# Discussion of Regional Models

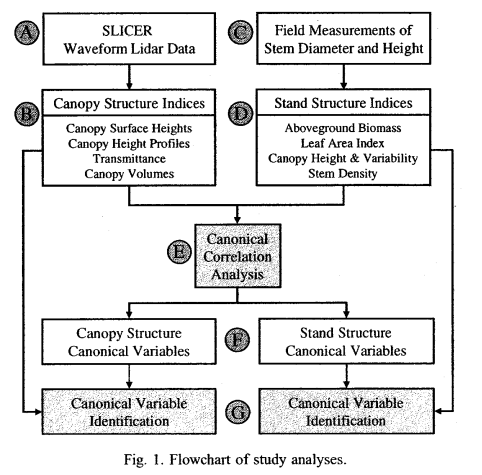
* + See Lefsky et al 2005, Lefsky et al 2002

Previous work tested the generality of lidar relationships over multiple sites (PNW Lefsky et al., 2005 A), and multiple forest cover types (high biomass forests Lefsky et al., 2002). 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. Lefsky et al 2005 A we were able to look at 5 sites within the Temperate Coniferous Needleleaf biome, using sites with varying environment and composition. They created equations that predicted stand structure variables (e.g. aboveground biomass and LAI) across an environmental and compositional gradient.

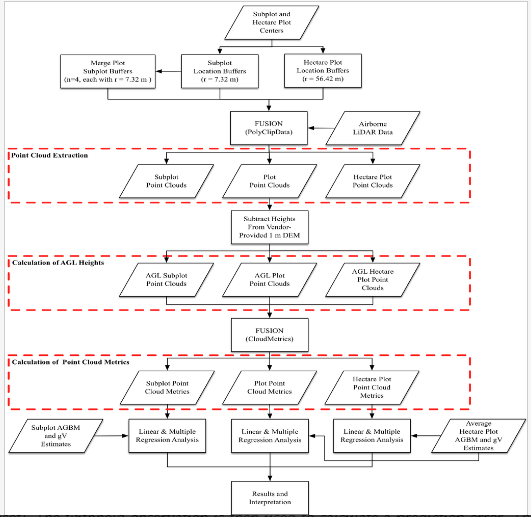
Aboveground biomass were consistently predictable along a productivity and species composition gradient from the true fir forests of Mt. Rainier to Ponderosa pine forests at Metolius, and at the high productivity forests of Cascade Head, the coast range, and H.J. Andrews (Lefsky et al 2005 A). (Lefsky et al 2005 A and 2002) offer 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. Lefsky et al 2005 A found broad consistency in lidar-stand structure relationship over this region, and a relative lack of importance of environmental conditions (Lefsky et al 2005 A). 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 (Lefsky et al 2005 A).

Given this wide range of conditions, and the earlier results of Lefsky et al. (2002 and 2005 A), 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 (Lefsky et al. 2002 and 2005 A) there has been an attempt to have a structural or temporal sequence of stands at one or more study locations.

Diagram:

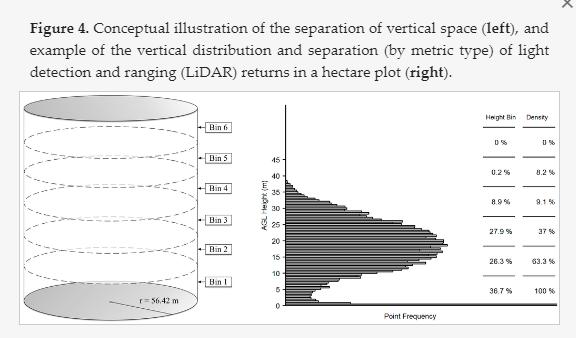
Lefsky et al 2005 B

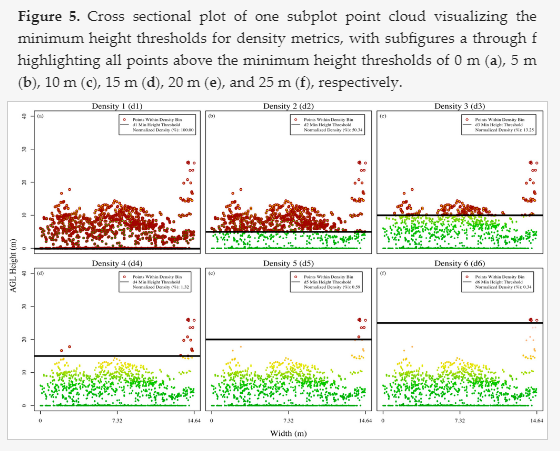
Sheridan workflow diagram:



* Covariate selection:
* Lidar instruments
  + In addition, waveform-recording lidar systems, such as the SLICER (Scanning Lidar Imager of Canopies by Echo Recovery, Blair et al., 1994; Harding et al., 1994, 2001). canopy surface height measurements, only uses the instrument's height measuring capability. A second set of measurements was made by transforming the raw waveform data into an estimate of the vertical distribution of the canopy- the canopy height profile (CHP). A third set of measurements described the transmittance of light in the canopy. A fourth was derived from a system for the measurement of canopy structure, the canopy volume method (CVM), which summarizes the total volume and spatial organization of filled and empty space within the canopy. Details of these methods can be found in Lefsky et al. (1999b).
  + Discrete-return lidar emits a small beam of light (centimeters in diameter), and records the positions from which the returned energy is greater than a certain threshold. Different systems can record from 1 to 5 discrete returns from each laser pulse. These systems are the ones commercially available (Baltsavias, 1999b), and are being used routinely to develop digital elevation models (Flood and Gutelius, 1997). Both types of lidar have been used successfully to estimate stand structural variables, such as mean height, total aboveground biomass, basal area, stem volume and stand density, in a variety of forest types (Means et al., 1999, 2000; Lefsky et al., 1999a,b, 2002a; Dubayah and Drake, 2000; Drake et al., 2002, 2003; Naesset and Bjerknes, 2001; Naesset and Økland, 2002; Naesset, 2002). The focus of most of the research has been on overcoming the saturation at high biomass levels that limits the use of passive sensors (Hall 2005)
* Variation by forest types:
  + different age classes exhibit distinct laser canopy height distribution forms [(Lefsky, Cohen, et al., 1999)](http://www.sciencedirect.com/science/article/pii/S0034425701002905" \l "BIB15).
  + Relationships of stand characteristics to lidar metrics vary between tree species [(Næsset, 1997a; Nelson, Oderwald, & Gregoire, 1997)](http://www.sciencedirect.com/science/article/pii/S0034425701002905" \l "BIB23)
  + crown shape differs between species and translates to differences in lidar metrics [(Nelson, 1997)](http://www.sciencedirect.com/science/article/pii/S0034425701002905#BIB28).
  + There were few patterns in the correlations that were consistent between all three biomes. This is due probably to the narrow range of canopy cover conditions observed in those plots. For the boreal coniferous site, the product of cover and several of the height indices performed better than the height indices alone. At the temperate deciduous site, the reverse was true, again due probably to the low range of canopy cover, and the resulting non-significant correlation between cover and biomass. At the temperate coniferous site there is no clear difference between the two sets of variables (heights alone and the products of height indices and cover indices). When all sites are considered together, mean height squared is the best overall predictor of above-ground biomass. (Lefsky et al 2002)
  + The equation for deciduous basal area failed at H.J. Andrews and Metolius, two sites with deciduous basal area less than 1.0 m2 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 (Lefsky et al 2005 A)
  + 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), temperate subalpine (Sherrill et al 2008)
* Stratified sample –
  + Commonly used stratification criteria such as age class and site quality, which is correlated with at least the dominant conifer tree species in Norway, may therefore be useful for an efficient stratification of inventories. In the present study, stratification according to age class and site quality was efficient to represent distinct forest types in practical inventories (Naesset 2002).
* Statistical approach
  + Issues influenced the analysis of this data set: 1) the potential for multicollinearity when a large number of independent variables are involved; 2) the choice of variables for, and the comprehensibility of, regression equations.
* Collinearity
  + Variance inflation factors (VIFs) were calculated to check for the presence of multicollinearity among the remaining predictor variables in the model. Predictor variables with VIFs greater than ten were considered an indicator of multicollinearity in the model [Sheskin 2007].
  + Latent variables – overfitting may occur, independent (validation) data used to determine the optimal number of latent variables (?? Maybe Naesset e tal 2005)
  + 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. 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 general trend in forest stand structure. Previous results (Lefsky et al. 2005a, 2005b) have indicated that using CCA components as independent variables results in more interpretable and parsimonious regression models. Sherrill et al 2008
  + Principal component analysis based on the correlation matrix was used to assess the presence of collinearity in the regression analysis. The square root of the largest eigenvalue divided by the smallest eigenvalue (condition number, κ) was used as a means for suggesting collinearity. A condition number larger than 30 has been proposed to indicate collinearity [(Weisberg, 1985)](http://www.sciencedirect.com/science/article/pii/S0034425701002905" \l "BIB34). Some of the models initially suggested by the stepwise selection procedure were subject to serious collinearity. The models selected for further analysis were therefore those indicated by the stepwise procedure that fulfilled the requirement of κ<30. The maximum condition number of the selected models was 11.7 [(Table 6)](http://www.sciencedirect.com/science/article/pii/S0034425701002905" \l "TBL6), which indicated no serious collinearity problems of the final models. (Naesset 2002)
* Lidar covariates:
  + In this work, canonical correlation analysis of coincident lidar and field datasets in western Oregon and Washington is used to define seven statistically significant pairs of canonical variables, each defining an axis of variation that stand and canopy structure have in common. The first major axis relates mean stand height, and related variables, to aboveground biomass. The second relates canopy cover and volume to leaf area index and stem density. The third relates canopy height variability to mean stem diameter and the basal area of deciduous species. Of the four remaining axes, three are related to contrasts between mature and old-growth stands. Canonical correlation analysis provides a method for ranking the importance of these effects, and for placing both canopy and stand structure indices within the overall covariance structure of the two datasets. In this sense, and for the study area involved, the first three factors (mean height, cover or leaf index area, height variability) represent the same kind of enhancement of lidar data that the tasseled cap indices [Crist, C.P., R.C. Cicone, f 984. A physically-based transformation of thematic mapper data-the TM tasseled cap. Lefksy et al 2005 B.

Nice graphics by Sheridan:

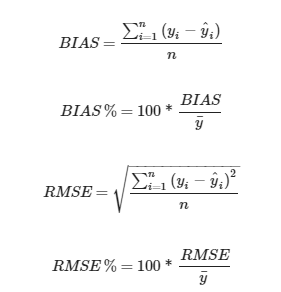


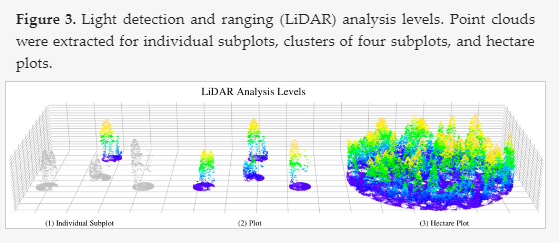


* Environmental covariates:
  + Three sets of statistical analyses were performed. The first set of analyses compared three methods for relating lidar-measured canopy structure and field-measured stand structure. The second set of analyses tested the ability of environmental (topographic, climate and edaphic) indices to explain the residuals from the first set of regression analyses. The second analysis involved a second round of CCA to relate residuals from each of the three regression analyses to topography, climate and soils. 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. (Lefsky et al 2005 A).
* Model Selection:
  + 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 variables (referred to here as SCV). (Lefsky et al 2005 A)
  + Canonical correlation analysis: Lefsky et al 2005 B use canonical correlation analysis to define pairs of canonical variables, each defining an axis of variation common to the canopy and stand structure datasets. In this way, the ranking of various effects can be understood as they relate to the explanation of variance in each dataset, and axes of variation that connect forest canopy and stand structure can be rigorously defined. Their goals are a better understanding of what each variable represents in the context of a suite of variables that define an axis of variation, and an understanding of what each axis of variation represents within the context of numerous axes of variation. In this way, we hope to avoid placing too much importance on any particular
* Models:
  + Naesset 2002
    - Young stands: mean height of last returns, proportion last returns > median/all
    - Mature, poor quality: 30th percentile of first; 50th percentile of first; proportion first returns > 60th percentile/all; proportion first returns > 90th percentile /all
    - Mature, good quality: 80th percentile of last; max first; proportion first returns > 60th percentile/all; proportion first returns > 90th percentile /all; proportion first returns > 50th percentile /all
* Goodness of fit:

Close agreement in magnitude between RMSE and RMSEcrossval suggests that the model is not overfit and suitable for generalization.

Iid - linear issues: Over predicted low values (Li et al 2008, Kim et al 2009)



* Plot size
  + The standard deviations between predicted and ground-truth values of the plots revealed by cross-validation were up to 160% larger than the corresponding standard deviations obtained for entire stands. The relationship between precision on small plots and entire stands illustrates two important aspects. First, small areas are subject to substantial inherent variation around canopy height quantiles leading to highly variable predictions (cf. [Magnussen & Boudewyn, 1998](http://www.sciencedirect.com/science/article/pii/S0034425701002905" \l "BIB17)). The size of the sample plots and stand grid cells should therefore not be too small. On the other hand, extended plot size will increase the inventory costs. Thus, traditional cost/benefit analysis could be used to balance precision and costs. (Naesset 2002)
  + Plot size has a strong influence on the model error (RMSE), with larger plots typically resulting in lower errors, although above 0.1 ha there is little influence on model coefficients (Mascaro et al. [2011b](https://link.springer.com/article/10.1007/s00442-011-2165-z#CR40)).
  + See also Zolkos, Sheridan (image below)
  + 
* Plots with small biomass, tree dbh threshold
  + The consequence of excluding trees with a small dbh from plots is that the biomass estimate may not be representative of the distributions of laser canopy heights, especially those with few trees.
* Coefficient of determination:

* + [Means et al. (2000)](http://www.sciencedirect.com/science/article/pii/S0034425701002905" \l "BIB19)reported R2 values for volume at .95–.97 for 2500 m2 plots:
  + 0.80–.93 200 m2 plots (Naesset 2002).
  + estimation of volume in stands with an average area of 1.5 ha [(Næsset, 1997b)](http://www.sciencedirect.com/science/article/pii/S0034425701002905" \l "BIB24) indicated smaller proportions of explained variation (46–89%) than the current trial.

Lidar biomass estimation reviews:

Zolkos et al 2013

S.J. Goetz, R.O. Dubayah. Advances in remote sensing technology and implications for measuring and monitoring forest carbon stocks and change Carbon Management, 2 (3) (2011), pp. 231-244.

Petrokofsky, G., Kanamaru, H., Achard, F., Goetz, S. J., Joosten, H., Holmgren, P., ... & Wattenbach, M. (2012). Comparison of methods for measuring and assessing carbon stocks and carbon stock changes in terrestrial carbon pools. How do the accuracy and precision of current methods compare? A systematic review protocol. *Environmental Evidence*, *1*(1), 6.

Rosette, J. Suárez, R. Nelson, S. Los, B. Cook, P. North Lidar remote sensing for biomass assessment Remote Sensing of Biomass — Principles and Applications (2012), p. 24

M.A. Wulder, J.C. White, R.F. Nelson, E. Næsset, H.O. Ørka, N.C. Coops, et al. Lidar sampling for large-area forest characterization: A review Remote Sensing of Environment, 121 (2012), pp. 196-209

forest and lidar over view:

R. Dubayah, J.B. Drake Lidar remote sensing for forestry Journal of Forestry, 98 (2000), pp. 44-46

Other:

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

Brown, P.M., D’Amico, D.R., Carpenter, A.T., Andrews, D., 2001. Restoration of montane ponderosa pine forests in the Colorado Front Range. Ecol. Restoration 19, 19–26.

Flannigan, M.D., Van Wagner, C.E., 1991. Climate change and wildfire in Canada. Can. J. Forest Res. 21, 66–72.

Fule´, P.Z., Waltz, A.E.M., Covington, W.W., Heinlein, T.A., 2001. Measuring forest restoration effectiveness in reducing hazardous fuels. J. Forestry 99, 24–29.

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 closedcanopy br

Houghton, R.A., Hackler, J.L., Lawrence, K.T., 2000. Changes in terrestrial carbon storage in the United States. 2. The role of fire and fire management. Global Ecol. Biogeogr. 9, 145–170

Kaufmann, M.R., Regan, C.M., Brown, P.M., 2000. Heterogeneity in ponderosa pine/Douglas fir forests: age and size structure in unlogged and logged landscapes of central Colorado. Can. J. Forest Res. 30, 698–711.M.A. Lefsky, D. Harding, W.B. Cohen, G. Parker, H.H. Shugart**Surface lidar remote sensing of basal area and biomass in deciduous forests of eastern Maryland, USA.** Remote Sensing of Environment, 67 (1999 A), pp. 83-98

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

S. Magnussen, P. Boudewyn**Derivations of stand heights from airborne laser scanner data with canopy-based quantile estimators.** Canadian Journal of Forest Research, 28 (1998), pp. 1016-1031

J.E. Means, S.A. Acker, J.F. Brandon, M. Renslow, L. Emerson, C.J. Hendrix**Predicting forest stand characteristics with airborne scanning lidar**. Photogrammetric Engineering and Remote Sensing, 66 (2000), pp. 1367-1371

Moore, M.M., Covington, W.W., Fule´, P.Z., 1999. Reference conditions and ecological restoration: a southwestern ponderosa pine perspective. Ecol. Appl. 9, 1266–1277.

R. Nelson **Modeling forest canopy heights: the effects of canopy shape**

Remote Sensing of Environment, 60 (1997), pp. 327-334

R. Nelson, R. Oderwald, G. Gregoire**Separating the ground and airborne laser sampling phases to estimate tropical forest basal area, volume, and biomass**

Remote Sensing of Environment, 60 (1997), pp. 311-326

Overpeck, J.T., Rind, D., Goldberg, R., 1990. Climate-induced changes in forest disturbance and vegetation. Nature 343, 51–53.

Perry, G.L.W., 1998. Current approaches to modelling the spread of wildland fire: a review. Prog. Phys. Geogr. 22, 222–245.

Weaver, H., 1959. Ecological changes in the ponderosa pine forest of the Warm Springs Indian Reservation on Oregon. J. Forestry 57, 15–20.

S. Weisberg Applied linear regression (2nd ed.), Wiley, New York (1985) 324 pp.

Beers, T. W., Dress, P. E., & Wensel, L. C. (1966). Notes and observations: aspect transformation in site productivity research. *Journal of Forestry*, *64*(10), 691-692.