

Predicting Live and Dead Tree Basal Area in Bark Beetle-Affected Forests from Discrete-Return LiDAR

Andrew T. Hudak¹, Ben Bright², Jose Negron³, Robert McGaughey⁴, Hans-Erik Andersen⁵, Jeffrey A. Hicke⁶

¹U.S. Forest Service, Rocky Mountain Research Station, ahudak@fs.fed.us

²University of Idaho, Department of Geography, bright@uidaho.edu

³U.S. Forest Service, Rocky Mountain Research Station, jnegrone@fs.fed.us

⁴U.S. Forest Service, Pacific Northwest Research Station, bmcgaughey@fs.fed.us

⁵U.S. Forest Service, Pacific Northwest Research Station, handersen@fs.fed.us

⁶University of Idaho, Department of Geography, jhicke@uidaho.edu

Paper Number: ####SL2012-139

1. Introduction

Recent bark beetle outbreaks in western North America have been widespread and severe. High tree mortality due to bark beetles affects the fundamental ecosystem processes of primary production and decomposition that largely determine carbon balance (Kurz *et al.* 2008, Pfeifer *et al.* 2011, Hicke *et al.* 2012). Forest managers need accurate data on beetle-induced tree mortality to make better decisions on how best to remediate beetle-affected forests and restore healthy ecosystem services (Negron *et al.* 2008). Discrete-return LiDAR measures canopy height and density, and LiDAR intensity provides some indication of the spectral reflectance and condition of canopy elements (foliage and branches) (Kim *et al.* 2009). LiDAR has been successfully applied to estimate biomass and carbon stocks in healthy forest (Hudak *et al.* 2012) and beetle-affected forest (Bright *et al.* 2012). A challenge in beetle-affected forests is that most airborne LiDAR has a single near infrared wavelength; i.e., LiDAR lacks the multispectral information useful for distinguishing between green, red, and grey trees. However, LiDAR intensity values may help distinguish between live green and dead red or grey trees. Moreover, mountain pine beetles (the most widespread bark beetle currently) and spruce beetles preferentially attack larger trees, so beetles impart a canopy structural signature that may be exploited (Coops *et al.* 2009).

Our objective is to predict Live and Dead Basal Area (BA) in beetle-affected areas in five states in the USA using canopy height, density, intensity, and topographic metrics derived from discrete-return airborne LiDAR data, tree measurements collected in field plots and summarized into plot-level estimates of Live BA and Dead BA using the Forest Vegetation Simulator (FVS), and the nonparametric Random Forests (RF) machine learning algorithm (Breiman 2001). Predicting both Live and Dead BA in bark beetle-affected forest, where live and dead trees are typically thoroughly mixed, has not been attempted before and should provide insight into the sensitivity of LiDAR to bark beetle effects on coniferous forest canopies.

2. Methods

2.1 Study areas

Areas of bark beetle-infested coniferous forests included in this study are from: 1) the Kenai Peninsula in Alaska (AK), 2) the Pinaleno Mountains of southeastern Arizona (AZ), 3) north-central Colorado (CO), 4) central Idaho (ID), and 5) central Oregon (OR). The commonality among these study areas is that both field plot and LiDAR data were collected at

the time of an active beetle outbreak in mature coniferous forest with each study area having trees of variable age, height, and density. The range of bark beetle-induced mortality was also highly variable across the field plots, ranging from 0 to 100% mortality.

2.2 Field plots

Field plot locations in AZ, CO, and ID were selected using stratified random (CO and ID) or systematic (AZ) sampling designs within the bounds of the LiDAR acquisitions. All plots were of fixed radius with radii of 12.6 m in AZ (0.05 ha), 8.0 m in CO (0.02 ha), and 11.3 m in ID (0.025 ha). Field plots in AK and OR were Forest Inventory and Analysis (FIA) plots at predetermined, systematic locations. In OR, these plots were augmented with USFS Region 6 (R6) Continuous Vegetation Survey (CVS) plots also situated along a systematic grid. Each FIA plot in AK and OR consists of 5 subplots of 7.3 m radius (0.11 ha total plot area) while the R6 plots in OR consist of 4 subplots of 15.5 m radius (0.38 ha total plot area). The subplot level data were combined and summarized at the plot level to preclude any issues of spatial autocorrelation, pseudo-replication, or inflated plot numbers. Live and Dead BA were summarized at the plot level from individual tree diameter at breast height (dbh) measurements and tree condition calls using the Forest Vegetation Simulator (FVS) while taking into account the different plot areas and then converted from English units of ft²/ac to metric units of m²/ha.

2.3 Data processing

Metrics of the height, intensity, and density of LiDAR points within the fixed-radius field plots were calculated using the CloudMetrics tool of FUSION LiDAR visualization and analysis software (McGaughey 2012). In AK and OR, the LiDAR points within the individual subplots were extracted from the point cloud and concatenated into a single file prior to calculating plot-level metrics. In all study areas, ground elevations from the LiDAR-derived digital terrain model (DTM) were subtracted from the point elevations before calculating canopy metrics. A height threshold of breast height (1.37 m) was used to separate ground and understory returns from overstory returns. Height and intensity metrics were statistics based on the distributions of height and intensity values associated with each return: percentiles (minimum, median, maximum, quartiles and interquartile range (IQ)), mode, mean, standard deviation, variance, skewness, kurtosis, coefficient of variation (CV), average absolute deviation (AAD), and canopy relief ratio (CRR; Pike and Wilson 1971). Density metrics were calculated as the percentage of both 1st and all returns >1.37 m (breast height, BH), mean, and mode; and within seven vertical height strata 0-6: <0.15m, 0.15-1.37 m, 1.37-5 m, 5-10 m, 10-20 m, 20-30 m, and >30m. Twenty topographic metrics, including elevation, slope, transformed aspect, profile and planiform curvatures, slope*cos(asp), slope*sin(asp), and heatload index were also calculated from digital terrain models (DTM) after aggregating them to 30-m resolution, using an Imagine add-on tool developed at the USFS Remote Sensing Applications Center (Ruefenacht In Prep).

2.4 Random Forests model development

Predictor variables were selected from the suite of candidate metrics generated from either FUSION (canopy height, intensity, density metrics) or the Imagine tool (topographic metrics). RF calculates variable importance values based on the number of nodes it forms within the random forest of regression trees. Those predictors exhibiting the highest Model Improvement Ratio (MIR) calculated from standardized importance values were selected (Evans and Cushman 2009, Evans *et al.* 2010). RF randomly selects variable subsets through a bootstrapping procedure that makes it robust to highly correlated data and resistant to overfitting (Breiman 2001). However, in the interest of parsimony, we also eliminated predictors if they had a Pearson correlation $r > 0.9$ with another predictor having a higher MIR value. RF randomly withholds one third of the data as an out-of-bag sample at every bootstrap iteration, so

it is not necessary to partition the data into model and validation datasets (Breiman 2001).

3. Results

In the AK, AZ, and OR field plots, dead trees comprised less than half of the BA measured and were a low proportion of Total BA in AK and AZ, whereas in CO and ID, more than half of the plot-level BA was from dead trees (Table 1).

Table 1: Summary Statistics of Plot-Level Live and Dead Basal Area (m²/ha)

Study area	Plots	Response	Minimum	1 st Quartile	Median	Mean	3 rd Quartile	Maximum
Alaska	194	Live BA	0.0	4.1	8.6	12.6	18.7	62.3
		Dead BA	0.0	0.0	1.5	3.9	5.7	24.5
Arizona	101	Live BA	0.0	9.3	32.8	35.5	57.0	126.3
		Dead BA	0.0	4.9	10.0	16.9	23.3	103.8
Colorado	113	Live BA	0.0	9.7	17.2	23.1	30.7	110.6
		Dead BA	0.0	36.8	50.8	54.9	70.6	126.7
Idaho	27	Live BA	5.0	8.7	14.3	15.3	21.2	29.5
		Dead BA	3.0	11.6	17.6	18.6	23.6	43.7
Oregon	190	Live BA	0.0	10.4	16.6	20.3	28.1	99.9
		Dead BA	0.0	0.0	1.4	3.3	4.3	27.2

Metrics of all types (height, intensity, density, and topographic) were selected based on MIR for the RF models (Table 2). In general, height and density metrics were most important in predicting Live and Total BA, whereas intensity metrics were most important in predicting Dead, %Live, and %Dead BA. An exception was OR, where intensity metrics were not important, perhaps because the dead trees comprised the lowest proportion of Total BA of any study area. Topographic metrics were generally the least important metrics in all models. Because intensity values could not be normalized across all study areas, we also tested the effect of omitting intensity metrics from the models; the result (not shown) was more density metrics selected in place of intensity metrics, but generally a lower number of selected predictors overall, and lower variance explained in all study areas except OR. Intensity metrics were more important in explaining Live BA in CO and ID, where tree mortality was higher, than in AK, AZ and OR. In AK, AZ, CO, and ID, RF Models predicting %Live and %Dead BA were almost identical in terms of important variables (Table 2) and variance explained (Table 3), which is not surprising because these responses are mutually dependent. In OR, where sampled tree mortality was lowest, RF models predicting %Live and %Dead BA were not identical.

In all study areas, the RF models explained more variance in Live BA than Dead BA (Table 3). In AK, AZ, CO, and ID, intensity metrics were very effective predictors of %Live and %Dead BA; omitting intensity metrics (not shown) reduced variance explained, especially for predicting dead BA. Models of %Live and %Dead BA explained more variance (51.2- 74.0% variance explained) than models predicting absolute Live, Dead, and Total BA (17.1- 67.9% variance explained) in these areas. In OR, where tree mortality was lowest and intensity metrics were comparatively unimportant, %Live and %Dead BA models explained less variance (13.6-24.5% variance explained) than RF models predicting absolute Live and Dead BA (27.3-59.1% variance explained). In AK, CO, ID, and OR, RF models predicting Total BA explained more variance than RF models predicting Live and Dead BA. The variance explained by RF for predicting Total BA in AZ was intermediate to that of predicting Live and Dead BA. In all study areas, omitting intensity metrics (not shown) did not greatly affect variance explained by RF

models predicting Total BA.

Table 2: LiDAR Metrics Selected for RF Models Based on MIR and Pearson $r < 0.9$

Study area	Response	Height	Intensity	Density	Topographic
Alaska	Live BA	Mode, 10 th	CV	%1 st >Mean, Strat0	Elev
	Dead BA	Max, CRR	Mean, StDev, CV	Strat4	Elev
	Total BA	Max, Mode, 30 th	Max	%1 st >Mean, Strat0, Strat4	Elev
	%Live BA	Mean, StDev, CV, Skew, CRR	Mean, StDev, CV, Skew, Kurt	%1 st >Mean, Strat4	Elev
	%Dead BA	Mean, StDev, CV, Skew, CRR	Mean, StDev, CV, Skew, Kurt	%1 st >Mean, Strat4	Elev
Arizona	Live BA	Mean, Mode, CV, 10 th , CRR	StDev, Skew	%All>BH, Strat0, Strat2	Elev
	Dead BA	Mean, Mode, CV, IQ, Skew	Max, Mean, Skew, Kurt	%1 st >BH, %All>Mode, Strat1	Elev, SlpCosAsp
	Total BA	Mean, Mode, CV, 10 th , CRR	StDev, Kurt	%1 st >BH, %All>Mode, Strat0, Strat2, Strat3, Strat4	Elev
	%Live BA	Mode, CV, 5 th , CRR	Mean, Skew, Kurt, AAD	%1 st >BH	Elev
	%Dead BA	Mode, CV, 5 th , CRR	Mean, Skew, Kurt, AAD	%1 st >BH	Elev
Colorado	Live BA	Max, Mean, Mode, Skew, 5th	Mean, Kurt	%1 st >BH, Strat2, Strat4	Heatload, Slope
	Dead BA	Mean, 99 th	Skew	%1 st >BH, %1 st >Mode, Strat4	Elev
	Total BA	Max, Mean, 25 th	Max, Mean	%All>BH, %1 st >Mode, Strat4, Strat5	Elev
	%Live BA	Min, Kurt	Mean, CV, Skew, Kurt	-	Slope
	%Dead BA	Min, Kurt	Mean, CV, Skew, Kurt	-	Slope
Idaho	Live BA	Max, CV, 1 st	CV, IQ, Kurt	%All>BH, %All>Mode, Strat1, Strat4, Strat5	Elev
	Dead BA	Mode, Skew, 30 th	Max, Mean, Skew, AAD	%1 st >Mode, Strat1	Elev, Slope, SlpCosAsp
	Total BA	Min, Max, Skew, AAD, 40 th , 75 th	-	%All>Mean, Strat5	-
	%Live BA	1 st	CV, IQ, Kurt	Strat1	Elev
	%Dead BA	1 st	CV, IQ, Kurt	Strat1	Elev
Oregon	Live BA	Mean, 20 th , 95 th , CRR	Mean, Kurt, AAD	%All>BH, Strat4	Elev
	Dead BA	Mean, Kurt, 1 st , 99 th	Mean, CV	%All>BH, Strat6	Elev
	Total BA	Mean, CV, Kurt, AAD	Mode, CV, Kurt	%All>BH, Strat4	Elev
	%Live BA	Mean, Skew, Kurt, 95 th , CRR	Mean, Mode, CV, Kurt	%All>BH, %All>Mean, Strat2, Strat3	-
	%Dead BA	Mode, CV, IQ, 1 st , 99 th , CRR	StDev, Skew, Kurt	%All>BH, %All>Mean, Strat2, Strat3	-

Table 3: % Variance Explained by RF to Predict Plot-Level Basal Area Responses

Study area	Response	Number of predictors	% Variance explained
Alaska	Live BA	6	67.2
	Dead BA	7	46.1
	Total BA	8	67.9
	%Live BA	13	59.7
	%Dead BA	13	59.1
Arizona	Live BA	11	50.2
	Dead BA	14	17.1
	Total BA	15	34.9
	%Live BA	10	74.0
	%Dead BA	10	73.4
Colorado	Live BA	12	44.3
	Dead BA	7	39.4
	Total BA	10	57.5
	%Live BA	7	51.6
	%Dead BA	7	51.2
Idaho	Live BA	12	58.0
	Dead BA	12	41.4
	Total BA	8	58.9
	%Live BA	6	67.7
	%Dead BA	6	67.7
Oregon	Live BA	10	59.1
	Dead BA	9	27.3
	Total BA	10	66.0
	%Live BA	13	24.5
	%Dead BA	13	13.6

4. Discussion

We attribute the lower variance explained for Dead BA compared to Live BA to greater penetration of the canopy by laser pulses after needles drop from dead trees, and thus a lesser amount of information than would be reflected from a green or red canopy of the same height. Tree condition calls (live or dead) at the time of field plot characterization did not coincide with the time of LiDAR acquisition; thus, we do not know whether dying or dead trees were green, red, or grey when the LiDAR was collected. Temporal mismatches between field and LiDAR data collections are likely greater in AK and OR, because only some FIA or CVS plots are characterized in a given year.

Based on this preliminary analysis, we conclude that it may be difficult to generalize a consistent subset of LiDAR metrics that work across these different affected forest ecosystems to predict bark beetle effects on BA. It is more likely that we may be able to generalize analytical approaches, if not specific metrics. In future work, we will include disturbance vectors derived from Landsat image time series in the analysis, to more accurately predict the BA response variables. Integrating temporal disturbance vectors from Landsat with airborne LiDAR metrics that capture fine spatial detail should prove a more powerful approach for predicting canopy structure effects of bark beetle-induced tree mortality.

5. Conclusion

This analysis tested the feasibility of predicting both Live and Dead BA in bark beetle-affected coniferous forests from discrete-return LiDAR. We found that LiDAR-derived height, intensity, density, and topographic metrics explained more variance in Live BA than Dead BA, and usually the most variance in Total BA, which we conclude is due to the greater sensitivity of LiDAR to needled trees than to grey trees.

Acknowledgements

Funding was provided by the Special Technology Development Program of the U.S. Forest Service Forest Health Protection division. Funding to the University of Idaho was through Joint Venture Agreement 11-JV-11221633-184.

References

- Breiman, L., 2001. Random forests. *Machine Learning*, 45, 5-32.
- Bright, B.C., Hicke, J.A. and Hudak, A.T. 2012. Estimating aboveground carbon stocks of a forest affected by mountain pine beetle in Idaho using lidar and multispectral imagery. *Remote Sensing of Environment*, 124: 270-281.
- Coops, N.C., Varhola, A., Bater, C.W., Teti, P. Boon, S., Goodwin, N., and Weiler, M., 2009. Assessing differences in tree and stand structure following beetle infestation using lidar data. *Canadian Journal of Remote Sensing*, 35: 497-508.
- Evans, J.S. and Cushman, S.A., 2009. Gradient modeling of conifer species using Random Forests. *Landscape Ecology*, 5, 673–683.
- Evans, J.S., Murphy, M.A., Holden, Z.A. and Cushman, S.A., 2010. Modeling species distribution and change using Random Forests. In: C.A. Drew, F. Huettmann, Y. Wiersma (Eds.). Chapter 8 in *Predictive Modeling in Landscape Ecology*. Springer.
- Hicke, J.A., Allen, C.D., Desai, A.R., Dietze, M.C., Hall, R.J., Hogg, E.H., Kashian, D.M., Moore, D., Raffa, K.F., Sturrock, R.N. and Vogelmann, J., 2012. Effects of biotic disturbances on forest carbon cycling in the United States and Canada. *Global Change Biology*, 18, 7-34.
- Hudak, A.T., Strand, E.K., Vierling, L.A., Byrne, J.C., Eitel, J., Martinuzzi, S. and Falkowski, M.J., 2012. Quantifying aboveground forest carbon pools and fluxes from repeat LiDAR surveys. *Remote Sensing of Environment*, 123: 25-40.
- Kim, Y., Yang, Z., Cohen, W.B., Lauver, C.L., and Vankat, J.L., 2009. Distinguishing between live and dead standing tree biomass on the North Rim of Grand Canyon National Park, USA using small-footprint lidar data. *Remote Sensing of Environment*, 113: 2499-2510.
- Kurz, W.A., Dymond, C.C., Stinson, G., Rampley, G.J., Neilson, E.T., Carroll, A.L., Ebata, T. and Safranyik, L. 2008. Mountain pine beetle and forest carbon feedback to climate change. *Nature*, 452, 987-990.
- McGaughey, R.J., 2012. FUSION/LDV: Software for LIDAR Data Analysis Visualization. USDA Forest Service, Pacific Northwest Research Station. 170 p.
- Negron, J.F., Allen, K., Cook, B. and Withrow, J.R. Jr., 2008. Susceptibility of ponderosa pine, *Pinus ponderosa* (Dougl. Ex Laws.), to mountain pine beetle, *Dendroctonus ponderosae* Hopkins, attack in uneven-aged stands in the Black Hills of South Dakota and Wyoming USA. *Forest Ecology and Management*, 254: 327-334.
- Pfeifer, E.M., Hicke, J.A., and Meddens, A.J.H., 2011. Observations and modeling of aboveground tree carbon stocks and fluxes following a bark beetle outbreak in the western United States. *Global Change Biology*, 17, 339-350.
- Pike, R.J. and Wilson, S.E., 1971. Elevation relief ratio, hypsometric integral, and geomorphic area altitude analysis. *Bulletin of the Geological Society of America*, 82, 1079–1084.
- Ruefenacht, B. (In Prep) Digital elevation model derivatives. USDA Forest Service Remote

Sensing Applications Center, Salt Lake City, Utah.