INTEGRATION OF LIDAR, LANDSAT ETM+ AND FOREST INVENTORY DATA FOR REGIONAL FOREST MAPPING.

Michael A. Lefsky, Warren B. Cohen, Andrew Hudak, Steven A. Acker, Janet L. Ohmann
Forest Sciences Laboratory
USDA Forest Service & Oregon State University
3200 SW Jefferson Way
Corvallis, OR, USA
lefsky@fsl.orst.edu

KEY WORDS: Lidar, Forest Inventory, SLICER, LVIS, VCL

ABSTRACT

Recent work has established the utility of waveform sampling lidar for predicting forest structural attributes. Nevertheless, serious obstacles to its wide-spread use still exist. They include the lack of waveform sampling lidar sensors capable of measuring forest canopy structure over large extents, and the practical difficulty of developing widely applicable relationships to predict forest stand structure indices (such as aboveground biomass) from measurements of canopy structure. While the advent of advanced devices such as NASA's LVIS and VCL sensors will allow the collection of larger datasets than previously possible, neither sensor is capable of collecting spatially comprehensive datasets at the regional scales critical for forest management. Therefore, methods to integrate data from these devices with conventional optical remote sensing products such as those from the Landsat Enhanced Thematic Mapper Plus (Landsat ETM+) sensor will play a critical role in the development of lidar remote sensing. In addition, it will be desirable to develop methods to facilitate the use of existing forest inventory datasets, which include substantial information on the height of dominant trees, to interpret lidar measurements of canopy structure. Preliminary results from a new study that incorporates both approaches (Lidar-ETM+ data fusion, incorporating forest inventory data for the interpretation of lidar data) to accomplish wall-to-wall mapping of forest structure and composition in Oregon and Washington suggest that these approaches are feasible, and that they will lead to more accurate maps of forest structure and composition.

1. INTRODUCTION

Lidar Remote Sensing

A variety of height-measuring remote sensing technologies are currently in development through NASA and other government and commercial remote sensing programs. These include small footprint laser altimeters, large footprint waveform digitizing lidar sensors (such as SLICER, Scanning Lidar Imager of Canopies by Echo Recovery, Harding et al., 1995; LVIS and VCL, Laser Vegetation Imaging Sensor and Vegetation Canopy Lidar, Dubayah, 1997), and inteferrometric SAR (Synthetic Aperture Radar) devices. These sensors have considerable advantages over traditional optical or radar remote sensing systems for the description of forested ecosystems. In particular, it has been demonstrated that laser altimetry and lidar systems (referred to here collectively as lidar sensors) are able to predict biophysical parameters, such as leaf area index (LAI) and aboveground biomass, with very high accuracy and without the asymptotic effect noted with other sensors (Lefsky et al. 1999a, Lefsky et al. 1999b). Recognizing the great potential utility of newer laser devices, NASA recently funded an Earth System Sciences Pathfinder satellite lidar mission—VCL. VCL, with a planned launch date of 2000, will obtain measurements from approximately 5% of the global land surface. This dataset, and similar datasets likely to follow, will greatly enhance our ability to map and monitor global forest ecosystems. The ETM+LIFE (Enhanced Thematic Mapper Plus Lidar for Forest Ecosystems) project is funded by NASA to address the challenges is using data from VCL for regional-scale mapping of forest structure.

Key Challenges in Regional Application of Lidar Remote Sensing

Despite the advantages that lidar remote sensing devices have for measuring forest structure, there are two key challenges in its widespread use for regional-scale forest mapping. First, there is a lack of documented relationships between the elements of forest canopy structure measured by lidar sensors, and the measurements that forest managers and researchers need; measurements such as stem basal area and volume, aboveground biomass, and leaf area index. The second challenge is the lack of systems for collecting spatially comprehensive, regional-scale lidar datasets. Until both of these challenges are addressed, forest inventory using lidar remote sensing will be severely limited.

Predicting Forest Structure

From our experience with lidar sensors, and the existing literature, we expect that it will, ultimately, be possible to use data from VCL to create a very accurate, spatially detailed (but not comprehensive) global dataset of forest stand structure attributes, including estimates of aboveground biomass and LAI. Such a database will be directly useful for global climate and carbon balance investigators. However, a substantial challenge in using VCL data will be the lack of documented relationships between lidar measurements of canopy structure and the forest structure attributes of interest. To develop these relationships,

existing studies with these devices have adopted a study design in which field measurements from a specific locale are directly compared to coincident remotely sensed data. While this design has been useful in demonstrating the potential of lidar sensors, it has numerous drawbacks. By studying a relatively small number of plots from a restricted area, the accuracy and precision in predicting forest stand attributes may be overestimated, both by the small sample size and the relative uniformity of species composition and environmental conditions over these small study areas. Studies over larger areas will be necessary to determine the actual accuracy of predictions of forest stand attributes using lidar measurements. In particular, differences in forest composition and productivity are likely to strongly influence the relationship between lidar-measured height and aboveground biomass. Given the global coverage of VCL, any attempt to generate these relationships using coincident field plots would be extraordinarily expensive. Moreover, if this kind of extensive fieldwork effort is undertaken, it should be preceded by a preliminary analysis of the geographic distribution of variability in these relationships, to efficiently distribute field samples. However, no research on this problem has yet been carried out. Therefore, now is an auspicious time to begin addressing this problem, so that when VCL is launched, guidance on the most effective sampling schemes is available.

An efficient solution to this problem would take advantage of forest inventory data collected by various regional, national and international institutions. Using forest inventory data, both stand and canopy structure attributes can be estimated. Stand structure, in this paper, is defined as those indices typically used to describe the size and development of forest stands, attributes such as aboveground biomass, basal area, and leaf area index. Canopy structure refers to measurements of the dimensions of the forest canopy, such as cover, maximum height and mean height of dominant and co-dominant trees. Using forest inventory data, stand structural attributes can be directly calculated (e.g. basal area), or estimated using allometric equations (e.g. aboveground biomass). Canopy structure, and lidar measurements of canopy structure, can be simulated by models using inventory parameters such as tree height and crown depth, along with ancillary crown geometry information. We can then estimate the relationship between stand and canopy structure. The advantage of such an approach is that the relationship can be estimated using forest inventory datasets over broad geographic regions, and validated using a relatively small number of plots with coincident field and lidar measurements. The validated relationship can be applied to lidar measurements from anywhere within the boundary of the original forest inventory dataset. The large dataset of forest inventory plots can be used to estimate and incorporate the influence of topography, soils, and climate on the relationships between canopy and forest stand structure attributes.

Fusion of lidar and conventional optical remote sensing

The second major challenge in the use of lidar data for regional forest mapping applications is the lack of spatially comprehensive datasets. VCL will collect data in one waveform wide transects over 5% of the Earth's land surface, but will not collect spatially comprehensive data in any one place. LVIS is capable of collected spatially comprehensive data, but its swath width is limited to 1km and the data are not yet generally available. The advent of commercial lidar systems has greatly increased the availability of lidar data. Nevertheless, the high cost and low spatial extent of these data are still prohibitive for regional-scale mapping applications. While commercial lidar technology is sure to improve, certain practical limitations will remain; for instance, the high scan angles associated with truly wide swath sampling from airborne platforms. We suggest that the most practical solution for regional forest mapping applications will involve a sampling approach, with conventional optical remote sensing serving as a spatial stratification, and lidar providing a measurement of forest structure within the spatial units and/or a supplement for field data collection for interpretation of the spectral data. This approach can make the most efficient use of lidar data, by allowing the conventional optical remote sensing to define spatial boundaries, a task for which it is more suitable than lidar, and using the lidar to define forest structure, a task for which it is most suitable. While this approach can be implemented using data from current sensors, increased efficiency will result when sensors, and sensor platforms, are designed for this type of sampling.

2. ETM+LIFE
(Enhanced Thematic Mapper Plus Lidar for Forest Ecosystems)

We are addressing both problems discussed above in the ETM+LIFE project, for the region of western Oregon and Washington, USA. ETM+LIFE has two key objectives

Objective 1. Statistically relate lidar waveforms to ground-measured forest structural attributes. The purpose of this objective is to identify which common forest inventory and ecological variables (e.g., total cover, basal area, mean and variability in tree size, tree density, aboveground biomass, and LAI) can be derived from lidar, and to determine the geographic variability in these relationships.

Objective 2. Develop alternative strategies for characterizing forest structure and composition over large landscapes using combined Landsat, lidar, ground, and environmental data. Because current, and near-term future, lidar sensors will obtain samples of the global terrestrial biosphere rather than complete coverage, it is important to devise a means to spatially extrapolate from these samples to broader geographic regions.

Objective 1. Statistically relate lidar waveforms to ground-measured forest structural attributes

Rather than directly relate lidar measurements of canopy structure to coincident field measurements of forest structure, we are adopting a modeling approach, in which forest inventory data from throughout the western PNW region are used to estimate the relationship between lidar measurable height indices and stand structure attributes such as aboveground biomass. Field data taken coincident with lidar are used only

for verification of the models. Lidar remote sensing is an obvious choice for modeling from forest inventory data because lidar directly measures the vertical and horizontal variability in canopy height, which is closely related to two key inventory parameters, tree height and cover. In fact, forest inventory measurement of height and lidar measurements of height are often direct related. In the eastern deciduous forest (Lefsky, 1997) and in the Pacific Northwest (PNW) (Lefsky et al., 1999b), there is a 1:1 relationship between SLICER and field estimates of maximum height, and we would expect that this relationship will hold for all forest types, due to the simple physical nature of the measurement. However, maximum height is potentially representative of the height of only a single tree, and therefore is often not well correlated with stand structural attributes. In the PNW another important field estimate of canopy height, mean height of dominant and co-dominant stems, also has a 1:1 relation with a SLICER measurement, the average height of waveforms in a plot or stand. As a consequence, relationships between these two-field measured heights (maximum height and mean height of dominant and codominant stems) and forest stand structural attributes can be derived from field inventory data, and directly compared to relationships derived from lidar data. However, the robustness of this relationship is likely due to the fact that the SLICER footprint is 8-12m wide, roughly equal to the crown size of a dominant tree in the PNW (Cohen et al., 1990). When footprint size exceeds the mean crown size, the relationship between mean waveform height and mean height of dominant and co-dominant trees, we

expect, will weaken, and require statistical or physical modeling.

Preliminary work on the problem of estimating the relationship between stand height and aboveground biomass suggest that this approach is viable. A dataset consisting of 7700 field plots has been assembled from the U.S. Forest Service's FIA (Forest Inventory and Analysis) and CVS (Current Vegetation Survey) programs, and the Bureau of Land Management's forest inventory program. Using allometric equations from the H.J. Andrews permanent plot program, supplemented with equations from the BIOPAK library (Means et al., 1994), aboveground biomass has been estimated for each of the over 700,000 inventoried trees on these plots, and estimates of total aboveground biomass calculated for each plot. Using an imputation technique, heights for each tree in the dataset which did not have a measured height were estimated using a dataset of trees with measured heights, stratified by DBH, species, crown class, and environmental conditions. Two heights were calculated for each stand, maximum tree height (HMAX) and the mean height of dominant and co-dominant species (HU). In addition, we calculated the values of these variables raised to the second power (HU2 and HMAX2), their product (HUMAX), the fraction of basal area that was deciduous (DECID_IMP) and the product of DECID_IMP and HUMAX (DIHUMAX). Stepwise multiple regression was then used to related these height indices to the aboveground biomass of each plot. Aspects of the resulting regression are presented in Figure 1. The height
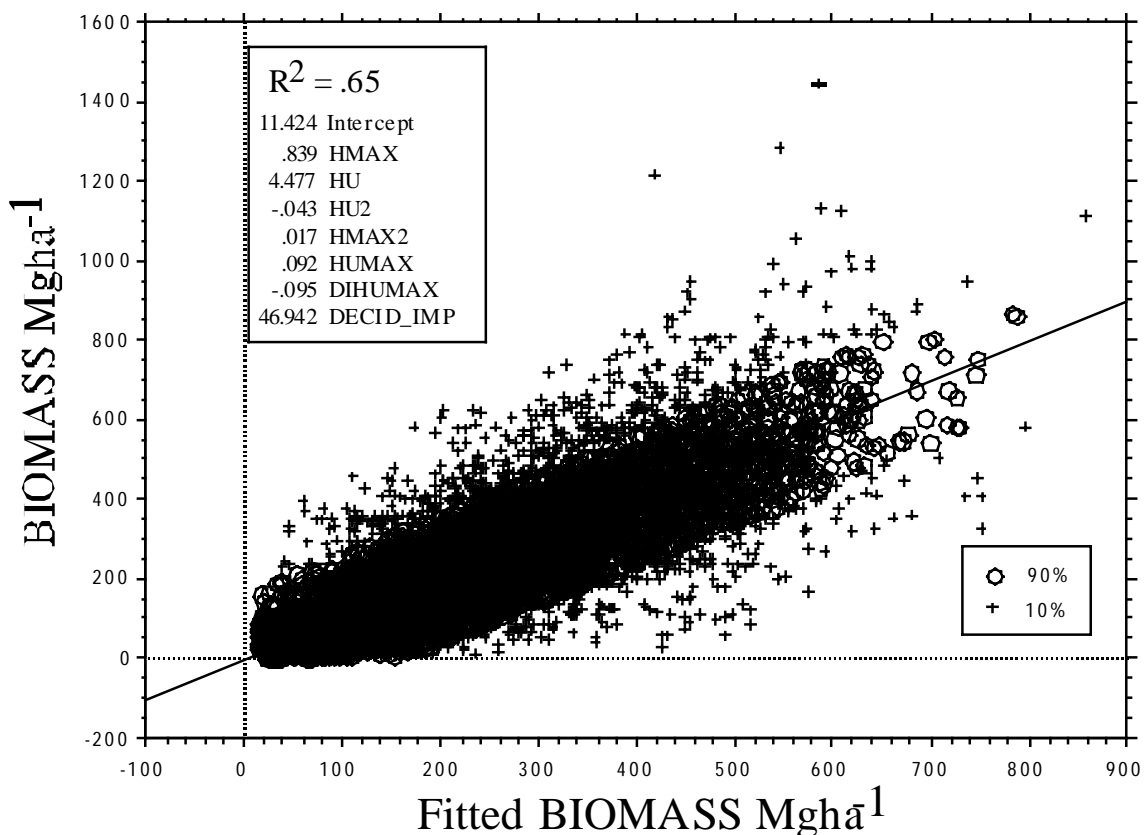
Fig. 1. Relationship between fitted aboveground biomass, estimated using height indices derived from forest inventory data, and observed aboveground biomass, for 7700 forest inventory plots from Western Oregon and Washington. Equation variables include HMAX (maximum height), HU (mean height of dominant and co-dominant trees), these variables raised to the second power (HU2 and HMAX2), their product (HUMAX), the fraction of basal area that is deciduous (DECID_IMP) and the product of DECID_IMP and HUMAX (DIHUMAX). Symbols indicate the 90% of data with the lowest absolute residual values, and the remaining 10%.

indices explained 65% of the variance in aboveground biomass, and all 7 indices were found to be statistically significant, although the HUMAX variable explained most of the variance.

A preliminary analysis of the effect of composition on the predicted biomass values has been performed, splitting the total dataset by the species which had the largest contribution to each plot's basal area. Species for which less than 100 plots were available were grouped in two "other" categories, one each for deciduous and coniferous species. This analysis indicates that a single height-biomass equation does not apply to all species. Intercepts and slopes for the resulting equations (relating the all-species predicted biomass to actual biomass) are as follows

Species	Intercept	Slope
PICO	24.877	0.726
PSME	13.312	0.95
TSHE	53.287	0.991
PIPO	-48.882	1.121
ALRU	-19.476	1.111
OTHD	-2.783	1.149
OTHC	-11.183	1.178
ABAM	-14.127	1.295
ABCO	-68.977	1.36
TSME	-26.681	1.482

Analysis of the relatively short-statured lodgepole pine (PICO) shows that the all-species height-biomass equation overestimates biomass, resulting in a regression slope of 0.726. The three most dominant species (Douglas-fir (PSME), western hemlock (TSHE), and ponderosa pine (PIPO), have, as would be expected, slopes near 1.0, and intercepts that are small relative to the range of values in aboveground biomass. Red alder (ALRU) and other deciduous species (OTHD) are next with slopes of 1.111 and 1.149, respectively. Species such as Pacific Silver Fir (ABAM), white fir (ABCO) and mountain hemlock (TSME), have the highest slopes, which appear to be associated with the high altitude conditions associated with these species. Most of the species making up the miscellaneous conifer species (OTHC) are also high altitude species, and they have a similar slope (1.178).

A preliminary analysis of the applicability of the forest inventory derived equations to lidar measured heights is available using data from the H.J. Andrews Experimental Forest (Lefsky et al., 1999b). Using 22 Douglas fir / western hemlock plots with coincident field and lidar measurements, we applied a simplified version of the equation presented in Figure 1 (omitting the deciduous composition variables, and using only plots dominated by either Douglas-fir or western hemlock) to lidar measured maximum height and mean height of waveforms. Figure 2 presents the relationship between aboveground biomass estimated using the forest inventory derived equation with lidar measured heights, and that estimated using data measured in coincident field inventory plots. The forest inventory equation explains 73% of variance, and the slope and intercept of a regression between observed and predicted biomass are 22.8 and 1.037, very close to an identity relationship. As field data from the 6 study areas (~100 plots) being measured as part of the ETM+LIFE project are processed, further validation of the inventory based predictions will be performed, for plots dominated by sitka spruce, red alder, ponderosa pine and high altitude sites dominated by Abies.

R^2=.728
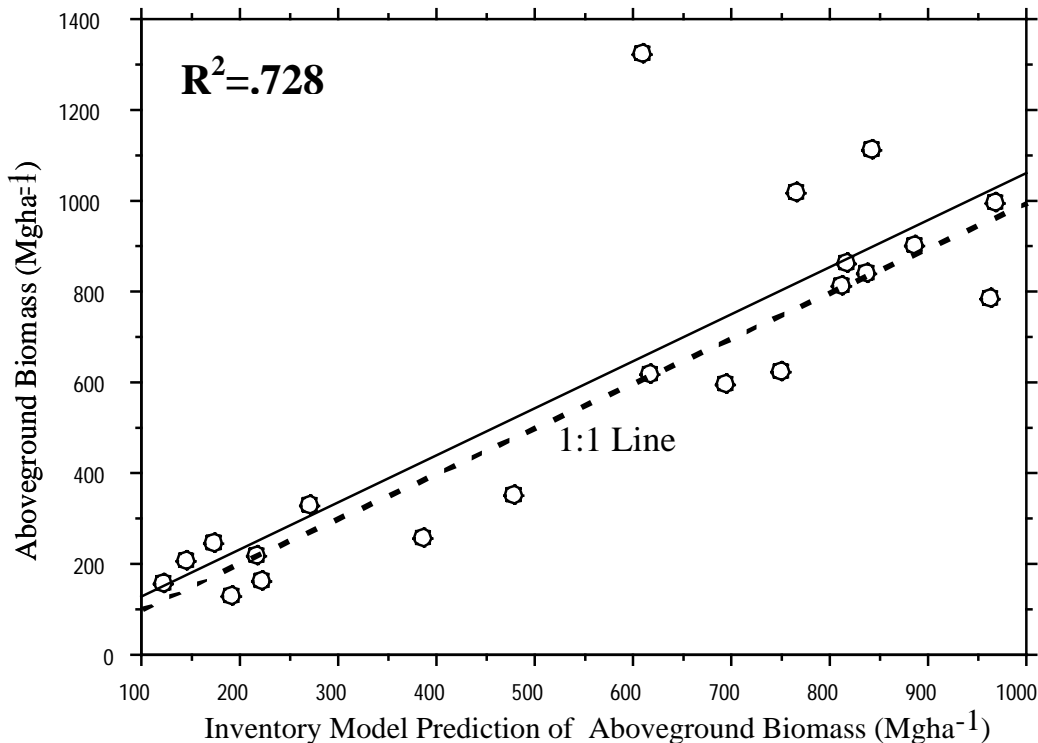
Fig.2. Comparison of forest inventory model of biomass, with lidar heights substituted for field measured heights, and observed biomass at the H. J. Andrews Experimental forest. Field Aboveground Biomass (Mg/ha)=22.878+1.037 * Predicted Biomass

In addition, we will be analyzing data for two other Douglas-fir / western hemlock study areas (Wind River Canopy Crane and Flynn Creek), at different points along a productivity gradient. Preliminary analysis from the forest inventory datasets suggest that there are 6 broad productivity zones in the PNW, and that the use of a single equation for predicting aboveground biomass in Douglas-fir / western hemlock communities could result in as much as a 30% bias for the most extreme of these zones. This result will be tested using data from the three Douglas-fir / western hemlock communities.

The results presented here are dependent on the 1:1 relationship between SLICER measurements and field measurements of height. In cases where there is a strong correlation between SLICER and field measurements of height that do not fall on a 1:1 line, simple regression based corrections could used. However, to most robustly relate tree-based height measurements from a variety of forests to lidar data from sensors with a variety of sampling methods, two steps are required. First, the geometry of individual tree crowns (estimated using forest inventory data) must be built into a three-dimensional model of canopy structure. These types of individual tree models of the canopy have been successfully used by Nelson (1997), Li and Strahler (1985), Sun and Ranson (1995), and Van Pelt and North (1996) for various purposes. Second, a model that simulates the interaction of a variety of lidar remote sensing devices with the modeled canopy structure is needed. When these two models are completed, we will be able to undertake analysis using more complex indicators of height and cover, through direct simulation of lidar waveforms.

In summary, we have preliminary evidence suggesting that it is possible to use forest inventory data to estimate the relationship between lidar measurements of canopy structure and forest stand structure attributes such as aboveground biomass. Overall, tree based stand height indices explain 65% of variance in aboveground biomass. This is a positive finding, in that although this result was expected, it has never been demonstrated over such a large area. Nevertheless, the amount of variability in this relationship is much higher than previously observed in studies that relied on smaller numbers of field plots. Further research is needed to characterize and reduce this variability. Species composition appears to be a significant effect on the relationship between stand heights and aboveground biomass, as does productivity.

Objective 2. Develop alternative strategies for characterizing forest structure and composition over regional landscapes using combined Landsat, lidar, ground, and environmental data.

We are evaluating three different strategies for integrating lidar and ETM+ data. Within each region, we are testing two-phase (double) sampling for stratification to estimate means and standard errors of structural attributes (e.g., biomass, LAI). Double sampling will allow us to use the less precise but spatially comprehensive estimates from ETM+ data (primary phase) as weighting factors for the more precise but spatially limited lidar-based estimates (secondary phase). Since the VCL

sensor will provide a systematic sample of the whole terrestrial biosphere, double sampling is a particularly important statistical procedure to be tested in this context.

Secondly, we are evaluating the effectiveness of lidar data in replacing ground data for parameterizing statistical and geo-statistical models (based on ETM+ imagery) of several individual forest structure attributes. Collecting ground data to support extensive remote sensing characterizations of vegetation is essential, but it is also expensive. Thus, we are interested in determining how effectively lidar data can be used in lieu of ground data for characterizations of forest vegetation using ETM+ data. Several independent, continuous data layers will be developed, one for each forest attribute of interest. Multi-temporal ETM+ imagery is being acquired to take advantage of phenological differences in vegetation.

Finally, we plan to use lidar data to parameterize ETM+ multivariate statistical models that will preserve the multiple-attribute covariance structure of related ground data. We are in the processing of evaluating a new multivariate method to predict forest structure from ETM+ data in combination with lidar data: the most similar neighbor (MSN) method of Moeur and Stage (1995). This multivariate method has the advantage of predicting several forest attributes simultaneously, and preserving the covariance structure among them. As such, the multivariate predictions based on ETM+ data are expected to be more ecologically realistic and retain the range of variability present in the higher quality lidar data, compared to several independent univariate models.

Currently, our work is focussed on the second task: determining the ability of lidar data to replace field data for parameterizing statistical and geo-statistical models (based on ETM+ imagery) of individual forest structure attributes. There are several motivations for this analysis.

Lidar data from the VCL mission will have high reliability relative to field data, which is prone to human error in the field, and in subsequent processing. The data will have excellent geolocation- which forest inventory datasets often do not. The global VCL dataset will be uniform and will cross political and administrative boundaries. Finally, the data from VCL will measure the physical canopy, which should be more directly related to the spectral indices and spatial pattern of ETM+ than forest structure attributes such as aboveground biomass. The use of lidar data may in fact give additional motivation to work on mechanistic modeling to relate forest structural attributes to the spectral and spatial qualities of imagery such as ETM+, by providing widespread estimates of the vertical structure of canopies which is related to both sets of attributes.

Statistical Fusion of Lidar and TM

Preliminary work on statistical analysis relating lidar and TM is being performed while we wait for a sufficient dataset of ETM+ data to be collected. In the most basic analysis we used a set of five Landsat TM images from March, May, June, July, and

August 1992, for a 20 x 20 km area around the Wind River Canopy Crane, in southern Washington state. Images were transformed to the tassled-cap Brightness, Greeness, and Wetness indices using standard transformation factors (Crist and Cicone, 1984), and combined to create a single, 15-band image (5 dates x 3 bands). The SMAP algorithm (Bouman and Shapiro, 1994) was used to segment the June image into patches that were relatively uniform in their spectral qualities. Six transects of SLICER data were collected at the site in a rosette pattern over the canopy crane, during September, 1995. Four transects were selected for model generation, the other two were designated for validation purposes only. Lidar estimates of mean stand height were calculated for each patch that had coincident lidar data. For these same patches, average spectral values were calculated for each of the 15 image bands. Stepwise multiple regression was used to predict mean stand height from the average spectral values of each patch. The resulting regression explained 71% of variance, and 9 of the 15 bands were found to be statistically significant.

Examination of a plot of observed versus predicted values (Figure 3) indicates that the model (generation) points and the validation points have very similar distributions, although the correlation coefficient associated with the validation dataset is smaller (58% vs 71%). Examination of the residuals from the validation dataset indicates that approximately 60% of predictions fall within 5 m of the observed stand height. The resulting equations can either be applied on a pixel-by-pixel basis (e.g., Cohen In Review) or can be applied on a patch-by-patch basis, which preserves their spatial structure. Finally, provided adequate models between image spectral qualities and height indices are found, equations relating those indices to forest stand structure attributes can be applied to the images, resulting in spatially comprehensive maps of, for instance, aboveground biomass. An example of one such image is presented in Figure 4.

3. CONCLUSION

Our preliminary results suggest that it is feasible to use forest inventory datasets for estimating the relationships between lidar remotely sensed height indices and elements of forest structure such as aboveground biomass. This type of application is likely to prove valuable for estimating such relationships and their regional variability, and to supplement datasets of coincident field measure stand structure and lidar measurements. Integration of lidar and ETM+ data is likely to be one important application for VCL data when data from that system become available.
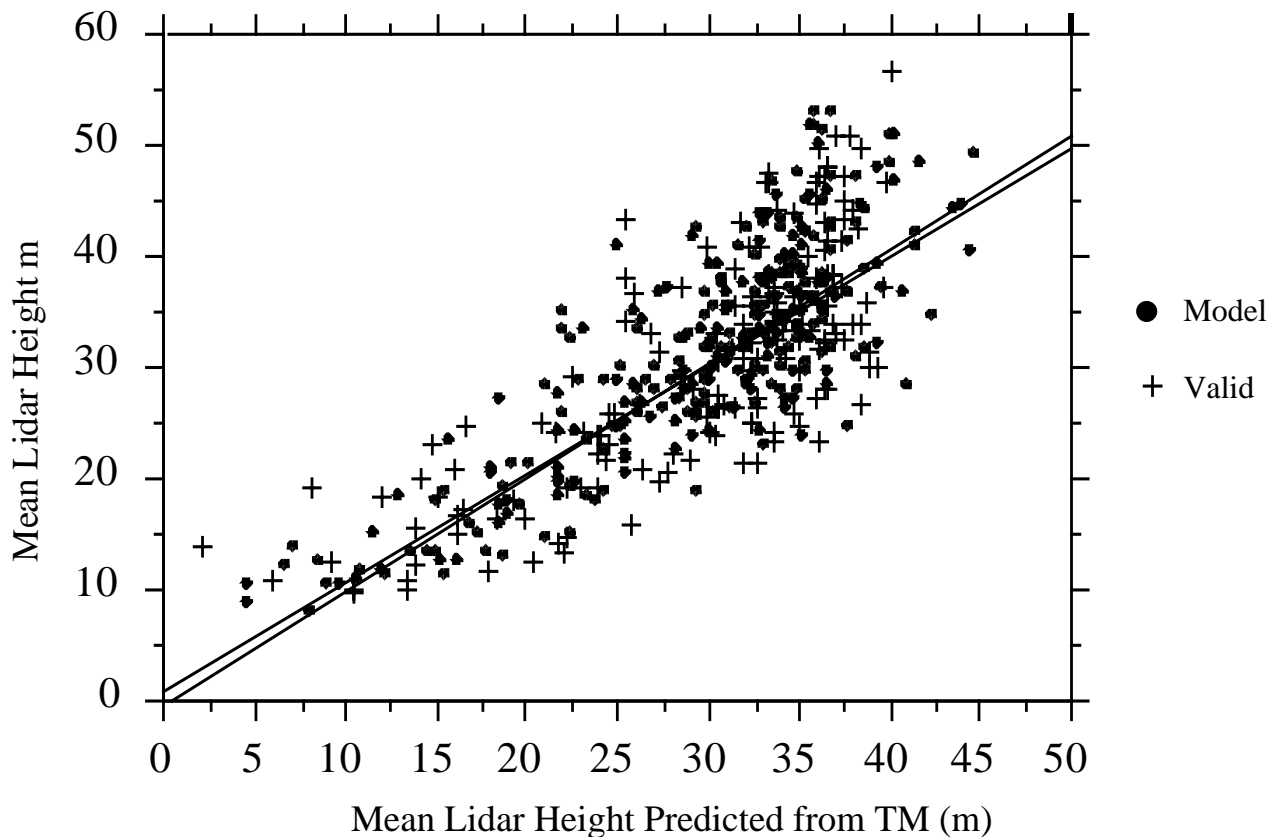
Fig. 3. Mean height predicted from TM spectral data, versus observed mean height. Mean height = .827 + .975 * Predicted Mean Height; R^2=.581 (Valid); Mean Height = -.518 + 1.025 * Predicted Mean Height; R^2 = .71 (Model)
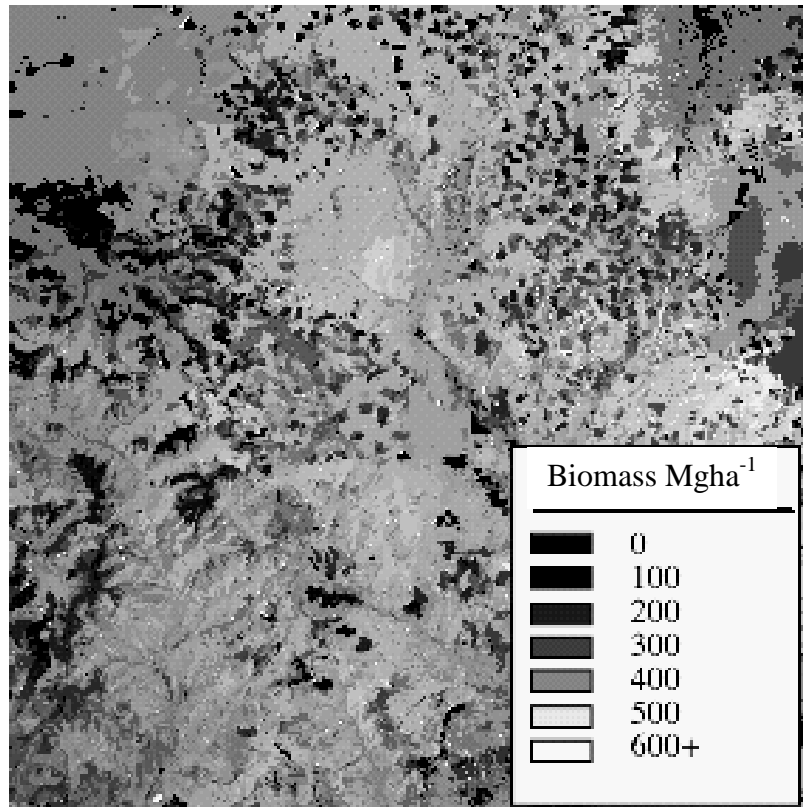
Fig. 4. Predicted biomass

4. REFERENCES

Bouman, C. and M. Shapiro, 1994. A multiscale model for bayesian image segmentation. *IEEE Transactions on Image Processing*, **3**(2):162-177.

Cohen, W. B., T. A. Spies, and G. A. Bradshaw, 1990. Semivariograms of digital imagery for analysis of conifer canopy structure. *Rem. Sensing of Environ.*, 34: 167-178.

Crist, E. P., and R. C. Cicone, 1984. A physically-based transformation of thematic mapper data--The TM tasseled cap. *IEEE Transactions on Geoscience and Remote Sensing,* GE-22: 256-263.

Dubayah, R., J. B. Blair, J. L. Bufton, D. B. Clark, J. JaJa, R. Knox, S. B. Luthcke, S. Prince, and J. Weishampel, 1997. The Vegetation Canopy Lidar mission, proceedings, land satellite information in the next decade II: sources and applications. *Am. Soc. for Photogramm. and Remote Sens.*, Bethesda, MD, 100-112.

Harding, D.J., J. B. Blair, E. Rodriguez, and T. Michel, 1995. Airborne laser altimetry and interferometric SAR measurements of canopy structure and sub-canopy topography in the Pacific Northwest. *Proc. Second Topical Symposium on Combined Optical - Microwave Earth and Atmosphere Sensing (CO-MEAS'95)*, 22-24.

Lefsky, M.A., 1997. Application of lidar remote sensing to the estimation of forest canopy and stand structure. Ph.D. thesis, Environmental Sciences, University of Virginia, Charlottesville, VA.

Lefsky, M.A., D. J. Harding, W. B. Cohen, G. G. Parker, and H. H. Shugart, 1999a. Lidar remote sensing of forest basal area and biomass: application and theory. *Rem. Sensing of Environ.,* 67:83-98.

Lefsky, M., W. Cohen, S. Acker, G. Parker, T. Spies, and D. Harding, 1999b. Lidar remote sensing of biophysical properties and canopy structure of forests of Douglas-fir and western hemlock. *Rem. Sensing of Environ.,* **70**(3):339-361.

Li, X. S., Strahler, A. H., 1985. Geometric-optical bidirectional reflectance modeling of a conifer forest canopy. *IEEE Transactions on Geoscience and Remote Sensing,* GE-24: 906-919.

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

Moeur, M. and A. R. Stage, 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. *Forest Science,* 41:337-359.

Nelson , R., 1997. Modeling forest canopy heights: the effects of canopy shape. *Rem. Sensing of Environ.*, 60:327-334.

Sun, G. and K.J. Ranson, 1995. A three-dimensional radar backscatter model of forest canopies. *IEEE Transactions on Geosciences and Remote Sensing,* GE-**33**(2):372-382.

Van Pelt, R. and M.P. North, 1996. Analyzing canopy structure in Pacific Northwest Old-growth forests with a stand-scale crown model. *Northwest Science,* 70:Special Issue 15-30.