmental variables). Equations were evaluated on the basis of two key statistics: the adjusted R^2 and the ratio of the root mean square error to the mean predicted value for each dependent value. For each statistic, the value obtained using direct stepwise multiple regression was treated as a reference value and results for the other analyses were presented in terms of their improvement over the reference value (improvement in R^2 is positive; for RSME lower values are desirable, so improvement in RMSE is negative). In most cases, these improvements were marginal when compared to the direct stepwise result; the average improvement in R^2 was just 0.06 and 0.09 for direct CCA estimation with or without environment, and 0.11 for SCV with or without environment. The average improvements in the ratio of RMSE to the mean predicted value were similar, except that the direct CCA method with environment was an improvement over direct CCA with canopy structure variables alone, and slightly better than either SCV result. However, these values were to some extent due to lower correlations and higher RMSE values for variables such as LAI and density. If the 9 equations where the stepwise regression explained less than 80% of variance were excluded, then the improvements in R^2 decreased to 0.0 and 0.01 for

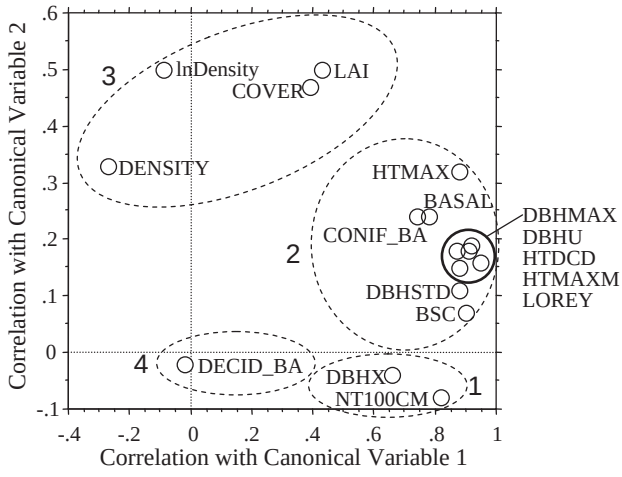

Fig. 1. Ordination diagram of stand structure variables plotted as a function of their correlation with canonical variables 1 and 2. Numbers identify clusters described in text.

Table 6
Correlations between canonical environmental factors 1 and 2 derived from both stepwise with direct CCA and SCV methods, with both (a) environmental and (b) stand structure indices

	SCV Environment 1	SCV Environment 2	Direct CCA Environment 1	Direct CCA Environment 2
(a)				
XX	−0.57	0.23	−0.28	0.38
YX	−0.53	0.13	−0.16	0.32
ANNPRE	0.34	−0.22	0.09	0.02
WTRPRE	0.43	**−0.26**	0.15	−0.06
SMRPRE	0.03	−0.04	−0.11	0.30
CVPRE	**0.79**	**−0.32**	**0.41**	−0.44
CONTPRE	**−0.79**	**0.36**	**−0.47**	0.41
ANNTMP	0.71	−0.23	0.23	**−0.50**
MAXTMP	0.35	0.02	0.09	−0.29
MINTMP	0.63	−0.25	0.19	−0.40
WTRTMP	0.65	−0.25	0.21	**−0.46**
SMRTMP	**0.73**	−0.16	0.22	**−0.51**
CVTMP	−0.49	0.24	−0.15	0.32
AWC	0.24	−0.04	**0.30**	0.23
DEPTH	**−0.85**	0.16	**−0.36**	0.32
SLOPE	−0.09	0.18	−0.11	0.32
ASPECT	−0.09	0.09	0.16	−0.10
ELEVATION	−0.58	**0.35**	−0.10	**0.51**
(b)				
BSC	−0.01	−0.05	0.25	0.14
BASAL	−0.04	0.07	0.16	−0.05
LAI	0.10	−0.03	**0.37**	0.16
HTMAXM	0.00	0.07	**0.30**	**0.23**
LOREY	0.11	0.18	0.06	**0.24**
NT100CM	**0.25**	**0.26**	**0.38**	0.09
DBHU	0.06	**0.33**	0.11	0.12
DBHSTD	0.14	0.23	0.15	−0.02
HTMAX	−0.04	**0.33**	0.03	0.14
HTDCD	0.04	0.24	0.12	**0.17**
COVER	0.00	0.03	0.07	**0.26**
DENSITY	**−0.26**	**−0.33**	**−0.40**	−0.08
LNDENSIT	−0.05	0.15	−0.14	0.03
DBHMAX	−0.02	0.20	−0.15	0.12
DBHX	**0.16**	0.08	0.08	0.08
DECID_BA	**0.32**	−0.21	−0.16	−0.16
CONIF_BA	0.00	0.18	0.26	−0.02

These factors are based on residuals from the CCA and SCV predictions of the stand structure indices. Bold numbers indicate top 25% correlations of stand structure indices with each canonical variable. Environmental variables are defined in Appendix I, Lefsky et al. (2005).

direct CCA (with and without environmental variables respectively) and 0.01 and 0.03 for analyses using SCV. In terms of this analysis of the entire dataset, both direct CCA estimation and SCV showed marginal improvement over stepwise multiple regression, and their results were very similar to each other.

3.2.6. Regression method comparison: Individual study areas

In addition to analysis of the equations predicting stand structure across the entire study region, analysis of the properties of the equations for each study area was also necessary. However with 5 areas, 4 methods, and 17 variables, describing all 340 of these equations would be prohibitively time-consuming. Therefore only summary statistics will be discussed. Table 8 provides average values for 9 statistics describing the performance of the methods in the five study areas.

To evaluate the applicability of the regression equations created using data from all study areas to each individual study area separately, regressions (with associated slopes and intercepts) for each area were calculated between the predictions from the overall regressions versus the observed values for each area. These results indicated that the direct CCA approach led to a smaller average deviation from the ideal value (1.0), but that a larger proportion of these site slopes significantly differed from zero. Further examination of equation slopes as a function of method and area indicated that most of the variability in equation slope (and the proportion of slopes significantly different

Table 7
Comparison of estimation methods

	Stepwise Adj. R^2	Change in R^2 with direct CCA estimation	Change in R^2 with direct CCA estimation with climate	Change in R^2 with SCV	Change in R^2 with SCVE*	Stepwise RMPV	Change in RMPV with direct CCA estimation	Change in RMPV direct CCA estimation with climate	Change in RMPV with SCV	Change in RMPV with SCVE*
BASAL	0.76	0.05	0.05	0.12	0.12	0.30	0.01	−0.07	−0.10	0.10
BIOMASS	0.92	−0.03	−0.02	0	0	0.26	0.01	−0.05	−0.07	0.07
CONIF_BA	0.67	0.14	0.16	0.21	0.21	0.40	0.02	−0.14	−0.16	0.16
COVER	0.57	0.04	0.1	0.10	0.10	0.22	0.02	−0.04	−0.03	0.03
DBHMAX	0.82	−0.01	0.01	−0.01	−0.01	0.23	0.01	−0.02	0.01	−0.01
DBHSTD	0.84	0.01	0.02	0.04	0.04	0.24	0.00	−0.03	−0.02	0.03
DBHU	0.82	0.03	0.04	0.02	0.03	0.24	0.00	−0.04	−0.01	0.02
DBHX	0.52	0.08	0.1	0.13	0.13	0.47	0.06	0.88	1.15	−1.15
DECID_BA	0.61	−0.06	−0.02	−0.11	−0.11	1.39	0.17	−0.46	−0.41	0.47
DENSITY	0.05	0.27	0.42	0.41	0.48	1.32	0.00	−1.16	−1.16	1.17
HTDCD	0.84	0.01	0.02	0.05	0.06	0.19	0.00	−0.03	−0.05	0.06
HTMAX	0.85	−0.04	−0.03	0.04	0.04	0.16	0.01	−0.03	−0.03	0.03
HTMAXM	0.87	0.01	0.03	0.04	0.04	0.17	0.03	0.10	0.07	−0.07
LAI	0.5	0.15	0.21	0.31	0.31	0.38	0.00	−0.23	−0.22	0.22
LNDENSITY	0	0.34	0.37	0.41	0.41		0.00	0.00	0.00	0.00
LOREY	0.9	0	0.01	0.06	0.06	0.14	0.01	−0.02	−0.05	0.05
NT100CM	0.76	0.03	0.07	−0.02	0.01	0.68	0.05	−0.16	0.01	0.03
Average	0.66	0.06	0.09	0.11	0.11	0.43	0.02	−0.09	−0.06	0.07
Average for R^2 values>0.8	0.86	0.00	0.01	0.03	0.03	0.21	0.01	−0.01	−0.02	0.02

SCVE is SCV with Environmental variables; RMPV is RMSE/MPV. MPV is Mean Predicted Value. R^2 values are bootstrapped.

from 1) is due to the Mt. Rainier site. Site intercepts (same as above but with the intercept of predicted vs. observed) follow a similar pattern, with the direct CCA approach having an average intercept closer to zero at all areas, but with the SCV resulting in intercepts that are up to 7% of the mean predicted value. However, the methods had roughly the same number of equations with intercepts that are significantly different from 0, indicating the larger intercepts were still within the 95 percent confidence intervals.

Table 8
Summary statistics for equation performance at individual sites

	Individual location slopes	Fraction of location slopes significantly different from one	Ratio of individual location intercepts to mean value	Fraction of location intercepts significantly different from one	$R2$_ratio	Ratio of variance of predictions and observations	RMSE as a percentage of Mean Predicted Value	Average site RMSE as fraction of overall RMSE	Ratio of bias to Mean Predicted Value
Stepwise	1.07	0.27	0.01	0.33	0.85	1.06	28.60%	0.980	−3.30%
Direct CCA	1.03	0.17	−0.02	0.23	0.88	1.02	24.12%	1.037	−1.48%
Direct CCA w/ climate	1.03	0.20	−0.02	0.25	0.88	1.03	22.95%	1.046	−1.18%
Swise w/ Canonical V.	1.06	0.13	−0.07	0.17	0.88	0.92	22.79%	1.030	0.72%
Swise w/ CV and climate	1.05	0.13	−0.05	0.19	0.87	0.92	22.36%	1.033	0.83%
Final	1.04	0.30	−0.30	0.09	0.89	0.96	22.00%	1.038	0.10

Notes:
1. Individual location slopes: The average slope of the regressions between values predicted by all−location equation and observed values for each location should be equal to 1.
2. Fraction of location slopes significantly different from one: The fraction of the slopes (for the regression between values predicted by all−location equation and observed values for each location) that are significantly different from one should be equal or close to zero.
3. Ratio of location intercepts to mean value: The intercepts of the regression between values predicted by all−location equation and observed values, when divided by the mean predicted value, should be equal or close to zero.
4. Fraction of location intercepts significantly different from one: The fraction of the intercepts (for the regression between values predicted by all−location equation and observed values for each site) that are significantly different from zero should equal or close to zero.
5. R^2_ratio: The average ratio of each location's bootstrapped r^2 values to overall bootstrapped r^2 value.
6. Four variables were excluded from this analysis, due to their low R^2 values: Density, LnDensity, NT100CM, and DECID_BA.

The R^2 ratio indicates the average ratio between area R^2 values and the overall equation's R^2 value; ideally it would be 1. When averaged over all areas, values for this statistic ranged between 0.85 and 0.88, indicating reasonable performance for all methods. Examination of area and method differences in this statistic demonstrated that there were area differences in the R^2 ratio. Specifically, HJA had an average R^2 ratio of 1.195, while values for Cascade Head (0.95), the Coastal Plots (0.85), and Rainier (0.9) approximated the overall value of 0.85, and Metolius was uniformly the lowest (0.65). While this may reflect differences in the intrinsic strength of the relationship between canopy and stand structure variables, it also reflected differences in the characteristics of these particular datasets, particularly the number of plots and range of values recorded in each set of stand variables.

Examination of each area's RMSE as a percentage of the mean predicted value indicated that three techniques (Direct CCA with environment, and SCV with and without environment) all have average values within less than 23% of the mean predicted value; the other two methods showed less precision in their predictions. Examination of this variable as a function of area and method indicated that one site (Cascade Head) had the lowest value for this variable (15%), while three sites (Coast Range, H.J. Andrews, and Rainier) had values approaching 25%. The Metolius area showed both area and method effects; on average the methods had values averaging 27% but the direct CCA methods had much higher values than those obtained with SCV, approaching 35% of the mean predicted values. The lower values at Cascade Head were due to the low number of younger stands found at that site, which increased the mean predicted value of most stand structure indices, and lowered the ratio of RSME to mean predicted value. The opposite effect may have occurred at Metolius—where stands were shorter and therefore the mean predicted values were relatively small. If the RMSE remains constant, then the RMSE/average predicted ratio will be large. This effect also applied to measurement error; a 1 m error in the lidar or field estimate of stand height would have a larger effect here than at Cascade Head.

Variance ratio and percent bias are both indicators of the ability of each method to preserve key qualities: the total variance and absolute values of the observed datasets. The method averages for these variables indicate that method had the most influence on these parameters. The direct CCA method resulted in higher variance ratios and negative biases, while the SCV methods resulted in lower variance ratios and positive biases. While the actual values involved indicate relatively small deviations from ideal values, it is instructive that each method has these distinctive patterns. Examination of these variables indicates that a combination of area and method effects is at work. A single area (Metolius) was the largest contributor of bias, with both SCV methods resulting in positive bias of 7.5% from the mean values, while the direct CCA methods resulted in a negative bias of -6%. Method is another contributor of variance; direct CCA results in

Table 9
Final models

Variable	Method	Bootstrapped Adj R^2	Number of sites with slopes significantly different from one	Number of sites with intercepts significantly different from zero	RMSE	RMSE (%)	Bias (%)	Ratio of variances	Mean Predicted Value	Sites with slopes or intercepts Significantly different from zero
BIOMASS	SCV	0.92	0	0	89.04	19.55	0.00	0.9626	455.56	
DBHMAX	Direct CCA w/Environment	0.83	0	0	20.99	21.33	0.15	1.0004	98.42	
DBHSTD	Direct CCA	0.85	0	0	4.41	21.55	−0.01	1.0005	20.45	
DBHU	SCV	0.84	0	0	12.34	22.56	0.00	0.9234	54.69	
HTDCD	SCV	0.89	0	0	4.75	15.89	0.00	0.9465	29.87	
HTMAX	SCV	0.89	0	0	6.47	14.08	0.00	0.9435	45.94	
LAI	SCV	0.81	0	0	1.53	23.76	0.00	0.8932	6.45	
NT100CM	Direct CCA	0.79	0	0	7.45	57.18	0.06	0.9998	13.03	
BASAL	Direct CCA	0.81	0	1	13.85	23.96	−0.62	0.9978	57.80	Metolius
CONIF_BA	Direct CCA	0.81	0	1	14.75	28.18	−0.79	0.9989	52.34	Metolius
DBHX	SCV	0.65	1	0	10.61	39.81	0.00	0.8181	26.65	Rainier
DECID_BA	Direct CCA	0.55	2	0	7.65	141.67	1.36	0.9797	5.40	HJA,Metolius
DENSITY	SCV w/Environment	0.53	1	1	996.97	91.49	0.00	0.7415	1089.67	Coast Range
LNDENSITY	SCV	0.41	1	1	1.06	16.61	0.00	0.6241	6.36	HJA,Metolius
LOREY	SCV	0.96	1	1	2.90	9.31	0.00	0.9799	31.16	Metolius
COVER	SCV	0.67	2	1	0.13	18.46	0.00	0.8226	0.71	Coast Range,HJA
HTMAXM	SCV	0.91	1	2	5.26	13.15	0.00	0.9572	40.02	Cascade Head

variance ratios that were ~1% above the results obtained using SCV at every area.

3.2.7. Equation selection

A comparison of the regression methods permits a few generalizations. First, both direct CCA and SCV slightly out-performed direct stepwise multiple regression in terms of variance explained and the R^2 ratio (Table 8). Large improvements were seen in RMSE as a percentage of mean predicted value and the fraction of equations with slopes or intercepts that significantly differed from 1 or 0, respectively. Secondly, both direct CCA and SCV performed similarly with respect to variance explained and root mean squared error. SCV outperformed direct CCA to a modest degree, especially with respect to slope, intercept, and bias.

In these tables, performance of each regression method was discussed with regard to their overall performance, and not the individual variables in question. However for final selection of regression equations, a different approach was used, specifically designed to pick the best equation for each variable. Two steps were involved. For each variable, the equation or equations with the least number of statistically significant site level deviations from the

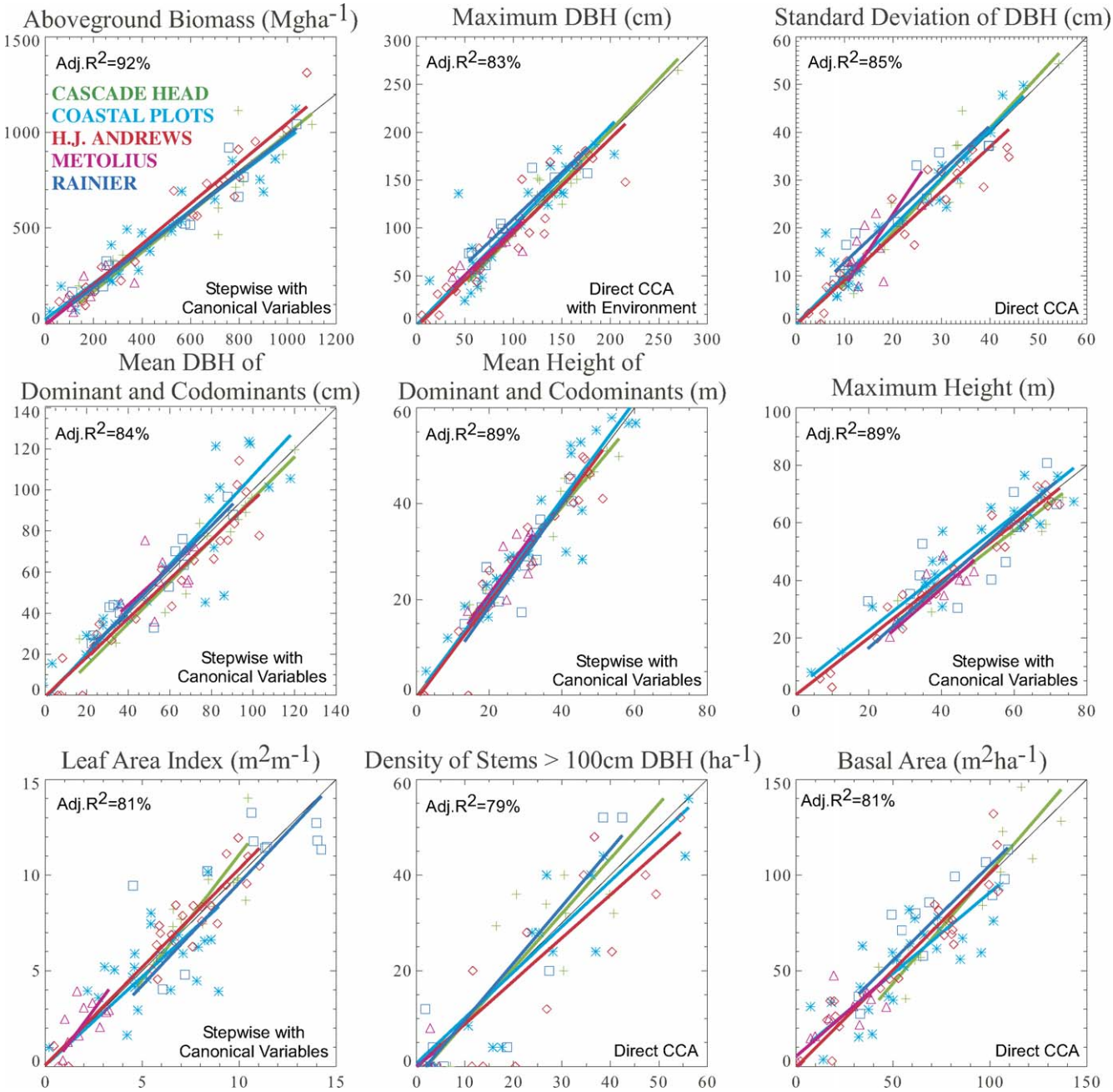

Fig. 2. Scatterplots of predicted vs. observed stand structure variables.

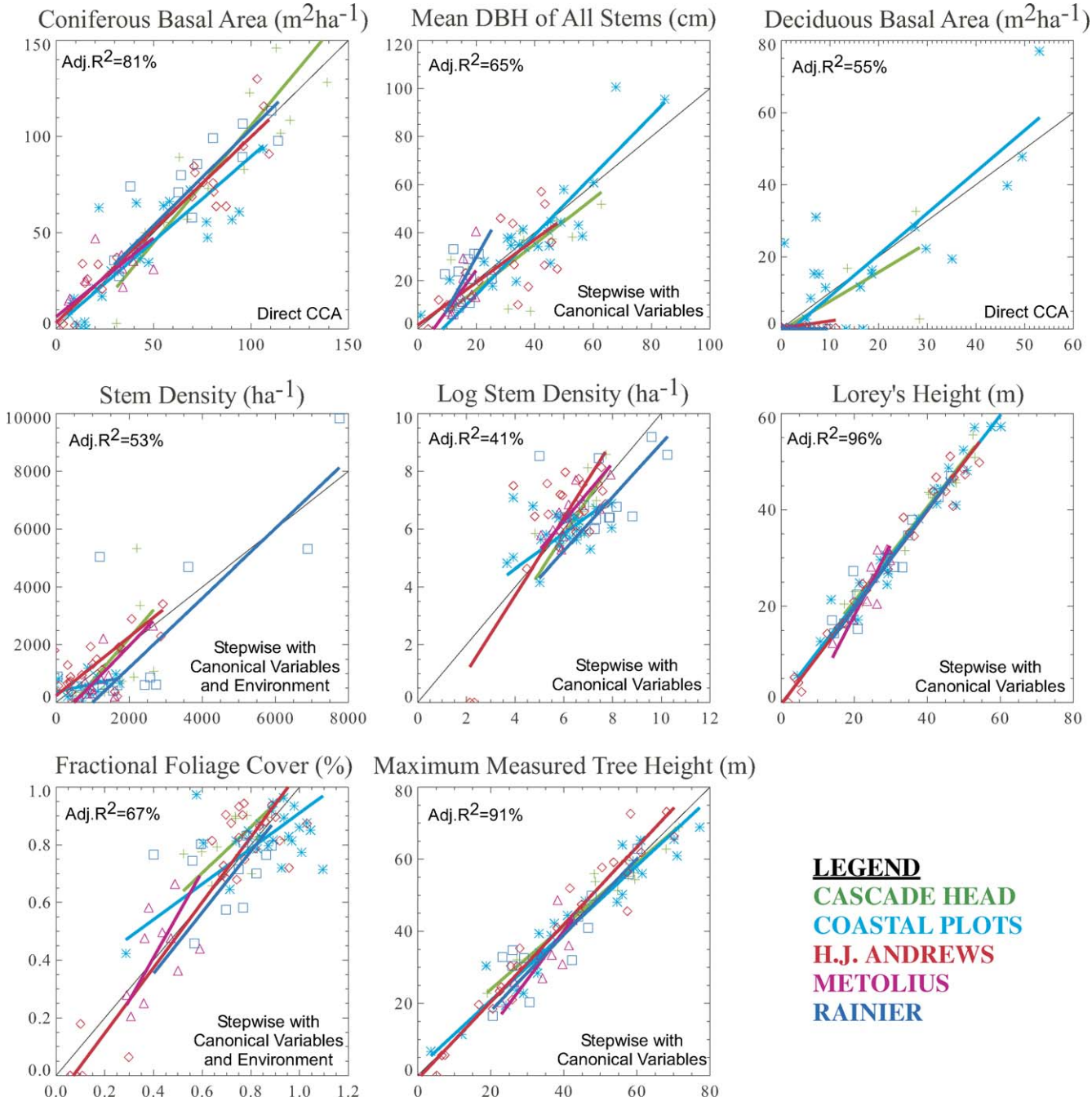

Fig. 2 (*continued*).

overall equation were selected. Second, adjusted R^2 values for these sites were compared, and the equation with the maximum R^2 was selected. If two equations had the same fractional R^2 (to 2 decimal places when expressed as a fraction), then equations without environmental data were preferred over those with environmental data. This rule allowed us to select an equation in each instance (Table 9, Fig. 2). In nine of seventeen cases, the equation was developed using SCV; five were developed with direct CCA, two were developed with SCV with environmental data, and 1 was developed with direct CCA with environmental data. Statistics by location for the combined dataset are presented in Table 9.

4. Discussion

4.1. Methods

The choice of methods to test for this study reflected a range of goals. Stepwise multiple regression was picked to represent the methods used in previous lidar papers (e.g.

Lefsky et al., 1999a,b, 2002). CCA using multiple independent variables and a single dependent variable is the method described in Cohen et al. (2003). In this work we were concerned with evaluating multivariate independent and dependent datasets, understanding the redundancy within each dataset, as well as estimating dependent variables. Combining these two goals, a third analysis approach (SCV), using stepwise multiple regression to predict dependent variables (i.e. stand structure) from canonical variables derived from the independent dataset (i.e. canopy structure), seemed to have a number of advantages. First, because the canonical variables were uncorrelated, the problem of using collinear variables in stepwise multiple regression was eliminated. The use of the canonical variables increased interpretability, as all collinear variables were considered in a single canonical variable and could be interpreted as contributing to the canonical variables in accordance with their correlation with them. This removed the problem encountered when multiple variables with similar correlation coefficients and F scores are available, but only one can be entered into a stepwise regression. This could easily lead to over-interpretation of the particular variables picked by the stepwise multiple regression, when other variables may be almost equally suitable.

It could be argued that, rather than using stepwise multiple regression to combine the canonical variables, a second CCA step could be used to combine the original canonical variables into estimates of the dependent variable—in essence performing the analysis of Cohen et al. (2003) using canonical variables in the place of the original variables. In this case, having created uncorrelated canonical variables, the capacity of stepwise regression to remove (from the regression equation) those variables that did not contribute to the estimation of the dependent variable was a decided advantage over the potentially ambiguous methods available to ascertain the contribution of independent variables used in CCA.

Summary statistics for the entire study area (Table 7) were useful for indicating the relative performance of different regression techniques, but masked the performance of each method at different sites (Table 8). Although the statistics reported in Table 8 give a reasonable representation of the performance of each method at individual sites, analysis of equation performance required a combination of graphical (box plot) and statistical (ANOVA) techniques, which proved cumbersome.

4.2. Variable selection

In contrast to the results of Lefsky et al. (1999a), in which canopy volume variables accounted for most of the independent variables picked in a stepwise regression, explanatory variables were evenly spread among the four methods for the description of forest canopies: canopy surface height measurements (each variable involved in an average of 1.89 equations), canopy height profile measurements (1.62), canopy transmittance indices (2.5), and canopy volume indices (1.83). This may be due to the inclusion of the canopy transmittance indices, which were not considered in Lefsky et al. (1999a), and which may capture much of the variance previously captured by the canopy volume indices.

4.3. Equation generality

Of the 17 stand structure variables considered in this paper, we were able to develop eight equations that were valid for all sites, including equations for two variables generally considered as highly important aboveground biomass and leaf area index. The other six equations that were valid for all sites were either related to height (which is most directly measured by lidar) or DBH (which should be closely related to height).

It is noteworthy that aboveground biomass and leaf area index were consistently predictable along a productivity and species composition gradient from the true fir forests of Mt. Rainier to Ponderosa pine forests at Metolius, and at the high productivity forests of Cascade Head, the coast range, and H.J. Andrews. This result offers a regional confirmation of the continental-scale hypothesis offered in Lefsky et al. (2002), in which the geographic generality of an equation predicting aboveground biomass was demonstrated. While the range of environmental conditions and composition examined in this paper is narrower than in Lefsky et al. (2002), the number of site locations examined is larger, and thus confirms the result for the Pacific Northwest region of the USA.

Of the nine equations that could not be generalized to all sites, four (basal area, conifer basal area, mean DBH of all stems, and Lorey's height) failed at either Metolius or Mt. Rainier, the sites having the most extreme differences in terms of composition, stand structure, and environmental conditions. Therefore, these variables were valid for the remaining locations, and probably for the stands of Douglas-fir/western hemlock and Sitka spruce/western hemlock in the Coast Range and western slopes of the Cascades in general. For these areas, we had a total of 12 equations that were applicable. In addition, the equation for deciduous basal area failed at H.J. Andrews and Metolius, two sites with deciduous basal area less than 1.0 m^2 ha^{-1}. It is possible that a more successful method for estimating deciduous basal area could be created using a combination of conventional optical remote sensing to detect the presence of deciduous trees (e.g. Maiersperger et al., 2001), and lidar to estimate their basal area.

One aid to investigating the potential generality of equations relating canopy and stand structure can be found in the forestry literature's site and yield tables. Site index tables relate stand age to the mean height of dominant and co-dominant trees—a standard index of productivity. Yield tables indicate the expected volume of a stand for a given stand age. Site index and yield tables both have high

variability due to productivity effects on the relationship of age to the height and yield variables. However, when mean height and yield are compared, the resulting relationships are free of most productivity and species effects (Fig. 3A and B), despite the fact that a relatively simple height index is being used. The fact that the yield variable is more consistent, as a function of productivity and composition, than either the mean diameter or basal area variables (also standard variables included in site index and yield studies, Fig. 3c and d), is consistent with the equations developed from our datasets. The site index and yield literature has a number of drawbacks including: only average and/or regressed values are reported for the variables of interest, the variables reported (e.g. yield) are not perfectly correlated with the variables of interest (e.g. aboveground biomass), and the low range of tree heights and yields. Nevertheless, they can be a useful tool to identify potential difficulties in developing equations for stands with varying composition and productivity, and have been an accurate indicator of generality in our experience.

4.4. Environmental variables

Of the 6 final equations (Table 9) that explained less than 80% of variance (and which may be expected to leave considerable room for model improvement), only density was substantially improved by adding the environmental canonical variables as part of the model. Stem density should be partially controlled by productivity, and therefore the inclusion of environmental variables makes sense. However, equations for the number of stems greater than 100 cm, and the mean DBH of all stems might also be expected to have a productivity component, but environmental variables did not have a substantial effect on these equations. For variables such as aboveground biomass and cover, it is reasonable to suggest that, given the high percentage of variance explained and direct physical relationship between dependent and independent variables, environmental variables would not contribute to the final equations. It is reasonable to hypothesize that there exist other direct linkages between the lidar measurements and many of these variables, which obviate the need for environmental effects.

Environmental effects may have been masked by the large range of stand structures included in this study. To check for this effect, the residuals for every variable were normalized separately by dividing them by the predicted value from the final model, and stepwise multiple regressions were run between environmental variables and the normalized residuals from each variable separately. Although there was a moderate chance of inflated results, predicting 17 variables from 18 independent variables with 86 cases, this analysis was meant to detect any possible environmental effects, and therefore this was considered an acceptable risk. Of the 17 dependent variables, in 10 cases the stepwise multiple regressions found no significant relationship, for an additional 6 cases the regressions

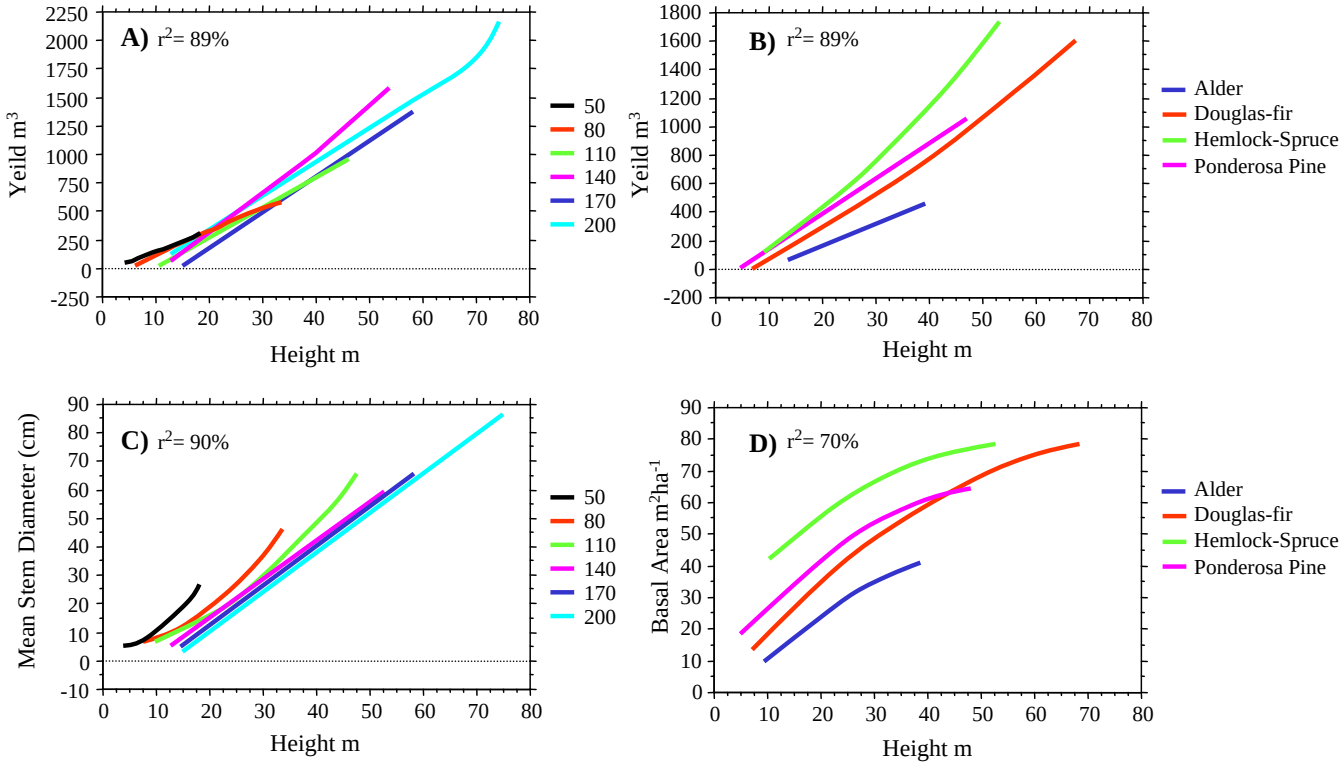

Fig. 3. Relationships between pairs of variables from site index and yield tables. Sources of data: Red Alder: Worthington et al. (1960), Douglas-fir: McArdle (1930), Western hemlock: Barnes (1962), Sitka Spruce And Western Hemlock: Meyer (1937), Ponderosa pine: Meyer (1961).

explained between 4% and 8%, and in one case the regression explained 15% of variance. Taking the last case, the variable being considered was maximum height (HTMAX) whose regression from the lidar canopy structure indices had already explained 89% of variance. Therefore, the 15% of variance in the normalized residuals could represent no more than 2.65% (0.15 * (1.0–0.89)) of the overall relationship between canopy structure, environment and the stand structure variable. This confirms the earlier conclusion that environment did not play a significant role in these relationships.

5. Conclusions

Previous studies have demonstrated the strong relationship between lidar measurements of canopy structure and indices of forest stand structure. Only one study (Lefsky et al., 2002) has attempted to test the generality of these relationships over multiple sites, and multiple forest cover types. Whereas Lefsky et al. (2002) were successful at developing a unified equation for predicting aboveground biomass in multiple biomes (Temperate Deciduous Broadleaf, Temperate Coniferous Needleleaf, Boreal Coniferous Needleleaf), there was no replication in each biome. In this work, we were able to look at 5 sites within the Temperate Coniferous Needleleaf biome, using sites with varying environment and composition.

We were able to create equations that predicted stand structure variables (e.g. aboveground biomass and LAI) across an environmental and compositional gradient that included open-canopy ponderosa pine on the east side of the Cascade range, Sitka spruce/western hemlock at Cascade Head, and true fir forests at Mt. Rainier. Given this wide range of conditions, and the earlier result of Lefsky et al. (2002) which included black spruce (*Picea mariana*), it is reasonable to ask if, in forests dominated by coniferous species, tree architecture is constrained to the point where a unified relationship between lidar measurements and stand structure might exist for these forests generally.

In existing studies of this type (including this one) there has been an attempt to have a structural or temporal sequence of stands at one or more study locations. This study found broad consistency in lidar-stand structure relationship over this region, and a relative lack of importance of environmental conditions. Therefore, it is likely that a modified sample design, in which plots with a range of structures or ages are distributed throughout a region or continent (without attempting to have complete sequences in every forest type), would be successful. In this type of study, analysis of residuals by forest type, and testing for the importance of environmental conditions would be used. Further analysis of this dataset will provide some guidelines for this type of study, which will be less plot intensive, and therefore, less expensive.

Acknowledgements

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.

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.*

Barnes, G. H. (1962). Yield of even-aged stands of western hemlock. *USDA Technical Bulletin*, *1273*. 52 pp.

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 Interational 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.

Carlson, T. N., & Ripley, D. A. (1997). On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sensing of Environment*, *62*, 241–252.

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.

Conrad, R., Gutmann, J. (1996). *Conversion equations between fork length and total length for Chinook Salmon (Oncorhynchus tshawytscha)*. Northwest Indian Fisheries Commission. Project Report Series No. 5. Olympia, WA. 32 pp.

Daly, C., Taylor, G., & Gibson, W. (1997). The PRISM approach to mapping precipitation and temperature. *10th Conf. on Applied Climatology* (pp. 10–12). Reno, NV: American Meteorological Society.

Drake, J. B., Dubayah, R., Knox, R., & Clark, D. (2002). Relationship between vertical canopy profiles and biomass in a Neotropical rainforest. *Remote Sensing of Environment*, *81*, 378–392.

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

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.

Healy, M. J. R. (1984). The use of R^2 as a measure of goodness of fit. *J. Roy. Statist. Soc. Ser. A*, *147*, 608–609.

Lefsky, M. A., Cohen, W. B., Acker, S. A., Spies, T. A., Parker, G. G., & Harding, D. (1999a). 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., Harding, D., Cohen, W. B., & Parker, G. G. (1999b). 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., 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). Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. *Remote Sensing of Environment*, *95*, 517–531.

Maiersperger, T., Cohen, W. B., & Ganio, L. (2001). A TM-based hardwood-conifer mixture index for closed-canopy forests in the Oregon Coast Range. *International Journal of Remote Sensing*, *22*, 1053–1066.

McArdle, E. R. (1930). The yield of Douglas fir in the Pacific Northwest. *USDA Technical Bulletin*, *201*. 64 pp.

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.

Meyer, Walter H. (1937). *Yield of even-aged stands of Sitka spruce and western hemlock*. USDA, Forest Service. Pacific Northwest Forest Experiment Station Technical Bulletin 544.

Meyer, Walter H. 1961. *Yield of even-aged stands of ponderosa pine*. USDA Technical Bulletin 630 (revised 1961).

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.

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, 4th ed., vols. 1 –2, Cary, North Carolina, USA, 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.

Turner, D., Cohen, W., Kennedy, R., Fassnacht, K., & Briggs, J. (1999). Relationship between leaf area index and Landsat TM spectral vegetation indices across three temperate zone sites. *Remote Sensing of Environment*, *70*, 52–68.

U.S. Department of Agriculture (1994). *State soil geographic (STATSGO) data base–data use information, miscellaneous publication number 1492*. Fort Worth, Texas: Natural Resources Conservation Service.

Van Huffel, S. (Ed.). (1997). *Recent advances in total least squares techniques and errors-in-variables modeling*. Philadelphia: Society for Industrial and Applied Mathematics.

Waring, R. H., & Schlesinger, W. H. (1985). *Forest ecosystems: Concepts and management*. Orlando, FL: Academic Press.

Waring, R. H., Way, J., Hunt, E. R., Morrissey, L., Ranson, K. J., Weishampel, J. F., et al. (1995). Imaging radar for ecosystem studies. *Bioscience*, *45*, 715–723.

Worthington, N. P., Johnson, F. A., Staebler, G. R., Lloyd, W. J. (1960). *Normal yield tables for red alder*. USDA Forest Service, Pacific Northwest Forest and Range Experiment Station, Research Paper, PNW-RS-36, 29 pp.