LIDAR INVENTORY AND MONITORING OF A COMPLEX FOREST

Jacob L. Strunk
Precision Forestry Cooperative
College of Forest Resources
University of Washington,
Seattle, WA 98195-2100
strunj@u.washington.edu

Stephen E. Reutebuch
Robert J. McGaughey
USFS, Pacific Northwest Research Station,
Seattle, WA 98195-2100
sreutebuch@fs.fed.us
bmcgaughey@fs.fed.us

Jeffrey R. Foster
Forestry Branch, Public Works
Fort Lewis Military Reservation
Fort Lewis, WA 98433
jeffrey.r.foster@us.army.mil

ABSTRACT

The Fort Lewis Military Base in west central Washington State covers 35,025 hectares, 23,404 of which have significant tree cover (forests, woodlands, savannas). The undeveloped portion of the base is highly diverse ecologically relative to other western Washington forests. It is west of the Cascades Mountains yet its ecology is a blend of that found on both the east and west sides of the Cascades. Forest management balances the protection of the unique ecological niches found here with military training, fire management, timber production and human uses. A sparse grid of permanent forest inventory plots is used to monitor forest growth and conditions. Augmentation of the permanent forest inventory with aerial LIDAR measurements is currently under examination by the Fort Lewis Forestry Branch. The basic LIDAR-based inventory technique is to model inventory parameters with plot based LIDAR metrics. Application of models to the landscape results in a spatially explicit complete inventory of the forest. Preliminary inventory analysis in which only basal area was considered resulted in a linear model with an R^2 of 0.86. The implication of this result is that the success found in modeling this variable at Fort Lewis may translate to similar success for other complex multi-use public forests where knowledge of landscape patterns is needed for planning and monitoring.

INTRODUCTION

Fort Lewis Military base is a complex mixture of forest types managed for an assortment of military (primary), civilian and ecological objectives. Monitoring of forest conditions to inform management decisions is performed with a relatively sparse, grid-based system of permanent fixed-area plots. Recent acquisition of high density discrete return aerial LIDAR for the entire installation provides complete, fine scale, spatially explicit information of forest canopy structure over Fort Lewis. In this study we evaluate the use of airborne LIDAR-based inventory techniques for Fort Lewis's forest landscape.

STUDY SITE

Fort Lewis Military Installation is located at the southeast edge of Puget Sound in western Washington State. The 35,025 hectare installation has 23,404 hectares of land with significant tree presence. The forests are dominated by coniferous forests, predominately dry Douglas-fir but include some moist forest types (Douglas-fir, redcedar, western hemlock, red alder, bigleaf maple, cottonwood, and a variety of additional minor deciduous and coniferous species). Oregon white oak and ponderosa pine woodlands and grassland ecosystems are interspersed with the coniferous forests, forming a complex forest-prairie mosaic.

METHODS AND MATERIALS

Continuous Forest Inventory Plots

The continuous forest inventory (CFI) plots on Fort Lewis are located on a grid of regularly visited fixed-area (0.081 hectare) circular plots. There are 131 plots located in the forested area of Fort Lewis. For this study, forested area was mapped from the LIDAR data assuming areas with vegetation lower than 5 m are non-forested (Fig. 1). Each plot is associated with a plot record that describes the individual trees on the plot. Each time a plot is revisited all trees on the plot are re-measured. Diameter is measured for every tree and height is recorded for the site tree. All of the trees are examined for defects. The CFI plots were last re-measured in 2005.

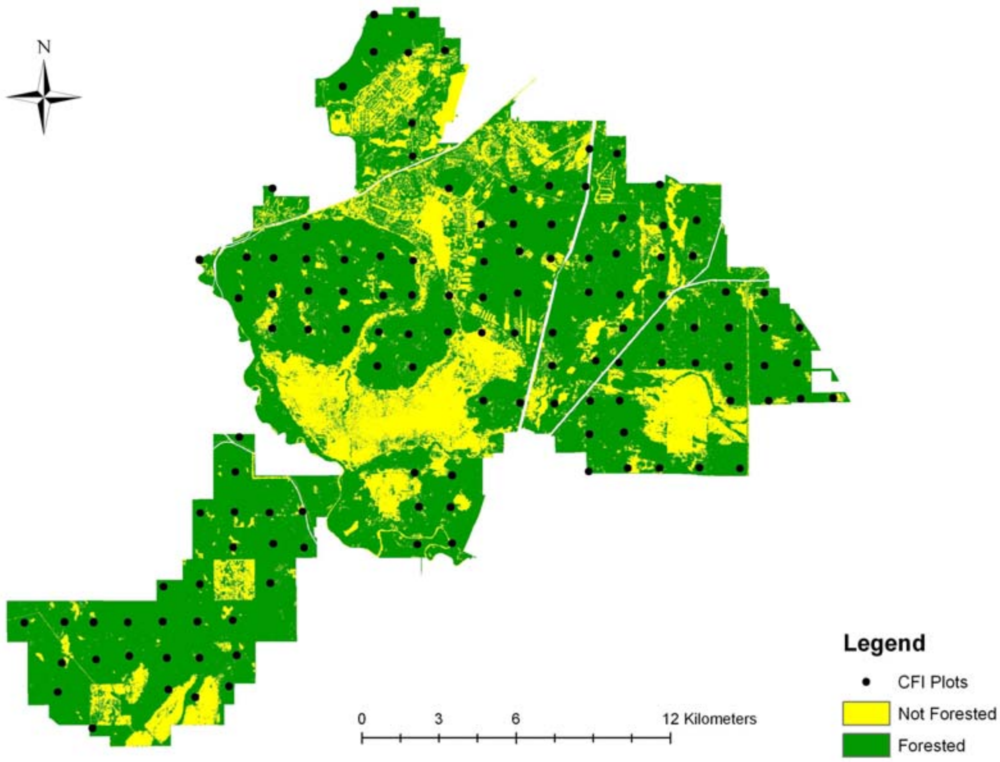

Figure 1. Fort Lewis forested and non-forested areas and CFI ground plot locations.

Plot Georeferencing

To properly align the ground sample plots with the LIDAR data, plot locations must be determined to within approximately +/- 1 m. This is complicated by the fact that precise plot positions under canopy are difficult to obtain with GPS (Clarkin, 2007). Approximate coordinates for the CFI plots are available but the range of error can exceed the diameter of the plot. In order to obtain the desired accuracy for plot coordinates georeferencing was performed using survey grade GPS units with differential post processing (Ensemble, 2002).

A Javad Navigation Systems MAXOR GGDT model receiver with GPS/GLONASS L1/L2 signal capabilities was used to determine plot coordinates. Receiver accuracy was assessed under canopy using these same receivers at Capitol State Forest within 50 km of Fort Lewis. Results at Capitol State Forest indicate that this receiver achieves positions with an RMSE of 0.86 m under coniferous canopy (Clarkin, 2007). The range of canopy densities at Fort Lewis is similar to that tested at Capitol State Forest. The positional accuracies measured at Capitol State Forest provide a reasonable indication of the range of accuracies that may be expected for GPS occupations at Fort Lewis.

LIDAR Data

In this study, a discrete-return, small-footprint airborne LIDAR dataset was acquired that consists of a three-dimensional "cloud" of 2.9 billion data points that describe the shape and structure of the terrain, infrastructure, and vegetation on Fort Lewis. LIDAR coverage was flown using the mission specifications shown in Table 1. The average density of LIDAR points for the base is 6.19 per square meter (Fig. 2).

Table 1. Fort Lewis LIDAR mission specifications.

Flight dates	September 19-21, 2005; leaf-on
LIDAR scanner	Optech ALTM 3100*
Scan pulse rate	71 kHz
Maximum returns per pulse	4
Scan angle	+/-14 degrees
Beam divergence	0.3 mrad
Flight above ground level	1,100 m
Flight line configuration	Opposing parallel lines
Flight line overlap	50% sidelap
Average LIDAR point spacing	<40 cm

* The use of commercial names is for the convenience of the reader and does not imply any endorsement by the USDA Forest Service.

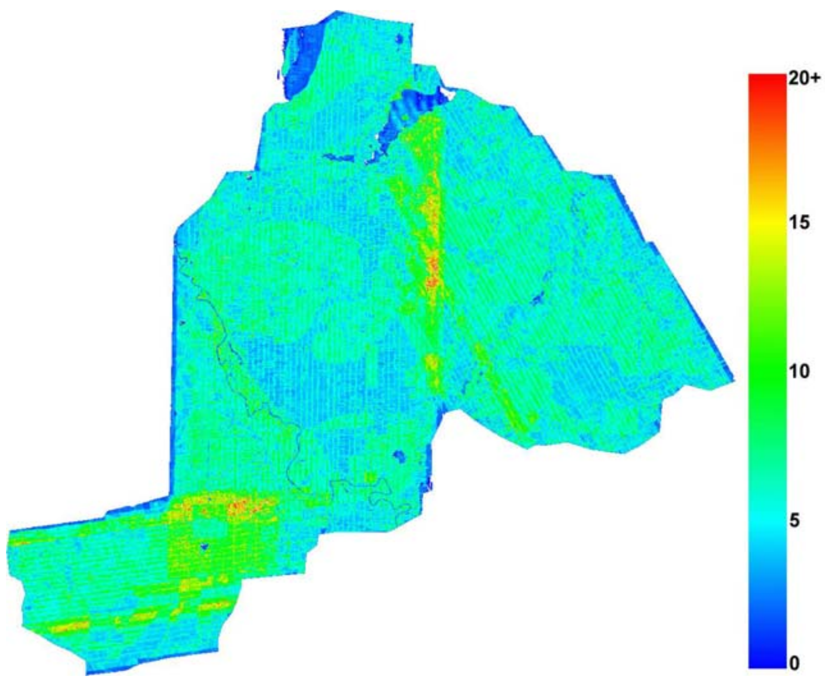

Figure 2. Color-coded map of LIDAR point density per square meter over Fort Lewis.

LIDAR Metrics

LIDAR metrics are descriptive structure statistics calculated from the raw XYZ LIDAR point cloud. Metrics were calculated with the program FUSION (McGaughey and Carson, 2003). The first step in generating LIDAR metrics is to select only the raw data corresponding to the plots of interest using the FUSION module ClipData. ClipData requires spatial parameters describing the plot location and size and then outputs the subset of LIDAR points that was collected within the plot area. CloudMetrics is then used to calculate metrics for the plot at a user defined resolution. A third utility, GridMetrics, calculates metrics for the entire landscape (Table 2). Fig. 3 illustrates how the canopy transparency metric is computed.

In addition to calculating canopy metrics, a raster of maximum heights for 15-m pixels was used to indicate forested areas. Analysis in ArcMap (ESRI, 2007) classified the raster into forested and non-forested pixels based upon the criteria that a forested cell must have a maximum height of greater than 5 meters. The result of this procedure was used to tabulate the total forested area and map forested and non-forested areas across Fort Lewis (Fig. 1).

Table 2. LIDAR metrics computed by FUSION utilities (McGaughey, 2007).

LIDAR Metric	Description
Total number of returns	Total number of discrete LIDAR measurements for a plot.
Minimum	Minimum height of all LIDAR data for a plot.
Maximum	Maximum height above ground of all LIDAR data for a plot.
Mean	Mean height above ground of all LIDAR data for a plot.
Median	Median height above ground of all LIDAR data for a plot.
Standard deviation	Standard Deviation of heights of LIDAR data for a plot.
Variance	Variance of heights of LIDAR data for a plot.
Interquartile distance	Interquartile distance for heights of LIDAR data for a plot.
Skewness	Skewness of distribution of LIDAR heights for a plot.
Kurtosis	Kurtosis of distribution of LIDAR heights for a plot.
AAD (average absolute deviation)	Average absolute deviation of LIDAR heights from the mean of LIDAR height for a plot.
Percentile height values (5th, 10th , 20th, 25th,…, 95th percentiles)	Height above ground at which a specified percentage of returns fall below.
Mean intensity	Mean intensity of those returns above a user specified height above ground.
Maximum intensity	Max. intensity of those returns above a user specified height above ground.
Canopy Cover	Percentage (0 - 100) of first returns 3 meters or more above ground.
Canopy Transparency	Percentage (0 - 100) of first returns above 6 meters after removal of points below 3 meters.

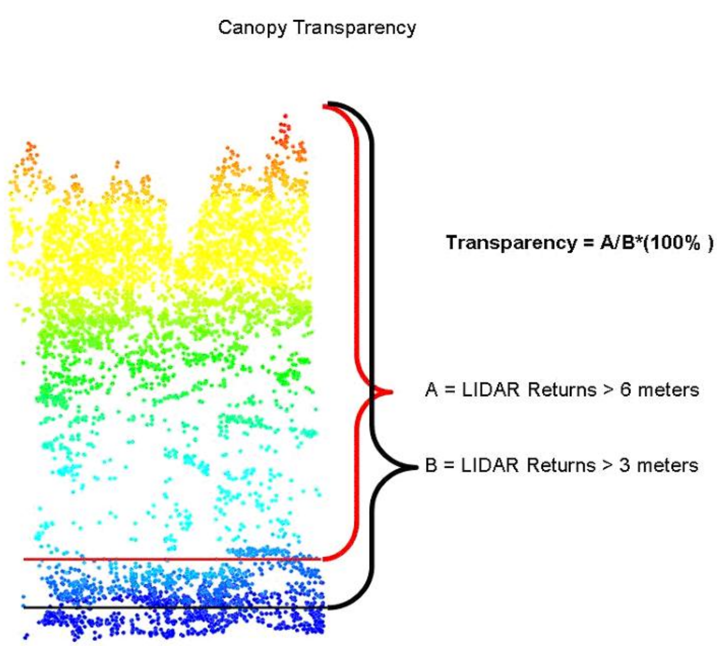

Figure 3. Example of FUSION canopy transparency metric computation.

REGRESSION ANALYSIS

Analysis of data was performed using the open source public domain statistical software R (R, 2006). The platform is a robust statistical programming environment capable of complex data manipulations and customizable with a wide variety of modules. Additional, non-standard R modules, Leaps (Lumley, 2006) and Ipred (Peters, 2007) were also used in this analysis.

The aerial LIDAR data are not sufficient for direct measurement of stem diameter. Instead stem diameter is predicted using regression analysis (Naesset, 2001) with LIDAR metrics that describe canopy size and vertical distribution. The resulting prediction model quantifies the allometric relationship between canopy parameters described by LIDAR metrics and stem basal area.

Response and Predictor Variables

The predictor variables for the regression analysis are the metrics derived from raw LIDAR point clouds for each plot. There are 44 LIDAR-based predictors that fall into three groups. The first group of metrics consists of point height above ground metrics. These describe the vertical distribution of raw LIDAR points relative to a given ground surface (usually a LIDAR-derived digital terrain model). The second group is derived from LIDAR intensity values. The last group of predictive metrics consists of measures of the relative density of points above a given height.

The response variable in this study is total basal area for a plot. This value, calculated per plot, is the sum of stems' cross-sectional areas measured at 1.37 meters height above the ground for all trees larger than 11.5 cm in diameter.

Automated Model Selection

There are many LIDAR-derived predictors from which to choose a model. Choice of an acceptable model was facilitated with a subset regression routine (Lumley, 2006). The optimized routine indicates potentially acceptable models ranked by the number of variables in the model and the specific model's coefficient of determination (R^2) as shown in Table 3.

Table 3. Comparison of the fit statistics and independent variables for three basal area models developed using the subset regression method.

Models	1	2	3
Independent variables	25^{th} Percentile Ht.**	75^{th} Percentile Ht.***	75^{th} Percentile Ht.***
	75^{th} Percentile Ht.***	Max Intensity^{+}	Canopy Cover***
	Mean Intensity**	Canopy Cover***	Canopy Transparency***
	Canopy Cover***	Canopy Transparency***	
	Canopy Transparency***		
Dependent variable	ln (basal area, m^2)	ln(basal area, m^2)	ln (basal area, m^2)
Model p-value	< 0.001	< 0.001	< 0.001
R^2 *ln(Basal Area)*	0.87	0.86	0.86
Stand. error *ln(Basal Area)*	0.29	0.30	0.30
Significance codes: "***" = 0.001; "**" = 0.01; "*" = 0.05; "+" = 0.1			

Cross Validation

The subset selection process is a biased procedure (Miller, 2002). The best subset regression model does not distinguish true correlation from bias resulting from random errors. This means that a best model in terms of residuals may not be the best predictive model.

The true error for the model is estimated using k-fold cross validation (Mevik and Cederkvist, 2004). K-fold cross validation is an iterative process in which the data are partitioned into k segments and the model is fitted using k-1 of the segments. In this study, we used five segments (20 percent of data withheld each iteration). The residuals for iteratively excluded segments are used to generate the root-mean-square-error (RMSE) for the model (Table 4).

Table 4. Comparison of residual-based RMSE using k-fold cross validation for basal area models.

Model	1	2	3
k-fold RMSE (m^2)	0.31	0.32	0.31
RMSE using entire dataset (m^2)	0.28	0.29	0.30

Selected Model

The best linear model is a description of the relationship between basal area and the selected metrics. There is not a consensus on the right way to select the best model (Miller, 2002). Each method has its own pitfalls. The method of best subset selection employed here is biased; however, the k-fold RMSEs are essentially equal to the model RMSE, indicating stability in the three models tested. Model 3 is the simplest of the models and the tradeoff for predictive ability is negligible. In this analysis emphasis was placed upon simplicity, hence model 3 is deemed the most acceptable model for our purposes (Eq. 1).

Eq. 1: $$\ln(BA) = -3.29 + .036X_1 + .018X_2 + .017X_3$$

Where, ln(BA) = predicted natural logarithm of basal area (m^2)
X_1 = 75th percentile height (m)
X_2 = canopy cover (percent)
X_3 = canopy transparency (percent)

Mapping Basal Area with Regression Model

The selected linear model (Eq. 1) is in units of ln(m^2). These units are not easily interpreted so it is necessary to back transform the model to basal area units of m^2. This process, however, introduces a systematic bias (Sprugel, 1983). A simple data dependent correction factor accounts for this bias in the back-transformation process. A second type of bias was introduced because the linear regression was performed on 809 m^2 plots but was applied to 900 m^2 area square pixels. Eq. 2 includes a composite correction factor for the two biases. Eq. 2 was used to predict basal area over the entire forested portion of Fort Lewis (Fig. 4).

Eq. 2: $$BA = CF * e^{(-3.29+.036X_1+.018X_2+.017X_3)}$$

where, BA = basal area (m^2)
CF = composite bias factor, in this case: $1.04_{(\text{log bias})} * 1.10_{(\text{pixel bias})} = 1.14$

Total Basal Area Estimates. The LIDAR estimate of total basal area for the entire Fort Lewis forested area was calculated by summing all pixel values in the predicted basal area raster (Table 5). A simple random sample (SRS) total basal area estimate based on the ground plots was calculated by multiplying the average plot's basal area (2.82 m^2 per plot) by the LIDAR-determined total forest area (26,224 hectares). The population size or total forested area required for the SRS ground estimate was calculated using LIDAR-derived canopy heights; therefore, it is likely more accurate than would be found without use of the LIDAR data.

Table 5. Predicted total basal area: SRS ground plot estimate versus LIDAR basal area model estimate mapped over total LIDAR-derived forested area on Fort Lewis.

			$CI_{95\%}$	
Method	SE (m^2)	Total Basal Area (m^2)	From	TO
SRS	49,355	913,086	816,351	1,009,821
LIDAR	26,256	663,471	612,007	714,934

The predictions displayed in Table 5 were performed on the entire base including housing areas or other regions with low density of trees. This may be responsible for the elevated SRS estimate of total basal area relative to the LIDAR estimate because the CFI plots were installed only in areas with significant presence of trees from a forest management perspective.

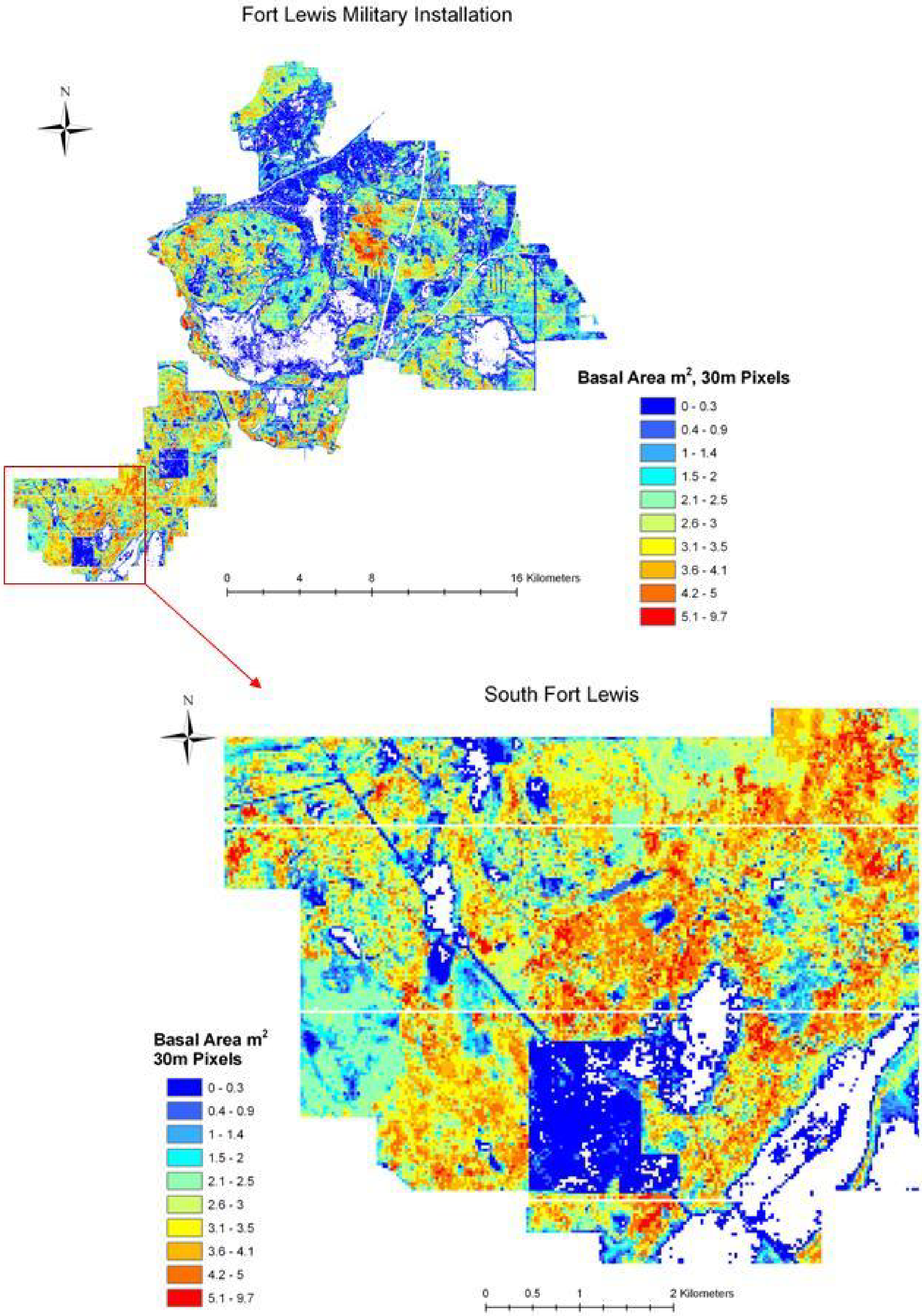

Figure 4. Basal area (per 30m pixel) for entire base (top); southwest section (bottom). White areas are non-forested. Note: blue rectangular areas are private non-forested ownerships within the base perimeter.

CONCLUSIONS

Results from this study indicate that LIDAR-based prediction of stand parameters is feasible even for complex forests. This is notable because most related studies in this field were performed on homogeneous managed forests. The LIDAR estimate of total basal area for Fort Lewis is significantly better than the SRS estimate and the standard error for the SRS is likely even a gross under-estimate of the true standard error. This is indicated by the fact that the SRS estimate does not include the likely better LIDAR estimate of total basal area. Basal area was the only parameter considered here, but future work will involve estimating parameters such as stem volume, biomass, quadratic mean diameter and crown base height. This work has implications in that managers of complex forest systems can measure and monitor forest change at very high resolution for the entire landscape.

ACKNOWLEDGEMENTS

Data and funding for much of this research was provide by Fort Lewis's Forestry Branch Environmental Division of Public Works. Additional funding and consultation were provided by the PNW Research Station's Silviculture and Forest Models team. Special thanks to Hans-Erik Andersen of the Anchorage Pacific Northwest Forest Inventory and Analysis research team for continued technical feedback. Last but not least, thanks to the Precision Forestry Cooperative and Corkery family for funding Jacob Strunk's graduate research appointment.

LITERATURE CITED

Clarkin, Tobey, 2007. Modeling global navigation satellite system positional error under forest canopy based on LIDAR-derived canopy densities. MS thesis, University of Washington, Seattle, WA.

Ensemble, 2002. Version 1.0, http://storage.javad.com/downloads/ensemble/PinSetup.exe (lasted accessed February 16, 2008).

ESRI ArcMap, 2007. Version 9.2, http://www.esri.com (lasted accessed February 16, 2008).

Lumley, Thomas, 2006. Leaps, version 2.7. http://rweb.stat.umn.edu/R/library/leaps/html/leaps.html

McGaughey, R. J. and Carson, W. W., 2003. Fusing LIDAR data, photographs, and other data using 2D and 3D visualization techniques, In: *Proceedings of Terrain Data: Applications and Visualization – Making the Connection*, October 28-30, 2003; Charleston, South Carolina: Bethesda, MD: American Society for Photogrammetry and Remote Sensing. pp. 16-24.

McGaughey, Robert J., 2007. USDA Forest Service, FUSION manual, version 2.60. http://www.fs.fed.us/eng/rsac/fusion/ (lasted accessed February 16, 2008).

Miller, Alan, 2002. *Subset selection in regression* (2nd ed.), Boca Raton, FL. Chapman & Hall/CRC.

Mevik, B.H., Cederkvist, H.R., 2004. Mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR), *Journal of Chemometrics,* 18(9):422-429.

Næsset, Erik, 2001. Predicting forest stand characteristics with airborne scanning laser
using a practical two-stage procedure and field data, *Remote Sensing of Environment*, 80: 88– 99.

Peters, Andrea, Hothorn, Torsten, 2007. Ipred: Improved Predictors, version 0.8-5, http://cran.r-project.org/web/packages/ipred (lasted accessed February 16, 2008).

R Development Core Team, 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Sprugel, D.G., 1983. Correcting for bias in log-transformed allometric equations, *Ecology,* 64(1):209-210.