Forest structure estimation and pattern exploration from discrete-return lidar in subalpine forests of the central Rockies

K.R. Sherrill, M.A. Lefsky, J.B. Bradford, and M.G. Ryan

Abstract: This study evaluates the relative ability of simple light detection and ranging (lidar) indices (i.e., mean and maximum heights) and statistically derived canonical correlation analysis (CCA) variables attained from discrete-return lidar to estimate forest structure and forest biomass variables for three temperate subalpine forest sites. Both lidar and CCA explanatory variables performed well with lidar models having slightly higher explained variance and lower root mean square error. Adjusted R^2 values were 0.93 and 0.93 for mean height, 0.74 and 0.73 for leaf area index, and 0.93 and 0.85 for all carbon in live biomass for the lidar and CCA explanatory regression models, respectively. The CCA results indicate that the primary source of variability in canopy structure is related to forest height, biomass, and total leaf area, and the second most important source of variability is related to the amount of midstory foliage and tree density. When stand age is graphed as a function of individual plot scores for canonicals one and two, there is a clear relationship with stand age and the development of stand structure. Lidar-derived biomass and related estimates developed in this work will be used to parameterize decision-support tools for analysis of carbon cycle impacts as part of the North American Carbon Program.

Résumé : Cette étude évalue la capacité relative d'indices simples obtenus avec le lidar (« light detection and ranging ») (c.-à-d. hauteurs moyenne et maximum) et de variables dérivées statistiquement de l'analyse de corrélation canonique (ACC) provenant d'impulsions discrètes du lidar d'estimer des variables de structure et de biomasse de la forêt pour trois stations de forêt subalpine tempérée. Tant les variables explicatives provenant du lidar que celles provenant de l'ACC ont produit de bons résultats mais les modèles basés sur le lidar expliquaient une portion légèrement plus grande de la variance et avaient une erreur quadratique moyenne plus faible. Les valeurs de R^2 ajustées pour la hauteur moyenne, l'indice de surface foliaire et tout le carbone dans la biomasse vivante étaient respectivement de 0,93 et 0,93, 0,74 et 0,73 et 0,93 et 0,85 pour les modèles basés sur le lidar et les modèles de régression explicative basés sur l'ACC. Les résultats de l'ACC indiquent que la principale source de variation dans la structure de la canopée est reliée à la hauteur, à la biomasse et à la surface foliaire totale de la forêt tandis que la deuxième plus importante source de variation est reliée à la quantité de feuillage dans la strate intermédiaire et à la densité des arbres. Lorsque l'âge du peuplement est mis en graphique en fonction des scores de chaque placette échantillon pour les ensembles canoniques un et deux, il y a une relation nette entre l'âge du peuplement et le développement de la structure du peuplement. La biomasse dérivée du lidar et les estimations qui y sont associées, que nos travaux ont permis de développer, seront utilisées comme paramètres dans les outils d'aide à la décision pour analyser les impacts du cycle du carbone dans le cadre du « North American Carbon Program ».

[Traduit par la Rédaction]

Introduction

North American forests have been estimated to have an annual carbon sink between 0.14 and 0.30 Pg C·year^{-1} (Pacala et al. 2001). Uncertainty of this magnitude is inconsistent with continental- to global-scale monitoring of carbon stocks and fluxes. Knowledge of aboveground carbon stocks from forest inventory and remote sensing provides a snapshot view of effects from land-use history and forest management and, through relationships with ecosystem function, a constraint on carbon flux (Cohen et al. 1996; Running et al. 1999). To address uncertainties in the effects of human alterations to forested systems, forest carbon managers are interested in combining spatially extensive but coarse-grained carbon estimates from remote sensing and forest inventory data with spatially intensive and finely resolved estimates from AmeriFlux flux tower carbon sites (Birdsey et al. 2004). By connecting these extensive- and intensive-scaled data sources, more accurate carbon budgets across management strategies and disturbance scenarios will be possible.

Although remotely sensed estimates of biomass in low- to moderate-cover forests (leaf area index (LAI) < 3) have been successful, biomass estimates in moderate- to high-

Received 25 October 2007. Accepted 9 April 2008. Published on the NRC Research Press Web site at cjfr.nrc.ca on 26 June 2008.

K.R. Sherrill[1] **and M.A. Lefsky.** Center for Ecological Applications of Lidar, Colorado State University, Fort Collins, CO 80523, USA.
J.B. Bradford. USDA Forest Service, Northern Research Station, Grand Rapids, MN 55744, USA.
M.G. Ryan. USDA Forest Service, Rocky Mountain Research Station, Fort Collins, CO 80526, USA.

[1]Corresponding author (e-mail: sherrill@cnr.colostate.edu).

 doi:10.1139/X08-059

Table 1. Lidar and field variables.

Variable	Description	Units
Lidar canopy surface height indices		
Maxh	Maximum stand height	m
Meanh	Mean stand height	m
QH	Quadratic stand height	m
Lidar canopy profile indices		
HP_{10}	Height at which 10% of points are below	m
HP_{90}	Height at which 90% of points are below	m
$CF_{2–4}$	Percentage of points in 2–4 m bin	%
$CF_{52–54}$	Percentage of points in 52–54 m bin	%
Stand structure		
Agemean	Mean tree or sapling age	years
Agemax	Maximum tree or sapling age	years
Height	Mean tree or sapling height	m
Allpla	Leaf area of trees, saplings and seedlings	$m^2 \cdot m^{-2}$
Treesha	Tree density	$Trees \cdot ha^{-1}$
Sapsha	Sapling density	$Saplings \cdot ha^{-1}$
Seedsha	Seedling density	$Seedlings \cdot ha^{-1}$
Batotal	Basal area of all trees and saplings	$m^2 \cdot m^{-2}$
Batree	Basal area of all trees	$m^2 \cdot m^{-2}$
BaLP	Basal area of lodgepole pine trees	$m^2 \cdot m^{-2}$
BaES	Basal area of Engleman spruce trees	$m^2 \cdot m^{-2}$
BaSF	Basal area of subalpine fir trees	$m^2 \cdot m^{-2}$
BaTA	Basal area of trembling aspen trees	$m^2 \cdot m^{-2}$
Balm	Basal area of limber pine trees	$m^2 \cdot m^{-2}$
BaSap	Basal area of all saplings	$m^2 \cdot m^{-2}$
Carbon pools		
Allcarbon	Total ecosystem carbon	Mg $C \cdot ha^{-1}$
Alllive	All carbon in live biomass	Mg $C \cdot ha^{-1}$
Abovelive	Carbon in aboveground live biomass	Mg $C \cdot ha^{-1}$
Treeaglive	Carbon in aboveground components of live trees	Mg $C \cdot ha^{-1}$
Saplive	Carbon in live saplings	Mg $C \cdot ha^{-1}$
Seedlive	Carbon in live seedlings	Mg $C \cdot ha^{-1}$
Veg	Carbon in understory vegetation	Mg $C \cdot ha^{-1}$
Alldead	All carbon in detrital biomass	Mg $C \cdot ha^{-1}$
Abovedead	Carbon in aboveground detrital biomass	Mg $C \cdot ha^{-1}$
Deadwoody	Carbon in detrital woody biomass	Mg $C \cdot ha^{-1}$
Treeagdead	Carbon in aboveground dead tree biomass (standing)	Mg $C \cdot ha^{-1}$
Stumps	Carbon in stumps (above and belowground)	Mg $C \cdot ha^{-1}$
CWD	Carbon in down woody material	Mg $C \cdot ha^{-1}$
Deadfines	All carbon in forest floor and mineral soil	Mg $C \cdot ha^{-1}$
Ff	Carbon in forest floor material	Mg $C \cdot ha^{-1}$
SoilC	Carbon in mineral soil	Mg $C \cdot ha^{-1}$
Allabove	All aboveground carbon	Mg $C \cdot ha^{-1}$

cover forests (LAI > 3) and in forests at more advanced succession stages (intermediate to mature) have been less successful (Waring et al. 1995; Lefsky et al. 2001). Light detection and ranging (lidar) is a remote-sensing technology that has demonstrated the capability to estimate biomass in high-cover forests. Using indices of stand height and details of the canopy's vertical structure, discrete-return lidar has been used in regression analysis to estimate forest biomass levels across a range of forest types including temperate mixed deciduous coniferous forests (Popescu et al. 2003; Lim and Treitz 2004), temperate deciduous (Patenaude et al. 2004), temperate coniferous forests (Hall et al. 2005; Hyde et al. 2007), and boreal mixedwood forests (Thomas et al. 2006). Although the ability to estimate biomass has been demonstrated for structurally similar forests, these techniques have not been applied to the type of temperate subalpine ecosystems addressed in this study.

The primary goal of this study is to use discrete-return lidar to develop equations for estimating components of forest structure and biomass in three small landscapes of subalpine coniferous forest. Specifically, we have the following objectives: (*i*) compare the performance of variables derived from canonical correlation analysis (CCA) to simple lidar variables in multiple regression analysis of forest structure and biomass; (*ii*) evaluate environmental and study site effects to assess the need for site specific equations; and

Table 2. Variables identified as significant in PCA and the corresponding eigenvalue for the field and lidar variables used in the subsequent CCA.

Variable	Eigenvalue
Field	
Abovelive	14.94
Treesha	4.02
Seedaglive	2.86
Batree	2.34
BaLm	1.92
SoilC	1.38
BaTA	1.16
Stumps	0.98
Height	
Agemean	
Lidar	
QH	14.50
$CF_{8–10}$	6.73
$CF_{30–32}$	3.58
$CF_{4–6}$	1.62
$CF_{34–36}$	1.04
HP_{20}	0.87

Note: Variables are listed in component order.

(*iii*) characterize the development of canopy and stand structure during stand development.

Methods

Study area

Coincident field and lidar data were collected at three study sites. The Niwot Ridge Long Term Ecological Research site is located approximately 35 km west of Boulder, Colorado, on the drier eastern side of the Continental Divide with plot locations occurring between 3035 and 3115 m. The Fraser Experimental Forest (FEF) is part of the USDA Forest Services Rocky Mountain Research Station located approximately 80 km west of Denver, Colorado, and 30 km southwest of the Niwot site on the wetter western side of the Continental Divide, with plots between 3030 and 3230 m. The Glacier Lakes Ecosystem Experiments Site (Glees) is located 60 km west of Laramie, Wyoming, in the Medicine Bow Mountains (approximately 160 km north-northwest of the Niwot and Fraser sites) with plots between 3175 and 3220 m. All three study sites are dominated by subalpine forest types, which include Englemann spruce (*Picea englemannii* Parry ex Engelm.), subalpine fir (*Abies lasiocarpa* (Hook.) Nutt.), lodgepole pine (*Pinus contorta* Dougl. ex Loud.), limber pine (*Pinus flexilis* James), and trembling aspen (*Populus tremuloides* Michx.) tree species.

Data

First- and last-return discrete lidar data was collected by Airborne 1 (El Segundo, Calif.) using an Optech ALTM 2025/2050. Data was collected from a fixed-wing aircraft at the FEF and Niwot sites on 8 October 2005 and at the Glees site on 23 July 2006. The accuracy of the lidar data was assessed using checkpoints collected over level unvegetated areas; vertical and horizontal accuracies at the Glees site were 0.08 and 0.16 m, whereas the FEF and Niwot sites had vertical and horizontal accuracies of 0.09 and 0.18 m. Mean point densities were 2.36, 1.57, and 2.36 returns·m^{-2}, with standard deviations of 1.02, 0.72, and 0.77 returns·m^{-2} at FEF, Niwot, and Glees, respectively. A "bare earth" digital elevation model (2 m horizontal spatial resolution) was developed using the Toolbox for Lidar Data Filtering and Forest Studies (TIFFS) program (Chen 2007) using both first- and last-return lidar data at each study site. TIFFS uses progressively larger window sizes for morphological operations (similar to Zhang et al. 2003) but without the assumption of a constant slope. Any lidar points above the bare earth surface are considered to have been returned from the vegetation canopy (Næsset 1997).

Thirty-six study plots (12 per site) were established using Forest Inventory Analysis style plot layouts (USDA 2005) to collect forest stand structure data. The plot layout consisted of four annular subplots of radius 8 m, with smaller nested subplots and line transects; three of the subplots are spaced 35 m away from a center subplot at 0°, 120°, and 240°; for the 36 study plots. this results in a total of 144 subplots. Subplot, species, location (distance and azimuth from subplot center), and diameter at breast height (DBH) were recorded for each tree located within the plot. Tree ages were estimated for the largest three to five trees per plot from increment cores. Tree heights were recorded for 10 trees per subplot and then regressed against DBH to generate predictive height equations, which were used in estimation of height for the remaining trees (Bradford et al. 2008). DBH and height were used with published species specific allometric equations to estimate aboveground biomass in foliage, branches, and stems as well as LAI (Bradford et al. 2008). Coarse woody debris (CWD) volume was quantified by establishing 8–15 m long transects at each subplot. Along each transect, fine CWD was quantified by recording the count of CWD with diameter <2.5 cm for the first 2 m, a count of CWD with diameter between 2.5 and 7.5 cm for the first 4 m, and the diameter and decay class (Arthur and Fahey 1990; Busse 1994) of all logs greater than 7.5 cm for the length of the transect (Brown 1971). Forest floor biomass was quantified by harvesting all organic material (other than standing biomass) above mineral soil within three 30 cm × 30 cm quadrats located 7 m from subplot center at 60°, 180°, and 300°. Litterfall was estimated by collecting litter twice a year in five 0.15 m^2 traps per subplot. Aboveground biomass of understory grasses and forbs was collected at peak biomass (late summer) from three 0.25 m^2 quadrats per subplot (Bradford et al. 2008). Plot-level field data were calculated as the means of the four subplots per plot. All field-derived forest stand structure variables are listed in Table 1.

Lidar variables used in regression analysis were calculated using the data contained within a 43.9 m circular buffer (footprint) around the center subplot's origin and using only returns above a specified ground threshold of 0.1 m. Lidar data within each area were used to derive three canopy height indices: maximum height, mean height, and quadratic mean height (Lefsky et al. 1999), and two sets of vertical canopy profile indices: the percentage of lidar points

Table 3. Correlations between field and lidar variables and adjusted R^2 and root mean square error (RMSE) values for the all variable

	Lidar variable											
Field variable	Maxh	Meanh	QH	HP_{10}	HP_{20}	HP_{30}	HP_{40}	HP_{50}	HP_{60}	HP_{70}	HP_{80}	HP_{90}
Agemean	0.64	0.80	0.84	–0.11	0.12	0.41	0.58	0.70	0.78	0.82	0.85	0.86*
Agemax	0.61	0.71	0.76	–0.12	0.11	0.38	0.51	0.61	0.67	0.71	0.74	0.77*
Height	0.68	0.94	0.95	0.12	0.38	0.65	0.75	0.84	0.89	0.93	0.95*	0.94*
Allpla	0.67	0.87*	0.85	0.19	0.46	0.76	0.82	0.83	0.84	0.83	0.82	0.80
Treesha	–0.10	0.45	0.33	0.87*	0.81	0.72	0.61	0.51	0.44	0.38	0.32	0.24
Sapsha	–0.43	–0.44	–0.48	0.06	–0.05	–0.22	–0.32	–0.41	–0.45	–0.46	–0.45	–0.46
Seedsha	–0.20	–0.42	–0.41	–0.25	–0.40	–0.44	–0.42	–0.40	–0.38	–0.38	–0.37	–0.38
Batotal	0.50	0.91	0.84	0.49*	0.71	0.91	0.94*	0.92	0.90	0.86	0.82	0.76*
Batree	0.53	0.91	0.86	0.45*	0.68	0.89	0.93*	0.93	0.91	0.87	0.83	0.78*
BaLP	–0.56*	–0.18	–0.28	0.38	0.24	0.04	0	–0.04	–0.10	–0.17	–0.23	–0.31
BaES	0.41	0.41	0.46	–0.01	0.03	0.25	0.28	0.29	0.32	0.36	0.43*	0.50*
BaSF	0.69	0.64	0.70	–0.12	0.08	0.34	0.42	0.51	0.58	0.64	0.68	0.72*
BaTA	–0.19	–0.16	–0.18	–0.02	0.05	–0.05	–0.12	–0.16	–0.17	–0.16	–0.16	–0.17
BaLm	–0.31	–0.19	–0.23	–0.09	0.03	–0.05	–0.1	–0.14	–0.17	–0.19	–0.2	–0.22
BaSap	–0.40	–0.29	–0.35	0.18	0.11	–0.06	–0.17	–0.26	–0.3	–0.31	–0.31	–0.32
Allcarbon	0.69	0.89*	0.87	0.16	0.48	0.76	0.85	0.88	0.88	0.87	0.84	0.81
Alllive	0.71*	0.93	0.91	0.23	0.51	0.80	0.88*	0.91	0.91	0.90	0.88	0.85
Abovelive	0.72*	0.93	0.91	0.22	0.50	0.79	0.87*	0.90	0.91	0.90	0.88	0.85
Treeaglive	0.72*	0.92	0.90	0.21	0.50	0.78	0.87*	0.90	0.90	0.89	0.87	0.84
Sapaglive	–0.36	–0.22	–0.27	0.34	0.18	0.01	–0.11	–0.20	–0.24	–0.26	–0.25	–0.25
Seedaglive	0.24	0.27	0.31	0.02	–0.04	0.02	0.06	0.14	0.22	0.28	0.34	0.37*
Veg	0.08	–0.42	–0.29*	–0.55	–0.70	–0.65	–0.61	–0.55	–0.47	–0.39	–0.31	–0.19*
Alldead	0.60*	0.75	0.73	0.05	0.40	0.65	0.74	0.76*	0.75	0.74	0.71	0.67
Abovedead	0.67	0.60	0.61	–0.21	0.14	0.44	0.57	0.60	0.61	0.60	0.59	0.57
Deadwoody	0.65	0.59	0.60	–0.20	0.15	0.45	0.57	0.60	0.60	0.59	0.58	0.55
Treeagdead	0.63*	0.64*	0.63	–0.13	0.23	0.52	0.64	0.66	0.66	0.64	0.61	0.57
Stumps	–0.17	–0.27	–0.30	0	–0.02	–0.12	–0.16	–0.21	–0.25	–0.28	–0.29	–0.31
CWD	0.43*	0.10	0.18	–0.38	–0.30	–0.14	–0.05	0.01	0.06	0.11	0.17	0.23
Deadfines	0.38	0.72	0.67	0.31	0.55	0.69	0.72	0.72*	0.71	0.69	0.66	0.62
FF	0.45	0.76*	0.71	0.42*	0.58	0.71	0.74	0.75	0.74	0.72	0.70	0.66
Soilc	–0.10	0.01	0	–0.19	0.01	0.05	0.03	0.03	0.02	0.02	0.01	–0.01
Allabove	0.75*	0.87	0.86	0.07	0.40	0.71	0.82*	0.85	0.86	0.85	0.83	0.80

Note: Variables with asterisks are significant ($p < 0.05$) in regression results for all of the variable models; na, not applicable.

in 2 m high bins ($CF_{0–2}$ through $CF_{36–38}$) and 10 height percentiles (HP_{10} – HP_{90}; the height below which the stated fraction of lidar points fall) (Table 1).

Statistical approach

Three issues influenced the analysis of this data set: the potential for multicollinearity when a large number of independent variables are involved; the choice of variables for, and the comprehensibility of, regression equations; and the need to identify the degree of covariance between the estimates of the dependant variables. Identification of that covariance indicates whether dependent variables represent unique sources of explanatory power for subsequent analyses or whether they are redundant.

To address multicollinearity, principle components analysis (PCA) was used to select a reduced set of explanatory variables for both dependant (field) and independent (lidar) data sets. Using the criteria of Isebrands and Crow (1975), all principle components with eigenvalues greater than 1.0 and the first component with an eigenvalue less than 1.0 were considered significant and retained. The field or lidar variables with the highest correlation with each significant component were identified and composed the reduced set of variables. Although they didn't meet the criteria, mean tree height and mean tree age were included as part of the reduced field data set because they are the most commonly used indicators of stand structure development. PCA reduction of the data set was also necessary to meet the criteria for a CCA, where having more variables in the combined data sets (field and lidar) than total number of samples should be avoided (McGarigal et al. 2000).

Remaining issues include variable choice, comprehensibility of the regression equations, and the covariance of the estimates of the dependant variables. To address these issues, CCA was used to further reduce data redundancy. The output from CCA is two (or more) sets of canonical variables with one set calculated from each of the multivariate data sets, in this case the lidar and field data sets. CCA maximizes correlations between each set of canonical variables, which is used to reveal the common structure between two (or more) multivariate data sets (McGarigal et al. 2000). Each set of canonical variables identifies general trends in a suite of variables rather than selecting specific independent variables, which often have correlations with more than one

model.

$CF_{0–2}$	$CF_{2–4}$	$CF_{4–6}$	$CF_{6–8}$	$CF_{8–10}$	$CF_{10–12}$	$CF_{12–14}$	$CF_{14–16}$	$CF_{16–18}$	Adjusted R^2	RMSE
–0.50	–0.58	–0.11	–0.01	0.12	0.36	0.72	0.79*	0.74	0.77	36.38
–0.38	–0.63	–0.18	–0.07	0.11	0.37	0.69*	0.72	0.66	0.62	77.15
–0.75	–0.60	0.04	0.26*	0.39	0.54	0.69	0.70	0.71	0.93	1.02
–0.58	–0.58	–0.20	0.05	0.29	0.51	0.72	0.71	0.71	0.74	1.80
–0.76*	–0.21	0.44	0.71	0.81	0.71	0.36	0.09	–0.07	0.87	277.34
0.11	0.58*	0.28	0.21	–0.03	–0.27	–0.49	–0.52	–0.47	0.32	885.96
0.46*	0.31	–0.24	–0.33	–0.34	–0.33	–0.30	–0.25	–0.22	0.19	1138.6
–0.83	–0.56	0.07	0.38	0.60	0.72	0.73	0.62	0.54	0.93	6.16
–0.79	–0.60	0.01	0.32	0.56	0.72	0.76	0.65	0.58	0.91	7.01
–0.27*	0.42	0.50	0.41	0.30	0.13	–0.19	–0.31	–0.41	0.43	0.25
–0.25	–0.43	–0.08	0.06	0.25	0.34*	0.33	0.26	0.26	0.43	0.13
–0.25*	–0.57	–0.30	–0.17	0	0.23	0.55	0.63	0.68	0.55	0.22
–0.11	0.40*	0.25	0.29	0.04	–0.16	–0.24	–0.24	–0.21	0.13	0.13
–0.09	0.30	0.56*	0.30*	–0.07	–0.23	–0.27	–0.24	–0.24	0.36	0.06
–0.12	0.52*	0.43	0.38*	0.14	–0.12	–0.38	–0.46	–0.44	0.39	2.22
–0.60	–0.60	–0.15	0.07	0.28	0.48	0.65	0.68	0.74	0.79	52.09
–0.67	–0.63	–0.10	0.14	0.36	0.56	0.71	0.70	0.72	0.93	15.96
–0.66	–0.63	–0.12	0.13	0.34	0.54	0.71	0.71	0.73	0.93	13.61
–0.65	–0.63	–0.12	0.12	0.34	0.54	0.71	0.71	0.73	0.92	14.44
–0.14*	0.42	0.37	0.32	0.23	0.02	–0.30	–0.43*	–0.41	0.24	1.52
–0.20	–0.17	0.07	0.08	0.08	0.09	0.09	0.15	0.25	0.11	1.15
0.71*	0.03	–0.41	–0.59	–0.58	–0.49	–0.27	–0.13	–0.03	0.67	0.35
–0.46	–0.50	–0.19	–0.02	0.17	0.36	0.52	0.59	0.69	0.62	34.9
–0.19	–0.38	–0.37	–0.29	–0.15	0.08	0.40	0.59	0.73*	0.52	19.36
–0.19	–0.37	–0.37	–0.28	–0.15	0.08	0.40	0.58	0.72*	0.51	23.62
–0.28	–0.41	–0.29	–0.20*	–0.07*	0.14	0.44	0.60	0.73	0.73	13.10
0.29	0.29	–0.20	–0.25	–0.21	–0.21	–0.26	–0.23	–0.17	na	na
0.24	–0.04	–0.45	–0.47*	–0.38	–0.22	0.02	0.20	0.33	0.29	5.42
–0.61	–0.51	0.05	0.27	0.46	0.56	0.52	0.45	0.46	0.51	21.89
–0.66	–0.51	0.04	0.30	0.53	0.63	0.54	0.45	0.45	0.61	18.41
0.03	–0.05	0.04	–0.03	–0.11	–0.08	0.02	0.05	0.09	na	na
–0.53	–0.58	–0.22	–0.02	0.18*	0.41	0.64	0.71	0.78	0.90	22.99

general trend in forest stand structure. Previous results (Lefsky et al. 2005*a*, 2005*b*) have indicated that using CCA components as independent variables results in more interpretable and parsimonious regression models.

In addressing the first objective, two sets of stepwise multiple linear regression models were developed for estimating the forest structure and biomass variables. One set of models was developed using the full set of lidar explanatory variables, and they are referred to as "all variable'' regressions. The all variable models used the full set of explanatory variables; thus, they represent the model set having the upper limit of possible explained variance. A second set of models, "CCA," used canonical components, derived from the PCA -reduced set of lidar indices, as explanatory variables in regression modeling.

To test for the existence of environmental and study site effects (the second objective), three climatic and two terrain variables were added to the variables used in the CCA models to yield a third set of models, which will be referred to as "CCA with environment." The terrain variables were mean slope and transformed aspect (Beers et al. 1966), and were derived from the lidar bare-earth digital elevation model. The climate variables (yearly precipitation, snow water equivalent, and mean temperature) were retrieved from the 1 km resolution PRISM data set (PRISM Group: www.ocs.oregonstate.edu/prism/index.phtml). ANOVA analyses on regression residuals for the CCA and CCA with environment models were then performed testing whether site effects were statistically significant both before and after environmental variables were added.

Following Lefsky et al. (2005*b*), plots of correlations between field-derived forest stand structure variables and the first two lidar canonical components (LCC_1 and LCC_2) were developed (the third study objective). The resulting ordination diagrams identify clusters of field data variables that have similar relationships to the respective lidar canonical. Graphs of forest age against LCC_1 and LCC_2 values were developed to evaluate forest structure pattern relationships with forest age. To increase understanding of the lidar canonicals, a qualitative visual assessment of the raw lidar point clouds was also performed using ArcScene's (ESRI, Redlands, Calif.) three-dimensional visualization for five plots located along the LCC_1 and LCC_2 axes for each data set. Selected plots are located at the extremes of the first

Table 4. Lidar variable correlations with lidar canonicals.

	Canonical					
Variable	LCC_1	LCC_2	LCC_3	LCC_4	LCC_5	LCC_6
Maxh	0.70	–0.51	–0.02	0.00	0.11	0.10
Meanh	0.98	0.02	–0.12	0.09	0.03	–0.01
QH	0.98	–0.14	–0.07	0.05	0.06	0.02
HP_{10}	0.19	0.80	–0.35	–0.14	0.05	0.08
HP_{20}	0.45	0.78	–0.36	0.21	0.10	0.05
HP_{30}	0.72	0.52	–0.32	0.20	0.05	–0.03
HP_{40}	0.82	0.31	–0.30	0.20	–0.01	–0.08
HP_{50}	0.90	0.14	–0.23	0.15	–0.04	–0.09
HP_{60}	0.95	0.02	–0.16	0.12	–0.05	–0.08
HP_{70}	0.98	–0.06	–0.09	0.08	–0.03	–0.05
HP_{80}	0.98	–0.13	–0.01	0.05	0.02	–0.01
HP_{90}	0.96	–0.22	0.05	0.00	0.11	0.05
$CF_{0–2}$	–0.76	–0.57	–0.15	–0.01	–0.04	–0.01
$CF_{2–4}$	–0.67	0.15	0.30	0.24	0.11	0.06
$CF_{4–6}$	–0.01	0.67	0.73	0.07	0.07	0.04
$CF_{6–8}$	0.25	0.84	0.38	–0.21	0.03	0.04
$CF_{8–10}$	0.43	0.76	–0.18	–0.45	–0.03	0.03
$CF_{10–12}$	0.60	0.48	–0.43	–0.40	–0.10	–0.01
$CF_{12–14}$	0.74	–0.09	–0.35	–0.04	–0.13	–0.10
$CF_{14–16}$	0.74	–0.39	–0.21	0.23	–0.13	–0.15
$CF_{16–18}$	0.75	–0.49	–0.10	0.37	–0.07	–0.11
$CF_{18–20}$	0.70	–0.46	–0.05	0.38	0.05	–0.08
$CF_{20–22}$	0.67	–0.44	–0.03	0.37	0.16	0.04
$CF_{22–24}$	0.62	–0.44	0.00	0.27	0.29	0.12
$CF_{24–26}$	0.53	–0.42	0.08	0.12	0.38	0.23
$CF_{26–28}$	0.44	–0.38	0.14	0.03	0.47	0.47
$CF_{28–30}$	0.31	–0.30	0.17	–0.06	0.62	0.58
$CF_{30–32}$	0.24	–0.26	0.19	–0.11	0.69	0.59
$CF_{32–34}$	0.19	–0.19	0.13	0.02	0.08	0.91
$CF_{34–36}$	0.15	–0.14	0.09	0.06	–0.18	0.95

two canonical variable axes and at the neutral loading (zero, zero) location.

Results and interpretations

Principal components analysis

To eliminate redundant variables, separate PCAs were performed for the field and lidar data sets. For the field data and lidar data sets, the first eight and six components were significant, explaining 92% and 93% of variance, respectively. Table 2 shows the field and lidar variables that were identified as having the highest correlation with each retained component. Pearson correlations between lidar and field variables are given in Table 3.

Canonical correlation analysis

The two sets of variables identified using PCA were then used in a CCA. Six canonical variables were retained for use in regression modeling in the CCA models, with canonical correlation coefficients ranging from 0.99 to 0.38. Using an *F* statistic based on Rao approximation and Wilks' Lambda tests, only canonicals one through three were statistically significant. Although the remaining canonicals were not statistically significant, they were retained for use in regression modeling as they have explanatory power in estimating the dependent variables. In all subsequent analyses, only the set of lidar-derived CCA components were considered. These components were calculated directly from the lidar data, with no mathematical contribution from the field plot data set.

Correlations between lidar and field data with lidar canonicals

Lidar canonical component 1 (LCC_1) correlates with mean stand height (r = 0.98), quadratic stand height (r = 0.98), and HP_{50}–HP_{90} (all r >0.90) (Table 4), indicating that it represents an index of mean stand height. In the field data set, LCC_1 is correlated with mean tree height (r = 0.97), all carbon in live biomass (r = 0.90), total basal area (r = 0.88), and total leaf area (r = 0.83) (Table 5). This high degree of correspondence with these variables is typical of lidar remote sensing (Lefsky et al. 2005*b*).

Lidar canonical 2 (LCC_2) correlates with the height percentiles HP_{10}, HP_{20}, and HP_{30} (r = 0.80, 0.78, and 0.52, respectively) and percentage of lidar observations $CF_{4–6}$, $CF_{6–8}$, and $CF_{8–10}$ (r = 0.67, 0.84, and 0.76, respectively); observations from just above the lower canopy. Tree density (r = 0.76) is the field variable most highly correlated with LCC_2, followed by basal area of lodegepole pine (r = 0.53) and the total carbon and basal area of saplings (r = 0.40 and 0.39, respectively). Carbon in understory vegetation, carbon in downed woody material, and seedling density have high

Table 5. Correlations of field variables with lidar canonicals and adjusted R^2 and root mean square error (RMSE) values for CCA models..

	Lidar canonical							
Field variable	LCC_1	LCC_2	LCC_3	LCC_4	LCC_5	LCC_6	R^2	RMSE
Agemean	0.83*	–0.28*	0.11	0	0	–0.01	0.74	38.57
Agemax	0.74*	–0.26*	0.01	0	–0.04	–0.05	0.59	80.17
Height	0.97*	–0.02	0.09	–0.02	0.03	–0.04	0.93	0.99
Allpla	0.83*	–0.04	–0.24*	0.17	0.05	0.02	0.73	1.85
Treesha	0.43*	0.76*	–0.09	–0.08	0.04	0.05	0.75	380.9
Sapsha	–0.45*	0.25	0.14	0.01	0.02	0.14	0.18	971.6
Seedsha	–0.45*	–0.27	–0.08	–0.08	–0.05	0.10	0.18	1147
Batotal	0.88*	0.30*	–0.18*	0.08	0.02	0.02	0.88	7.92
Batree	0.88*	0.25*	–0.20*	0.08	0.01	0	0.87	8.56
BaLP	–0.19	0.53*	0.20	–0.04	–0.08	–0.04	0.26	0.29
BaES	0.46*	–0.13	0.03	–0.35*	0.06	0.14	0.30	0.14
BaSF	0.65*	–0.34*	–0.10	0.08	0.04	–0.04	0.52	0.23
BaTA	–0.15	0.20	0.15	0.06	–0.01	–0.01	0	na
BaLm	–0.17	0.29	0.48*	0.27	–0.02	–0.05	0.21	0.07
Basap	–0.29	0.39*	0.21	0	0.05	0.15	0.13	2.65
Allcarbon	0.85*	–0.02	–0.20*	0.22*	0.02	–0.06	0.79	51.35
Alllive	0.90*	0.02	–0.17*	0.14*	0.05	0	0.85	24.06
Abovelive	0.90*	0	–0.17*	0.15*	0.05	0	0.84	20.58
Treeaglive	0.89*	0	–0.18*	0.15*	0.05	–0.01	0.83	21.48
Sapaglive	–0.22	0.40*	0.12	–0.07	0.09	0.23	0.14	1.62
Seedaglive	0.32	–0.12	0.21	–0.14	0.01	0.25	0	na
Veg	–0.38*	–0.63*	0.02	–0.15	0.05	0.03	0.52	0.42
Alldead	0.71*	–0.06	–0.22	0.28*	–0.02	–0.12	0.55	37.95
Abovedead	0.54*	–0.33*	–0.23	0.37*	0	–0.14	0.50	19.88
Deadwoody	0.53*	–0.31*	–0.24*	0.39*	0	–0.14	0.53	23.05
Treeagdead	0.58*	–0.22	–0.21	0.40*	–0.05	–0.18	0.47	21.01
Stumps	–0.35*	–0.03	–0.29	0.18	0.22	0.02	0.10	18.57
CWD	0.08	–0.54*	–0.15	0.05	0.18	0.12	0.27	1.67
Deadfines	0.70*	0.23	–0.13	0.09	–0.04	–0.06	0.48	22.54
Ff	0.74*	0.24*	–0.17	–0.02	0.04	0.05	0.58	19.25
SoilC	0.01	0.01	0.05	0.23	–0.19	–0.22	ns	na
Allabove	0.82*	–0.12	–0.20*	0.24*	0.03	–0.05	0.76	36.5

Note: Variables with asterisks are significant ($p < 0.05$) in regression results for the CCA models. na, not applicable; ns, not significant.

negative correlations with LCC_2 (r = –0.63, –0.54, and –0.27, resectively). Positive values of LCC_2 appear to represent forest layers within the midcanopy and the associated suppression of vegetation below this height (plots with mean height <3.5 m have negative LCC_2 values). Increases in stem density appear to be associated with delayed recruitment in these stands.

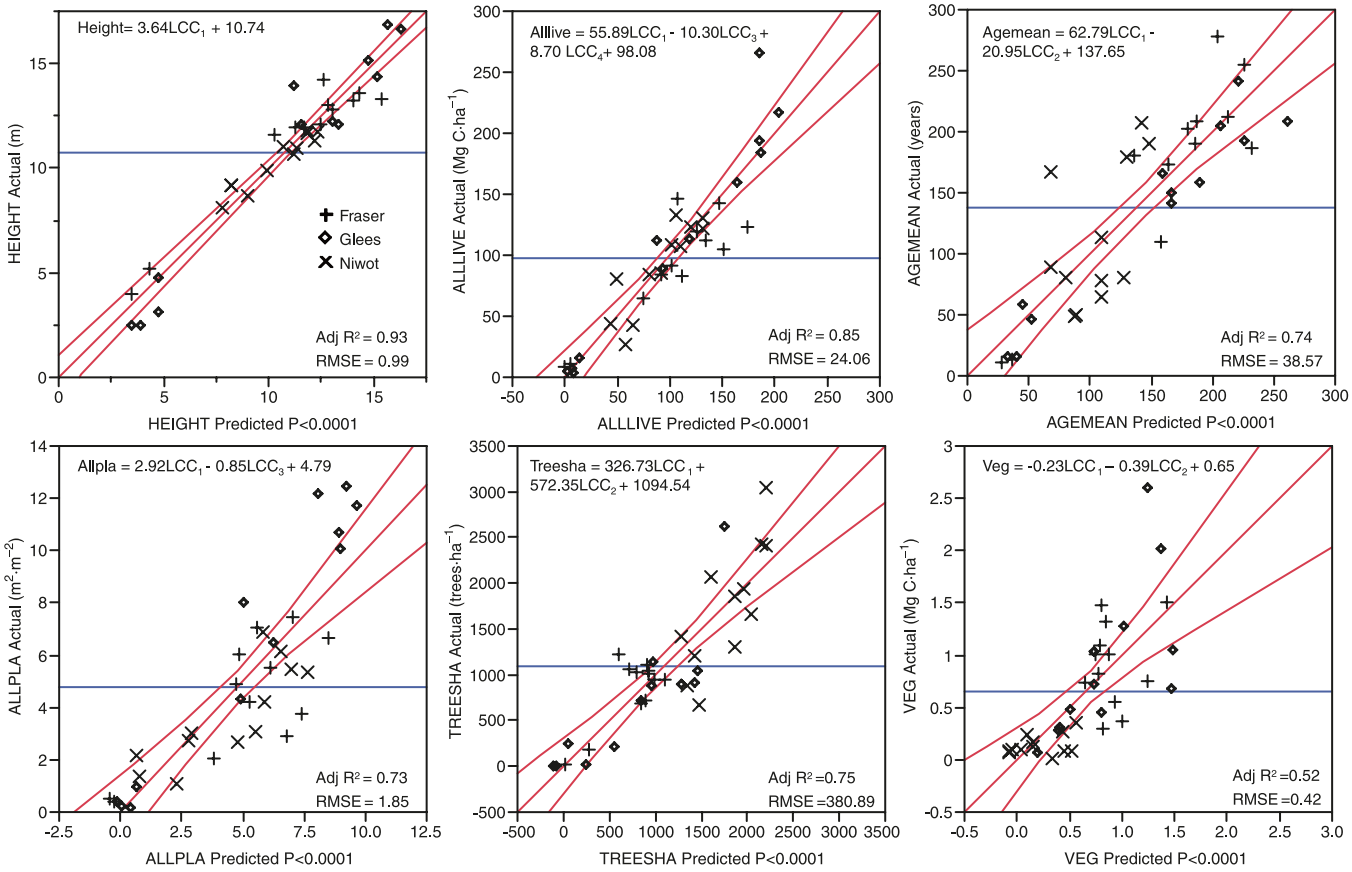
Relative to the first two canonical components, LCC_3 is not clearly correlated to fundamental forest attributes, such as forest height and forest density. LCC_3 is most highly correlated with $CF_{4–6}$ and $CF_{6–8}$ (0.73 and 0.38), suggesting that, as with LCC_2, LCC_3 is picking up understory and low canopy characteristics in the height range associated with sapling sized trees. Basal area of limber pine (0.48) is the field variable with the highest correlation to LCC_3. Correlations with lidar canonicals LCC_4 through LCC_6 show a progression of increasing correlation with increasing forest canopy height bin variables, which may be an artifact of the analysis method (Lefsky 1997). LCC_4 is correlated with $CF_{14–16}$ through $CF_{20–22}$ (r = 0.23–0.38) and negatively associated with lower canopy height bins $CF_{8–10}$ and $CF_{8–12}$ (r = –0.45 and –0.40, respectively). LCC_5 is correlated with the higher range of $CF_{24–26}$ through $CF_{30–32}$ (r = 0.38–0.69), and LCC_6 is highly correlated with the range of $CF_{26–28}$ through $CF_{34–36}$. LCC_4, LCC_5, and LCC_6 are most highly correlated with the field variables carbon in aboveground dead biomass (r = 0.40), carbon in stumps (r = 0.22) and carbon in live seedlings (r = 0.25), respectively.

Comparison of canonical and lidar variable regressions

Stepwise regression with all lidar variables

All variable regression results (which used all lidar variables as explanatory variables) are shown in Table 3. Adjusted R^2 values range from a high of 0.93 for mean tree height (RMSE < 1 m), basal area of all trees and saplings, all carbon in live biomass, and carbon in aboveground live biomass to nonsignificant for carbon in mineral soil and carbon in stumps. Carbon in aboveground components of live trees, basal area of all trees, and all aboveground carbon variables are the next three highest R^2 models (R^2 = 0.92, 0.91, and 0.90, respectively). Several other dependent varia-

Fig. 1. Regression graphs, equations, adjusted R^2, and root mean square error (RMSE) values for mean height (Height); sum of leaf area for trees, saplings, and seedlings (Allpla); all carbon in live biomass (Alllive); tree density (Treesha); mean tree or sapling age (Agemean); and carbon in understory vegetation (Veg) field variables using lidar canonical variables (LCCs) as explanatory variables.



bles are predicted well enough to be useful including tree density (R^2 = 0.87), mean age (R^2 = 0.77), and total leaf area (R^2 = 0.74). A third set of variables are estimated more poorly but represents variables that are not commonly accessible through remote sensing, including dead tree aboveground carbon (R^2 = 0.73), understory herbaceous vegetation (R^2 = 0.67), carbon in forest floor and mineral soil (R^2 = 0.61), and carbon in forest floor material (R^2 = 0.51).

Stepwise regression with lidar canonicals

Modeling results with the lidar canonicals as explanatory variables, (CCA models) are shown in Table 5. Adjusted R^2 values range from 0.93 for mean tree height to nonsignificant for basal area of trembling aspen and carbon in live seedlings. Basal area of all trees and saplings, basal area of all trees, and all carbon in live biomass are the next-highest R^2 models (R^2 = 0.88, 0.87, and 0.85, respectively). Regression graphs and the corresponding regression equations are given for mean height; sum of leaf area for trees, saplings, and seedlings; all carbon in live biomass; tree density; mean tree or sapling age; and carbon in understory vegetation in Fig. 1. The set of dependant variables that can be estimated from the CCA models is the same as for the all variable analysis, although the percentage of variance explained is slightly less in nearly every case (Table 6). The developed regression equations were used to develop spatial maps (30 m resolution) of the stand structure and carbon pool variables at each study site. The spatial maps developed with the canonical regression equation for all carbon in live biomass (Table 5, Alllive) are shown in Fig. 2.

Summary of regressions

The all variable models have the best overall model fit as measured by explained variance (adjusted R^2) and the ratio of mean predicted value (RMPV), (RMSE divided by the mean predicted value). Differences between the CCA and all Variable model R^2 values are shown in Table 6. The all variable models have R^2 values that are, on average, 0.08 higher than the CCA models. The all variable models have lower (reduced relative model error) RMPV ratios relative to the CCA models. The mean RMPV value for the CCA models with R^2 values greater than 0.80 is 0.21, whereas the mean RMPV value for the corresponding all variable models is slightly lower (improved) at 0.15.

Evaluation of environment and study site effects

Stepwise regression with environmental variables and ANOVAs

The CCA with environment regressions models were developed to evaluate environmental influences across the three study sites. In the CCA with environment models, increases in adjusted R^2 relative to the CCA models range from zero (no significant environmental variable) to as high

Table 6. Model comparisons using adjusted R^2 values for all variables, CCA, and CCA with environment regression models.

	Adjusted R^2						ANOVA on residuals	
Variable	All variable	CCA	CCAwith environment	All variable – CCA	RMPV × all variable	RMPV × CCA	CCA	CCA with environment
Agemean	0.77	0.74	0.77	0.03	0.26	0.28	ns	ns
Agemax	0.62	0.59	0.59	0.03	0.34	0.36	ns	ns
Height	0.93	0.93	ne	0.00	0.09	0.09	ns	ns
Allpla	0.74	0.73	0.82	0.01	0.56	0.39	0.005	ns
Treesha	0.87	0.75	ne	0.12	0.25	0.35	ns	ns
Sapsha	0.32	0.18	0.33	0.14	0.66	0.72	0.01	ns
Seedsha	0.19	0.18	ne	0.01	0.95	0.96	ns	ns
Batotal	0.93	0.88	ne	0.05	0.14	0.18	ns	ns
Batree	0.91	0.87	0.88	0.04	0.17	0.21	ns	ns
BaLP	0.43	0.26	0.54	0.17	0.78	0.91	0.001	ns
BaES	0.43	0.30	ne	0.13	0.81	0.88	ns	ns
BaSF	0.55	0.52	0.77	0.04	0.61	0.64	0.01	ns
BaTA	0.13	ns	0.09	0.13	3.94	na	na	ns
BaLm	0.36	0.21	0.41	0.15	2.86	3.33	0.02	ns
Basap	0.39	0.13	0.19	0.26	0.64	0.76	ns	ns
Allcarbon	0.79	0.79	0.83	0.00	0.21	0.21	ns	ns
Alllive	0.93	0.85	0.88	0.08	0.16	0.25	ns	ns
Abovelive	0.93	0.84	0.88	0.09	0.17	0.26	ns	ns
Treeaglive	0.92	0.83	0.87	0.09	0.19	0.28	0.04	ns
Sapaglive	0.24	0.14	0.23	0.10	0.59	0.63	0.003	ns
Seedaglive	0.11	ns	0.18	0.11	0.83	na	na	ns
Veg	0.67	0.52	ne	0.15	0.54	0.65	ns	ns
Alldead	0.62	0.55	ne	0.07	0.23	0.26	ns	ns
Abovedead	0.52	0.50	0.55	0.03	0.86	0.89	ns	ns
Deadwoody	0.51	0.53	0.58	–0.02	0.86	0.84	ns	ns
Treeagdead	0.73	0.47	ne	0.27	0.94	1.51	0.03	ns
Stumps	ns	0.10	ne	–0.10	na	10.32	ns	ns
CWD	0.29	0.27	ne	0.02	0.64	0.20	ns	ns
Deadfines	0.51	0.48	ne	0.03	0.18	0.19	ns	ns
Ff	0.61	0.58	0.64	0.04	0.30	0.32	ns	ns
SoilC	ns	ns	0.17	na	na	na	na	ns
Allabove	0.90	0.76	0.82	0.14	0.22	0.35	ns	ns

Note: RMPV, ratio of mean square error to the mean predicted value for all variable and CCA models. Regression model residual tests for analysis of variance (ANOVA, $F_{2,33}$) by site and significance level, for the CCA and CCA with environment" models. na, not applicable; ne, no significant environmental variables; ns, not significant.

*RMPV is RMSE/MPV. MPV is mean predicted value.

as 0.28 for basal area of lodgepole pine (Table 6). The CCA with environment models having greater than 0.05 increases in adjusted R^2 values all occur in models having relatively low explained variance ($R^2 \leq 0.76$) in the CCA models.

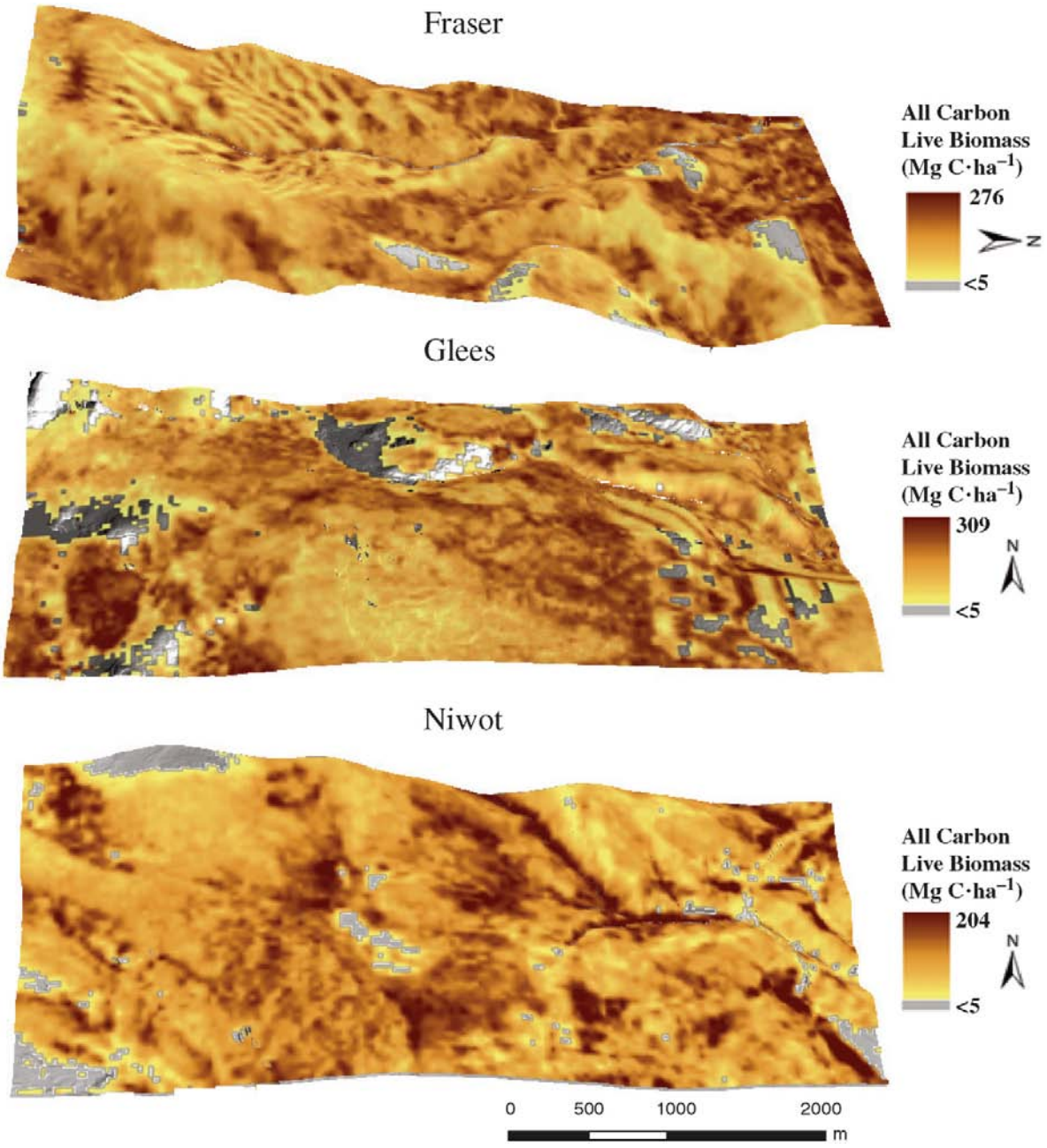
ANOVAs by study site were performed on the CCA and CCA with environment model residuals to determine if significant site differences still existed after environmental differences were accounted for in the CCA with environment" models. ANOVA by study site for the CCA and CCA with environment models were significant for eight and no models, respectively (Table 6). Significant site differences exist for the total leaf area, basal area of lodgepole pine, basal area of limber pine, basal area of subalpine fir, carbon in live saplings, sapling density, carbon in aboveground components of live trees, and carbon in aboveground dead tree biomass variables in the CCA models. Site differences for species-specific basal areas are expected because species distribution is not uniform across sites, for example, limber pine is absent from Glees and Fraser. ANOVA results on the CCA with environment model residuals were nonsignificant for all variables. This implies that the environmental site differences observed in the significant CCA model ANOVAs, are incorporated in the CCA with environment models. By the inclusion of the two terrain and three climatic variables, a single equation can be used to capture study site variability.

Examination of canopy and stand structure development

Field data correlation graphs with lidar canonicals 1 and 2

The degree of correlation between each field data variable and the first two canonical components were graphed (Fig. 3), and four distinct clusters were observed. In this space, the distance between variables indicates the degree of statistical independence between them (larger distances indicating lower correlation). The first cluster variables are the most reliable estimates with high degrees of correlation with LCC_1 and moderately positive to moderately negative correlations with LCC_2. This cluster is related to numerous

Fig. 2. Spatial map of all carbon in live biomass (Alllive) for the Fraser, Glees, and Niwot study sites using CCA explanatory variable regression equation (see Table 5).



variables related to forest height including forest age, biomass, and basal area. Cluster two consists of the tree density variable, which is moderately correlated with LCC_1 and highly correlated with LCC_2.

Cluster three consists of sapling variables and the basal area of lodgepole pine, limber pine, and trembling aspen. This cluster has moderate negative correlations with LCC_1 and moderate positive correlations with LCC_2, indicating these variables are highest at shorter stand heights and higher levels of midstory development. It is necessary to stress that the species relationships observed in canonical space are picking up forest structure characteristics associated with the sampled study plots; thus, observed patterns are not intrinsically tied to species type. For example, lodgepole pine can be associated with mature forest having high stand heights and minimal sapling presence.

Lastly, cluster four consists of numerous understory related variables including seedling variables and carbon in stumps, downed woody debris, mineral soil, and understory vegetation. Cluster four has a large correlation range, with LCC_1 correlations from neutral to moderately negative and LCC_2 correlations from neutral to highly negative.

Mean age graphs by lidar canonical variables 1 and 2

Canonicals LCC_1 and LCC_2 are strongly related to structural correlates of forest succession within our three study areas. Graphs of forest age by LCC_1 and LCC_2 values show the relationship between mean forest age and forest structure

Fig. 3. Field data variable correlations with lidar canonicals 1 (LCC_1) and 2 (LCC_2).

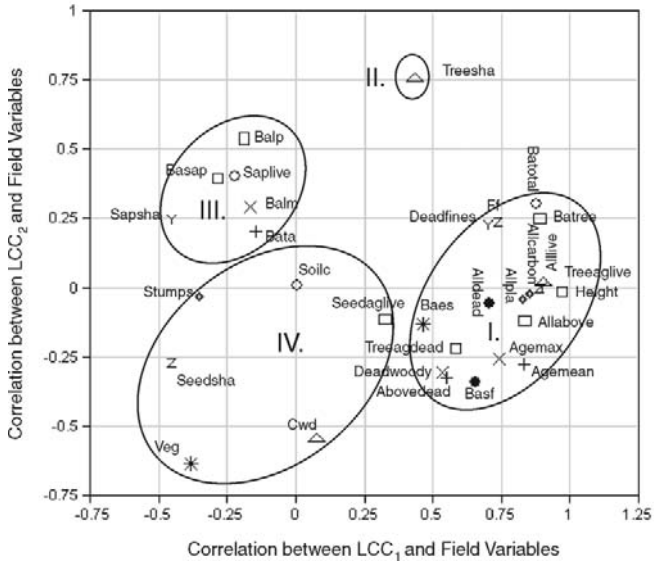
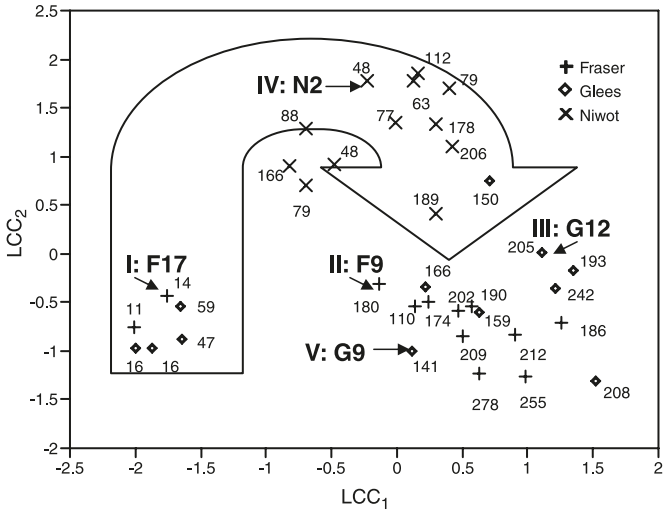


Fig. 4. Mean tree age plotted against LCC_1 and LCC_2 values. Points are plots within the three sites. Point I (F17: $LCC_1 = -1.77$, $LCC_2 = -0.43$), point II (F9: $LCC_1 = -0.14$, $LCC_2 = -0.31$), point III (G12: $LCC_1 = 1.11$, $LCC_2 = 0.01$), point IV (N2: $LCC_1 = -0.24$, $LCC_2 = 1.76$), and point V (G9: $LCC_1 = 0.12$, $LCC_2 = -1.00$). The arrow tracks the development of forests from young to mature forests.



(Fig. 4); three age-related clusters are apparent. Plots with negative LCC_1 and LCC_2 values are generally young in age (mean age <50 years), intermediate-aged plots (mean age between 50 and 175 years) have neutral LCC_1 and positive LCC_2 values, and lastly, the mature cluster (>175 mean age) have positive LCC_1 and neutral to negative LCC_2 values. LCC_1 explains the majority of the variance in stand age (69%); the inclusion of LCC_2 explains an additional 5% of variance, which is related to midstory development. Two Niwot sites with mean tree ages of 48 years are outliers from the young cluster having positive LCC_2 values of 1.76 and 0.91. Relative to the other young sites these two plots have high sapling densities.

Discussion

Much of the research investigating forest canopy and stand structure complexity has been performed in high-productivity forest types such as tropical rain forests (e.g., MacArthur et al. 1966; and Drake et al. 2002) and old-growth Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) forests of the Pacific Northwest (Weiss 2000 and Parker et al. 2004). Older forests of these types have high canopy cover and stand height. Using multivariate analysis methods, this study investigated forest structure patterns and compared similar patterns in subalpine forest, which are shorter and have lower canopy cover and biomass. The mean aboveground live biomass of forests in this study is 80 Mg C·ha^{-1}; the mean aboveground live biomass of coniferous stands in the Pacific Northwest (Lefsky et al. 2002) is 300 Mg C·ha^{-1}. Important to this study was the analysis of the large forest structure data set consisting of 32 dependent variables, of which 15 and 17 were related to stand structure and carbon pool measures. Analysis of this robust data set allowed for the simultaneous evaluation of covariance between numerous forest variables that are influenced by forest structure complexity.

Comparison of canonical and lidar variable regressions

Regression analysis results

The forest height related variables: forest height, biomass, basal area, and forest age are known to have a relationship with lidar-derived indices. Numerous studies have shown a strong relationship between discrete-return forest canopy height measurements and field-based canopy height measurements (Næsset 1997; and Popescu et al. 2002). A strong height and biomass relationship was also observed in this study with modeling results for height and biomass variables having R^2 values >0.90. A strong relationship between leaf area measurements and lidar has been shown using height percentiles and height indices (Magnussen and Boudewyn 1998; Riaño et al. 2004), first and last lidar return ratio (Morsdorf et al. 2006; Solberg et al. 2006) and using waveform data (Lefsky et al. 1999). This LAI relationship was also observed in this study with modeled LAI R^2 values of 0.74, 0.73, for the all variable, and CCA models.

The all variable regression models on average have slightly higher explained variance and lower RMPV values relative to the CCA regression models. The slightly better model fit in the all variable models was expected and may be a product of overfitting the models. The strength of CCA is its use in multivariate analysis to evaluate forest structure patterns that are not apparent when evaluating variables at a singular level. This said, a fundamental weakness and strength of using CCA is the loading of multiple variables per canonical. This loading of multiple variables allows for the assessment of covariance but also makes interpretation and verification of the derived canonical values more difficult (Kikkawa 1982). For example, the LCC_2 canonical is highly correlated with both tree density and basal area of lodgepole pine, making interpretation of the canonical more difficult.

Table 7. Correlation values of mean tree age and tree density (with total ecosystem carbon, all carbon in live biomass, and all carbon in aboveground live biomass.

	Total ecosystem carbon	Carbon in live biomass	Carbon in aboveground live biomass
Mean tree age			
Field	0.67	0.71	0.70
All variable	0.93	0.87	0.87
CCA	0.90	0.92	0.92
Tree density			
Field	0.30	0.42	0.40
All variable	0.45	0.44	0.45
CCA	0.47	0.48	0.48

Note: Correlations are presented for each combination of variables: the correlation between the original field estimates and the correlation of estimates from the all variable and CCA methods.

When two variables have a close proximity in the LCC_1 and LCC_2 ordination space, their estimates will be more highly correlated with each other than the original two variables. This is true for both the all variable and CCA estimates, but the correlation of the estimates will be higher using the CCA estimates. The increased correlation of the estimated variables is due to similarity of each variable's relationship to the canonical components that are used to estimate them. This tendency is demonstrated by Table 7, which reports the level of correlation between three of the cluster I variables (total ecosystem carbon, all carbon in live biomass, and carbon in aboveground live biomass) with two variables: mean tree age (another cluster I variable) and a cluster II variable (tree density). As expected, field estimates of mean tree age have a relatively high correlation with the other cluster I variables with R^2 values ranging from 0.67 to 0.71, but correlations of the all variable and CCA estimates of mean tree age are substantially higher, with R^2 values of 0.87–0.93 and 0.90–0.92, respectively. In contrast, the correlation between the tree density estimates (from either the all variable or CCA analyses) were only slightly higher than the correlation between the field estimates of these variables.

Examination of canopy and stand structure development

Canonical correlation analysis

Four distinct clusters were identified in the LCC_1–LCC_2 ordination graph (Fig. 3). As determined from the mean age graph and correlations between the field and lidar variables, LCC_1 is highly correlated with forest height, biomass, and forest age. Plots having high LCC_1 values are intermediate to mature in age (>175 years) with high forest height and biomass. Conversely, plots having low LCC_1 values are young in age (<50 years) with low forest height and biomass values. LCC_2 tracks the successional development of forests from young to mature (Fig. 4). In low LCC_1 plots, LCC_2 tracks the transition from young plots with seedlings (low LCC_2) to intermediate-aged plots (high LCC_2) with saplings and low-stature lodgepole pine, limber pine, and aspen, which in these plots are primarily young, and relatively dense short stands. For high LCC_1 plots (mature and old), variation in LCC_2 is more subtle but differentiates between newly mature stands with high stem density (high LCC_2 values) to older stands (low LCC_2 values) having high dead woody material from increased mortality and high basal area of spruce and fir (which are the oldest trees in our study plots). The loading of multiple forest structure components (e.g., sapling density, tree density, dead wood material, and basal area of spruce) on LCC_2 is an example of the major advantage of using canonicals to study forest structure relative to singular lidar indices.

The sapling and basal area variables associated with cluster III (Fig. 3) all have explained variances <0.50. Sapling and seedling variables (predominately <2 m) contribute few lidar returns in dense multistoried canopy forests, and any ability to estimate their abundance is due to their correlation with overstory canopy structure. We expect that these correlations would be modified by numerous site factors such as forest age, forest life form (deciduous versus coniferous), forest composition, and terrain slope.

In these subalpine forests, a substantial fraction of total ecosystem carbon is stored in forest floor material (Bradford et al. 2008), so estimating forest floor carbon with lidar would be a valuable carbon assessment tool. The forest floor variables associated with cluster IV (Fig. 3) are poorly explained by the lidar variables in regression models. Carbon in understory vegetation is the only variable with R^2 values >0.50 (0.67 for all variable and 0.52 for CCA). This suggests that, although there may be an underlying level of forest structure influencing these variables (as observed by the formation of the cluster), the lidar canonicals and indices used in this analysis poorly explain the cluster IV forest floor variables.

Conversely, the forest floor variables in cluster I (carbon in forest floor and mineral soil and carbon in forest floor material) have moderate explained variance values (R^2 = 0.51 and R^2 = 0.48 and R^2 = 0.61 and R^2 = 0.58) in the respective all variable and CCA models. Soil carbon is poorly estimated so that there is no difference between the correlation of LCC_1 and LCC_2 and the two variables. It is only the forest floor carbon component that is being tracked by these two variables, and this apparently monotonically increases with stand age.

Estimation of forest understory variables with discrete-return lidar represents an area for potential research in the future, especially in light of technology advancements, which are resulting in increased lidar sensor capacity (e.g., greater return density and full waveform returns). Current

Fig. 5. Lidar point clouds for the five selected field plots in Fig. 4 representing the range of observed LCC_1 and LCC_2 values.

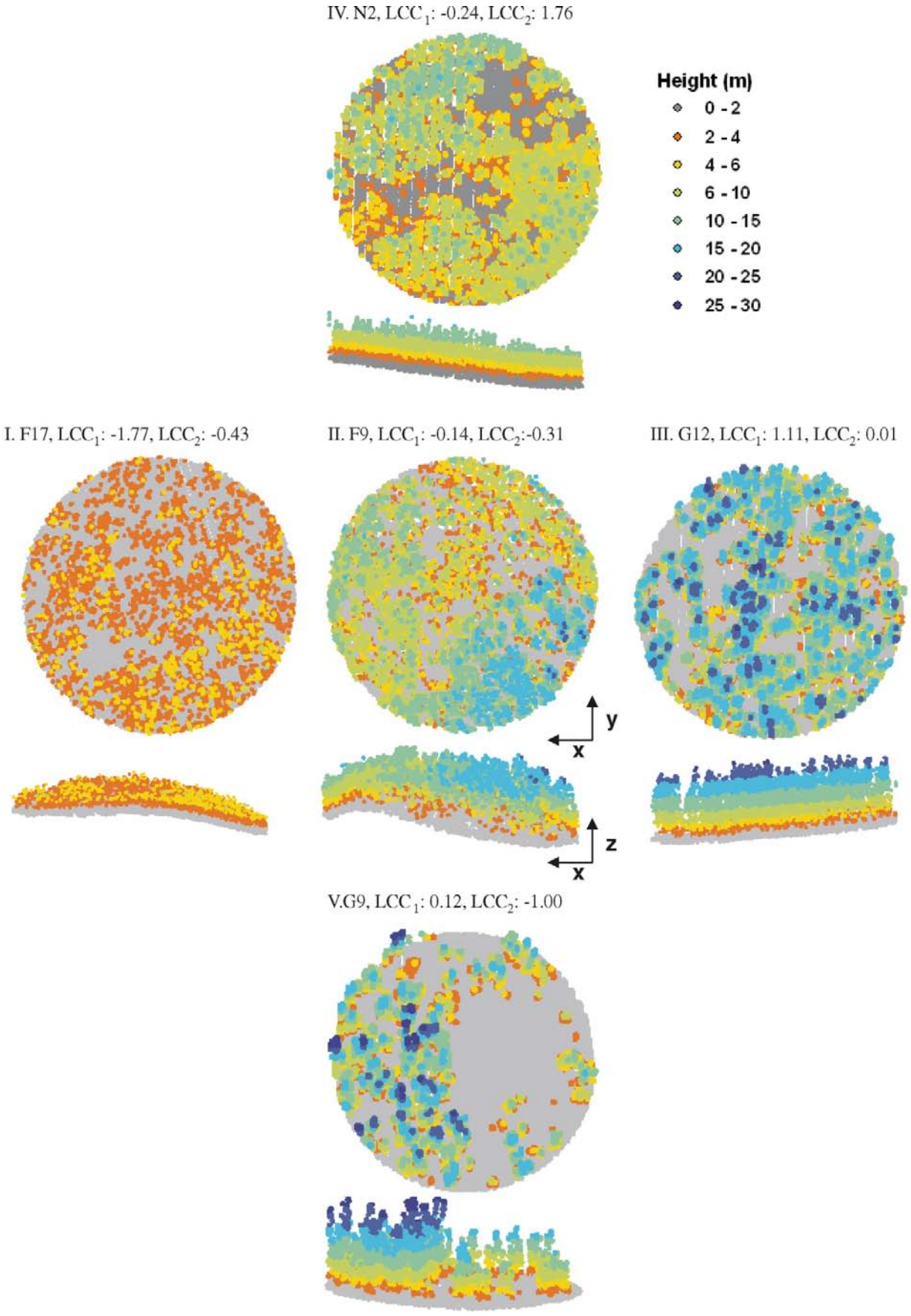
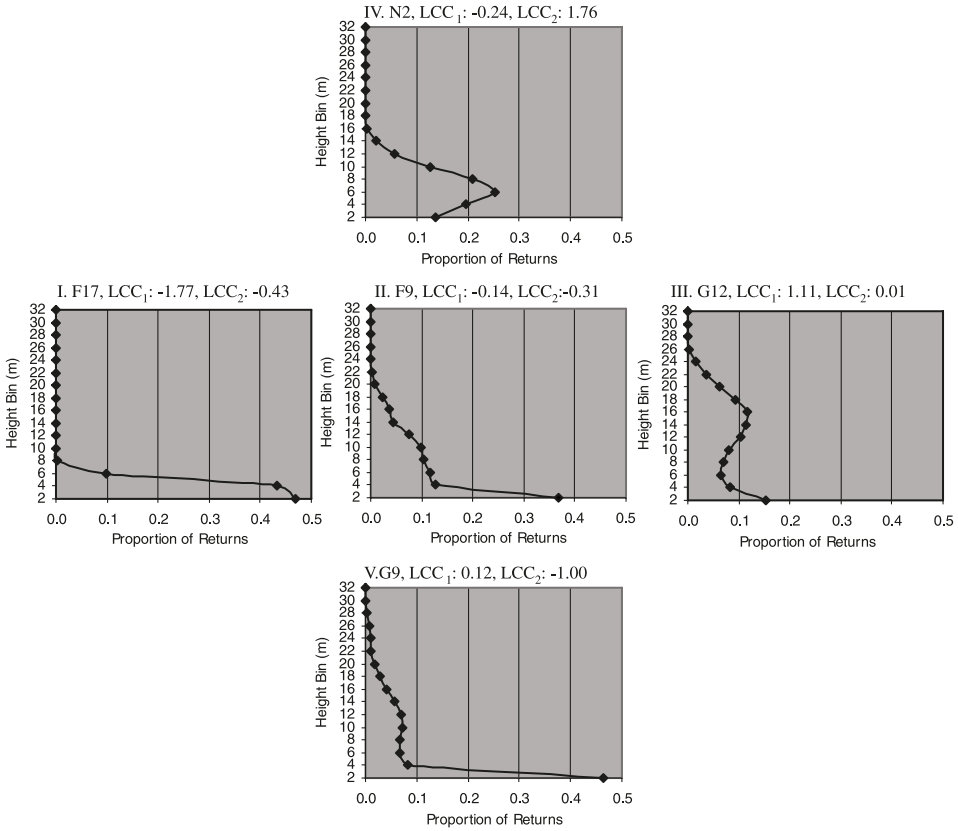


Fig. 6. Vertical return distributions for the five selected field plots in Fig. 4 representing the range of observed LCC_1 and LCC_2 values.



research at the Fraser, Niwot, and Glees sites is studying the relationship between lidar-derived height-variability measures and understory variables such as carbon in CWD, and carbon in stumps (Barry et al.[2]).

Lidar canonical variables 1 and 2 visual assessments

Five study plots were selected from the graph of the LCC_1 and LCC_2 values by plot location (Fig. 4). For each selected plot the coincident lidar point cloud and vertical return distributions are examined (Figs. 5 and 6). Vertical return distributions (apparent foliage profiles) have been used to evaluate differences in vertical and horizontal canopy structure across structurally diverse forest plots (Coops et al. 2007). Visualization was performed to assess the ability of the lidar height data to qualitatively assess the LCC_1 and LCC_2 values and to increase understanding of the loadings on the first two canonicals. Mature plots with high LCC_1 loadings and high forest heights should be readily visible in the lidar returns. Intermediate-aged plots with high tree density may be more difficult to visually recognize. Of particular interest is whether the lidar point cloud data can be used to differentiate high and low tree densities.

Point I (Figs. 5 and 6) has a negative loading for both LCC_1 and LCC_2 with values of –1.77 and –0.43. Visually the loadings for both LCC_1 and LCC_2 are apparent. This site has recently been logged (1983 clearcut) and is composed of young regeneration (mean tree age 14 years) with a majority of lidar returns within 7 m of the ground surface. This plot is short in height with low tree density (187 trees·ha^{-1}) and has low total basal area and tree biomass values that contribute to the high negative LCC_1 value. Point I has low tree density (reduced LCC_2) but high lodgepole pine basal area (BaLP = 1) (positive LCC_2) offsetting each other to yield a moderately negative LCC_2 value.

Point II is a mature plot (mean tree age of 180 years) with LCC_1 and LCC_2 values of –0.14 and –0.31. This plot has areas of moderate forest height (10–21 m), which is observable in the southeastern quarter of the plot. The remaining plot areas are dominated by pockets of low to moderate forest height (2–12 m), contributing to the near-neutral LCC_1 value. For the point II plot visually, it is difficult to determine the plot's tree density. Similar to point I, point II has high basal area of lodgepole pine (BaLP = 0.53), which increases the LCC_2 value, and conversely the lower relative tree density (721 trees·ha^{-1}) decreases the LCC_2 value.

Point III is a mature plot with a mean tree age of 205 years, with LCC_1 and LCC_2 values of 1.11 and 0.01. This site appears to be a relatively tall forest with many returns >18 m and little apparent undergrowth along the forest floor. Visual inspection confirms that this is a mature stand with dominant trees between 18 and 27 m tall. Comparison of points III's and II's vertical return distribution (Fig. 6) shows that point III has a greater proportion of returns in the middle to upper canopy above 10 m. A site having little to no loading on LCC_2 is expected to have moderate tree density. From the point cloud, it can be seen that the plot does have forest gaps with little to no understory vegetation (saplings and seedlings) interspersed with

[2] P.J. Barry and M.A. Lefsky. Forest CWD relationship to Lidar-derived canopy structure. In preparation.

areas having some lower height (<10 m), likely sapling sized trees, overall helping to explain the neutral LCC_2 value.

Point IV has LCC_1 and LCC_2 values of –0.24 and 1.76 and is a young stand (mean tree age of 48 years). Visual examination of both the point cloud and vertical return distribution shows an intermediate aged forest with low to moderate forest heights (~2–17 m), which is in agreement with the slightly negative LCC_1 value. This plot has a high tree density (1679 trees·ha^{-1}) and a high proportion of returns in the lower to middle canopy range (2–10 m) as indicated in the vertical return distribution. The point cloud indicates that the majority of the plot consists of lower to middle canopy returns associated with intermediate stand age development. As previously mentioned, this young-aged plot (<50 years) with a high LCC_2 value is an outlier from the other young aged plots because it has a high tree density and is closer in forest structure to an intermediate-aged plot.

Point V is intermediate in age (mean tree age 141 years) with LCC_1 and LCC_2 values of 0.12 and –1.00. From the point cloud visually, it is apparent that this plot has high horizontal spatial variability, with the western half consisting of a mature forest and the eastern half being predominately nonforested. Point V's vertical return distribution in the midcanopy ranges between 2 and 10 m is noticeably lower than point IV's (high positive LCC_2) with all bin percentages being less than 0.082. The large trees that do occur at this plot are tall (~17–27 m) and old (maximum tree age 332 years); however, the density is low (211 trees·ha^{-1}).

Examination of these point clouds indicates the importance of horizontal spatial variability in forest structure. Within plots, there can be considerable variance in forest height. For example, point V goes from bare ground in the eastern half of the plot to relative low-density mature forest in the western half. It is important to note that this spatial variability can result in misleading canonical values. Although not included in this paper, it is worth mentioning that subplot level analyses were performed and resulted in findings similar to those of the plot level indicating that, although visually heterogeneous, the same patterns we examine here are applicable at the subplot scale as well.

Conclusions

This study applies an existing methodology (Lefsky et al. 2005*a*) for analysis of forest canopy complexity and subsequent regression analysis, to lidar data sets with discrete return lidar-derived canopy indices. Graphs of the correlation between the first two canonical variables and field-derived forest structure variables reveal distinct groupings of forest structure. Visualization techniques help increase the interpretability of the derived canonicals.

Modeling results support the use of explanatory variables derived from canonical correlation analysis in the modeling of field-derived forest structure and forest biomass variables at three subalpine forest sites, suggesting that CCA can be used for enhancement of lidar data (Lefsky et al. 2005*a*) similar to tasseled cap indices (Crist and Cicone 1984) for other optical remote sensing data types. Overall, the best modeling results (slightly higher R^2 and lower RMPV) are achieved using the independent lidar indices. Of particular interest is the successful development of regression equations for the forest height (R^2 = 0.93) and biomass-related variables (R^2 > 0.92 for all variable, and R^2 > 0.79 for CCA models).

Acknowledgements

This research was supported by funding from the USDA Forest Service Northern Global Change Program and US NASA Carbon Cycle Science research grant CARBON/04-0225-0191.

References

Arthur, M.A., and Fahey, T.J. 1990. Mass and nutrient content of decaying boles in an Engelmann spruce – subalpine fir forest, Rocky Mountain National Park, Colorado. Can. J. For. Res. **20**: 730–737. doi:10.1139/x90-096.

Beers, T.W., Dress, P.E., and Wensel, L.C. 1966. Aspect transformation in site productivity research. J. For. **64**: 691–692.

Birdsey, R., Kolka, R., Smith, M.-L., Ryan, M., Hollinger, D., Heath, L., and Hoover, C. 2004. Landscape carbon monitoring and analysis at the Experimental Forest Network. *In* Proceedings, 3rd Annual Conference on Carbon Sequestration, 3–6 May 2004, Alexandria, Va.

Bradford, J.B., Birdsey, R.A., Joyce, L.A., and Ryan, M.G. 2008. Tree age, disturbance history, and carbon stocks and fluxes in subalpine Rocky Mountain forests. Glob. Change Biol. In press.

Brown, J.K. 1971. Planar intersect method for sampling fuel volume and surface area. For. Sci. **17**: 96–102.

Busse, M.D. 1994. Downed bole-wood decomposition in lodgepole pine forests of central Oregon. Soil Sci. Soc. Am. J. **58**: 221–227.

Chen, Q. 2007. Airborne lidar data processing and information extraction. Photogramm. Eng. Remote Sens. **73**: 109–112.

Cohen, W.B., Harmon, M.E., Wallin, D.O., and Fiorella, M. 1996. Two decades of carbon flux from forests of the Pacific Northwest. Bioscience, **46**: 836–844. doi:10.2307/1312969.

Coops, N.C., Hilker, T., Wulder, M.A., St-Onge, B., Newnham, G., Siggins, A., and Trofymow, J.A. 2007. Estimating canopy structure of Douglas-fir forest stands from discrete-return LiDAR. Trees (Berl.), **21**: 295–310. doi:10.1007/s00468-006-0119-6.

Crist, E.P., and Cicone, R.C. 1984. A physically-based transformation of thematic mapper data—the TM tasseled cap. IEEE Trans. Geosci. Remote Sens. **22**: 256–263. doi:10.1109/TGRS.1984.350619.

Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, B., Hofton, M.A., Cahzdon, R.L., Weishampel, J.F., and Prince, S.D. 2002. Estimation of tropical forest structural characteristics using large-footprint lidar. Remote Sens. Environ. **79**: 305–319. doi:10.1016/S0034-4257(01)00281-4.

Hall, S.A., Burke, I.C., Box, D.O., Kaufmann, M.R., and Stoker, J.M. 2005. Estimating stand structure using discrete-return lidar: an example from low density, fire prone ponderosa pine forests. For. Ecol. Manage. **208**: 189–209. doi:10.1016/j.foreco.2004.12.001.

Hyde, P., Nelson, R., Kimes, D., and Levine, E. 2007. Exploring LiDAR–RaDAR synery-predicting aboveground biomass in a southwestern ponderosa pine forest using LiDAR, SAR and InSAR. Remote Sens. Environ. **106**: 28–38. doi:10.1016/j.rse.2006.07.017.

Isebrands, L.G., and Crow, T. 1975. Introduction to uses and interpretation of principal component analysis in forest biology. USDA For. Serv. Gen. Tech. Rep. NC-17.

Kikkawa, J. 1982. Ecological association of birds and vegetation in wet tropical forests of Australia. Aust. J. Ecol. **7**: 325–345.

Lefsky, M.A. 1997. Applications of lidar remote sensing to the estimation of forest canopy and stand structure. Ph.D. dissertation, University of Virginia, Charlottesville, Va.

Lefsky, M.A., Harding, D., Cohen, W., Parker, G., and Shugart, H. 1999. Surface Lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA. Remote Sens. Environ. **67**: 83–98. doi:10.1016/S0034-4257(98)00071-6.

Lefsky, M.A., Cohen, W.B., and Spies, T.A. 2001. An evaluation of alternate remote sensing products for forest inventory, monitoring, and mapping in Douglas-fir forests of western Oregon. Can. J. For. Res. **31**: 78–87. doi:10.1139/cjfr-31-1-78.

Lefsky, M.A., Cohen, W.B., Harding, D.J., Geoffrey, G.P., Acker, S.A., and Gower, S.T. 2002. Lidar remote sensing of aboveground biomass in three biomes. Glob. Ecol. Biogeogr. **11**: 393–399. doi:10.1046/j.1466-822x.2002.00303.x.

Lefsky, M.A., Hudak, A.T., Cohen, W.B., and Acker, S.A. 2005*a*. Patterns of covariance between forest stand and canopy structure in the Pacific Northwest. Remote Sens. Environ. **95**: 517–531. doi:10.1016/j.rse.2005.01.004.

Lefsky, M.A., Hudak, A.T., Cohen, W.B., and Acker, S.A. 2005*b*. Geographic variability in lidar predictions of forest stand structure in the Pacific Northwest. Remote Sens. Environ. **95**: 532–548. doi:10.1016/j.rse.2005.01.010.

Lim, K.S., and Treitz, P.M. 2004. Estimation of above ground forest biomass from airborne discrete return laser scanner data using canopy-based quantile estimators. Scand. J. For. Res. **19**: 558–570. doi:10.1080/02827580410019490.

MacArthur, R., Recher, H., and Cody, M. 1966. On the relation between habitat selection and species diversity. Am. Nat. **100**: 319–332. doi:10.1086/282425.

Magnussen, S., and Boudewyn, P. 1998. Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators. Can. J. For. Res. **28**: 1016–1031. doi:10.1139/cjfr-28-7-1016.

McGarigal, K., Cushman, S., and Stafford, S. 2000. Multivariate statistics for wildlife and ecology research. Springer-Verlag, New York.

Morsdorf, F., Kötz, B., Meier, E., Itten, K.I., and Allgöwer, B. 2006. Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. Remote Sens. Environ. **104**: 50–61. doi:10.1016/j.rse.2006.04.019.

Næsset, E. 1997. Estimating timber volume of forest stands using airborne laser scanner data. Remote Sens. Environ. **61**: 246–253. doi:10.1016/S0034-4257(97)00041-2.

Pacala, S.W., Hurtt, G.C., Baker, D., Peylin, P., Houghton, R.A., Birdsey, R.A., Heath, L., Sundquist, E.T., Stallard, R.F., Ciais, P., Moorcroft, P., Caspersen, J.P., Shevliakova, E., Moore, B., Kohlmaier, G., Holland, E., Gloor, M., Harmon, M.E., Fan, S.M., Sarmiento, J.L., Goodale, C.L., Schimel, D., and Field, C.B. 2001. Consistent land- and atmosphere-based U.S. carbon sink estimates. Science (Washington, D.C.), **292**: 2316–2320. doi:10.1126/science.1057320.

Parker, G.G., Harmon, M.E., Lefsky, M.A., Chen, J., Van Pelt, R., Weiss, S.B., Thomas, S.C., Winner, W.E., Shaw, D.C., and Franklin, J.F. 2004. Three-dimensional structure of an old-growth *Pseudotsuga–Tsuga* canopy and its implications for radiation balance, microclimate, and gas exchange. Ecosystems (N.Y. Print), **7**: 440–453.

Patenaude, G., Hill, R.A., Milne, R., Gaveau, D.L.A., Briggs, B.B.J., and Dawson, T.P. 2004. Quantifying forest above ground carbon content using LiDAR remote sensing. Remote Sens. Environ. **93**: 368–380. doi:10.1016/j.rse.2004.07.016.

Popescu, S.C., Wynee, R.H., and Nelson, R.F. 2002. Estimating plot-level tree heights with lidar: local filtering with a canopy-height based variable window size. Comput. Electron. Agric. **37**: 71–95. doi:10.1016/S0168-1699(02)00121-7.

Popescu, S.C., Wynne, R.H., and Nelson, R.F. 2003. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. Can. J. Remote Sens. **29**: 564–577.

Riaño, D., Valladares, F., Condes, S., and Chuvieco, E. 2004. Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests. Agric. For. Meteorol. **124**: 269–275.

Running, S.W., Baldocchi, D.D., Turner, D.P., Gower, S.T., Bakwin, P.S., and Hibbard, K.A. 1999. A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data. Remote Sens. Environ. **70**: 108–127. doi:10.1016/S0034-4257(99)00061-9.

Solberg, S., Næsset, E., Hanssen, K.H., and Christiansen, E. 2006. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. Remote Sens. Environ. **102**: 364–376. doi:10.1016/j.rse.2006.03.001.

Thomas, V., Treitz, P., McCaughey, J.H., and Morrison, I. 2006. Mapping stand-level forest biophysical variables for a mixedwood boreal forest using lidar: an examination of scanning density. Can. J. For. Res. **36**: 34–47. doi:10.1139/x05-230.

US Department of Agriculture (USDA). 2005. Forest inventory and analysis national core field guide, version 3.0 edition. USDA, Washington, D.C.

Waring, R.H., Way, J., Hunt, R., Morrissey, L., Ranson, K.J., Weishampel, J.F., Oren, R., and Franklin, S.E. 1995. Imaging radar for ecosystem studies. Bioscience, **45**: 715–723. doi:10.2307/1312677.

Weiss, B.S. 2000. Vertical and temporal distribution of insolation in gaps in an old-growth coniferous forest. Can. J. For. Res. **30**: 1953–1964. doi:10.1139/cjfr-30-12-1953.

Zhang, K., Chen, S.-C., Whitman, D., Shyu, M.-L., Yan, J., and Zhang, C. 2003. A progressive morphological filter for removing nonground measurements from airborne LIDAR data. IEEE Trans. Geosci. Remote Sens. **41**: 872–882. doi:10.1109/TGRS.2003.810682.