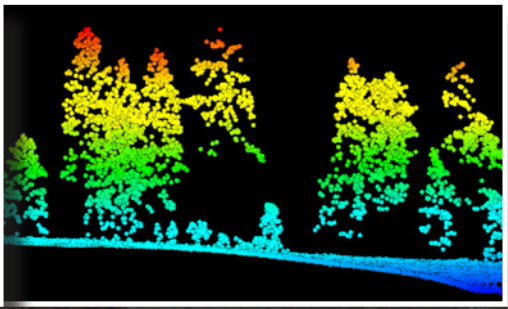
United States Department of Agriculture

Forest Service

Creating Lidar Canopy Structure Layers and Inventory Models for Evaluating Northern Goshawk Habitat Quality

May 2015

RSAC-10070-RPT1



Abstract

The northern goshawk *(Accipiter gentilis)* is listed as a sensitive species and there is a need to understand the relationship between long-term goshawk demographic performance and three-dimensional (3-D) structure of its forest habitat. Before the availability of lidar, it was not possible to relate detailed landscape-level forest structural data to the demography of a species. The Kaibab Plateau in Arizona benefits from a unique combination of high quality lidar data and extensive, long-term field data describing northern goshawk demographic performance. This allows for an in-depth investigation of how the 3-D canopy structure, as sampled by lidar, can help identify which of the 3-D structural elements of forest vegetation confer habitat quality for the goshawk. In this phase of the project we used FUSION software to generate a collection of lidar-derived 3-D canopy structure derivatives, at a 20 m resolution, that describe the height and density of the canopy across the entire Kaibab Plateau. In addition, lidar forest inventory models were generated using an area-based approach to model forest inventory parameters of interest, with R^2 values ranging from 0.19 to 0.75. The next phase of this project will use the 3-D canopy structure layers and landscape inventory models as a baseline to explore and understand the links between 3-D canopy structure and goshawk demographic performance on the Kaibab Plateau.

Keywords

Lidar, goshawk, wildlife habitat, canopy structure, forest inventory modeling, landscape modeling, Kaibab plateau

Authors

Brent Mitchell is a remote sensing specialist employed by RedCastle Resources at the Remote Sensing Applications Center (RSAC) in Salt Lake City, Utah.

Mark Beaty is a remote sensing specialist employed by RedCastle Resources at RSAC.

Richard Reynolds is a research wildlife biologist at the Forest Service Rocky Mountain Research Station in Fort Collins, Colorado.

Tom Mellin is a remote sensing coordinator working at the Forest Service Southwestern Regional Office in Albuquerque, New Mexico.

Andrew T. Hudak is a research forester at the Forest Service Rocky Mountain Research Station in Moscow, Idaho.

Abigail Schaaf is the sponsored projects group leader and remote sensing specialist, employed by RedCastle Resources at RSAC.

Haans Fisk is the program leader for the Remote Sensing Evaluation, Application, & Training Program at RSAC.

Mitchell, B.; Beaty, M.; Reynolds, R.; Mellin, T.; Hudak, A.T.; Schaaf, A.; Fisk, H. 2015. Creating lidar canopy structure layers and inventory models for evaluating northern Goshawk habitat quality. RSAC-10070-RPT1. Salt Lake City, UT: U.S. Department of Agriculture, Forest Service, Remote Sensing Applications Center. 21 p.

Table of Contents

Introduction

The northern goshawk *(Accipiter gentilis)* is an apex predator that lives in the sub-canopy of forests throughout the United States and Canada. It is a sensitive species or a species of special concern in six of the nine U.S. Department of Agricultural Forest Service Regions, all United States Fish and Wildlife Service Regions, six Bureau of Land Management national conservation lands, many U.S. State jurisdictions, Canadian Provinces, and is a threatened species in British Columbia. To date, the Forest Service has experienced three decades of expensive litigation in response to concerns that forest management affects goshawk viability by reducing its habitat quality. This has created a need to understand the relationship between long-term goshawk demographic performance and 3-D structure of its forest habitat.

On the Kaibab National Forest the northern goshawk (hereafter referred to as the "goshawk"), demographic performance has been studied for 21 years (1991 to 2011) in 125 territories across the Kaibab Plateau in Arizona. Associated field data contains information pertaining to territory occupancy, mate and territory fidelity, reproduction, turnover, recruitment, immigration, and juvenile and adult survival rates. Based on syntheses of the habitat needs of the goshawk and the habitats of their bird and mammal prey, Reynolds and others (1992) developed a set of management recommendations to sustain goshawk populations through the management of forest structure.

Lidar technology has proven capable of directly measuring the vertical and horizontal forest structure and sub-canopy structure across large landscapes (Lefsky and others 2002, Reutebuch and others 2005). Before advances in remote sensing technology—i.e., lidar—natural resource managers were limited to characterizing wildlife habitat based on field data of a limited spatial extent in conjunction with optical remote sensing data that does not characterize vertical structure within the forest canopy (Vierling and others 2008). This method restricts managers in accurately characterizing 3-D forest canopy structure across large landscapes. Given the increased availability of lidar data, numerous studies have recently been conducted that explore lidar-derived 3-D canopy structure information and how it relates to wildlife demographic data. Several studies have demonstrated the use of lidar data as a predictive tool for species distributions based on known natural history of the species. For example, Nelson and others (2005) suggest great potential in identifying tall, dense forests that are the preferred habitat of the endangered Delmarva fox squirrel *(Sciurus niger cinereus)*. Alternatively lidar can be used to better understand habitat selection by species that have a known distribution. Broughton and others (2006) used existing territory maps of marsh tits *(Poecile paustris)* and documented substantial vegetation structural differences between the occupied territories and the adjacent areas not occupied by the birds. Their results indicated that marsh tits occupied sites composed of mature trees with a sub-canopy shrub layer and avoided sites containing many small, young trees. Vogeler and others (2013) and Hagar and others (2014) compared occupancy rates of the brown creeper *(Certhia americana)* and marbled murrelet *(Brachyramphus marmratus)*, respectively, with lidar canopy descriptor statistics describing height and density, both of which were relatively strong predictors of occupancy. Goetz and others (2007) illustrated links between vertical forest structure diversity and bird species diversity.

There has also been work that indicates there is a strong agreement between vegetation structural measurements derived from lidar and ground field measurements; this suggests that lidar measurements can supplement field-derived vegetation structural measurements used to characterize avian habitat (Hyde and others 2005, 2006). Lidar data have not only been used to describe the physical structure of animal habitat but also to indicate the quality of existing habitat across large landscapes. Hill and others (2004) and Hinsley and others (2002) illustrated this capability by associating canopy height to nesting chick body mass, which is a surrogate for breeding success and territory quality. This relationship allowed them to predict habitat quality across the entire lidar acquisition.

The Kaibab Plateau in Arizona has a unique combination of high quality lidar data and extensive, long-term field data describing northern goshawk demographic performance. This allows for an in-depth investigation of how the 3-D canopy structure, as sampled by lidar, can help identify which of the 3-D structural elements of the forest vegetation confer habitat quality for the goshawk. Before the availability of lidar, the ability to relate detailed landscape-level forest structural data to the demography of a species was unavailable.

To test the efficacy of the management recommendations and to empirically identify the forest structural elements that confer habitat quality for the goshawk, Richard Reynolds from the Rocky Mountain Research Station and colleagues proposed to acquire lidar data for the entire study area in order to relate the long-term demographic performance (total reproduction, survival, site fidelity) of goshawks on each of the 125 territories to the 3-D forest structure of those territories. In the summer of 2012 high density lidar data was acquired covering the entire Kaibab Plateau. The objective of this Remote Sensing Steering Committee-sponsored project was to provide lidar-derived 3-D canopy structure derivatives and plot-based lidar forest inventory models and associated GIS layers for the Kaibab Plateau goshawk study area. In addition, technology transfer efforts have and will be provided by the Remote Sensing Applications Center (RSAC) to ensure the cooperators are informed about the

above-mentioned products and apply them in an appropriate manner when building northern goshawk habitat and demographic models.

Study Area

The study area for this project was the Kaibab Plateau, located in Arizona and encompassed by the Kaibab National Forest and Grand Canyon National Park-North Rim (figure 1). The Kaibab Plateau comprises 1800 km^2 of ponderosa pine and mixed conifer forests, predominately above 2,200 m. The landscape also includes meadows and forest openings created from disturbances such as high-severity fire, wind throw, and timber harvesting. The Kaibab Plateau provides suitable habitat for the goshawk, which occupies the principal forest types of ponderosa pine, mixed conifer and spruce-fir of the Southwest (Reynolds and others 1992).

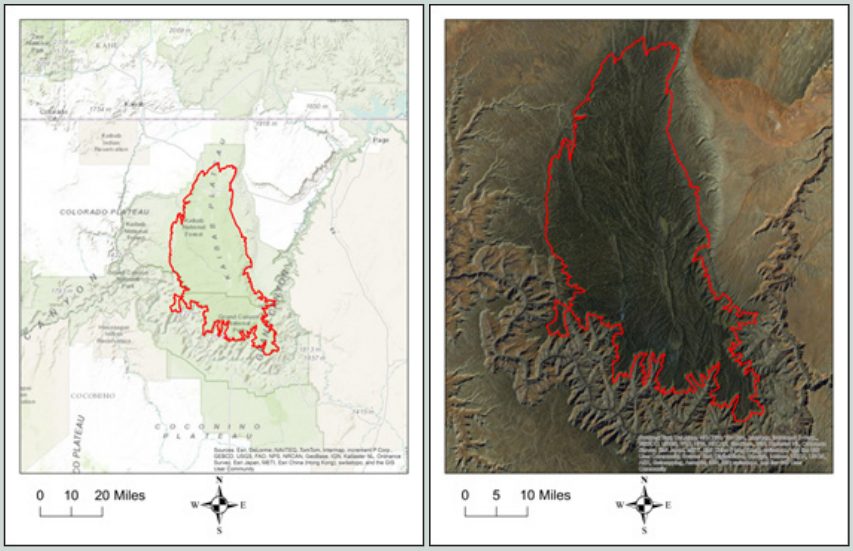


Figure 1—The study area for this project is the Kaibab Plateau, located in Arizona within the Kaibab National Forest and the Grand Canyon National Park. The figure on the left provides a regional perspective of the study area while the figure on the right provides a more detailed large-scale perspective of the landscape with NAIP imagery as the background.

Data

The data required for this phase of the project included fully preprocessed airborne lidar data and associated field plot data. The two datasets were collected during two consecutive growing seasons. The field plot data were collected with protocols designed to ensure they were accurately geolocated and with a fixed-radius plot design so the plots would be appropriate for use with high-resolution lidar data.

Field Data

In addition to the lidar-derived 3-D canopy structure derivatives (relative measures of the height and density of the canopy), it was also desirable to generate forest inventory models (timber volume, biomass, basal area, etc.) and explore their potential relationship with goshawk habitat. To do this successfully, lidar-specific field data were needed so relationships between the lidar plot measurements and field-derived inventory plot measurements could be modeled.

To ensure the training data (field measurements) were appropriate for model development, the Kaibab Plateau was divided into three main land cover types based on forest stand data: forest, woodlands (pinyon-juniper), and burned areas. The forest land cover type was given highest priority for field plot collection as forested land covers the majority of the Kaibab Plateau and makes up the most significant portion of the goshawk habitat. In addition, lidar-derived inventory models have been shown to work well in forest environments. A limited number of plots were collected in the burned areas and no plots were collected in the woodlands as that class made up only a small fraction of the land cover within the study area.

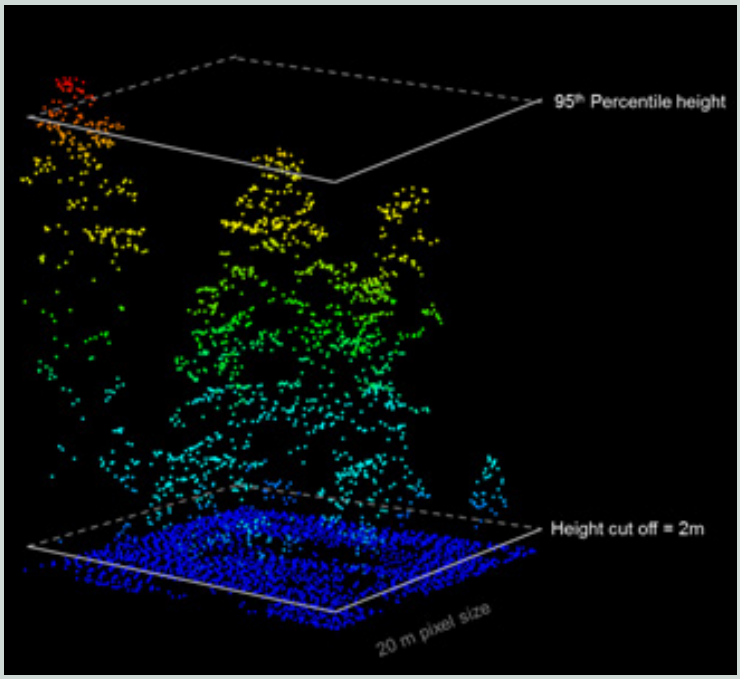
Sampling Strategy

Stratified sampling has been shown to be more efficient than random sampling for lidar forest inventory modeling (Hawbaker and others 2009). A stratified sample design produces a greater range of attribute variability and minimizes extrapolation beyond the range of the observed field data in the resulting predictive regression models.

Two lidar-derived 3-D canopy structure derivatives—dominant height (95th percentile height), illustrated in figure 2, and vegetation density (percentage of all returns above 3 m), illustrated in figure 3—were used to create structural classes for the stratified sampling scheme for the forested land cover type. Four equal-area quartiles were created for both the dominant height and vegetation density layers. Sixteen structural strata were then created to include all combinations of dominant height and vegetation density. Data was collected from a minimum of six field plots within each stratum with a total of ninety-seven forested plots across the landscape (table 1).

In recently burned areas, where the canopy structure is less variable, a simpler stratification was used, with two equal-area classes of height and density. Data was collected from 19 plots in those areas. All field data was collected in the summers of 2013 and 2014 from 0.1 acre (.04 ha) circular plots.

95th Percentile height of all returns > 2 m



95th Percentile landscape metric

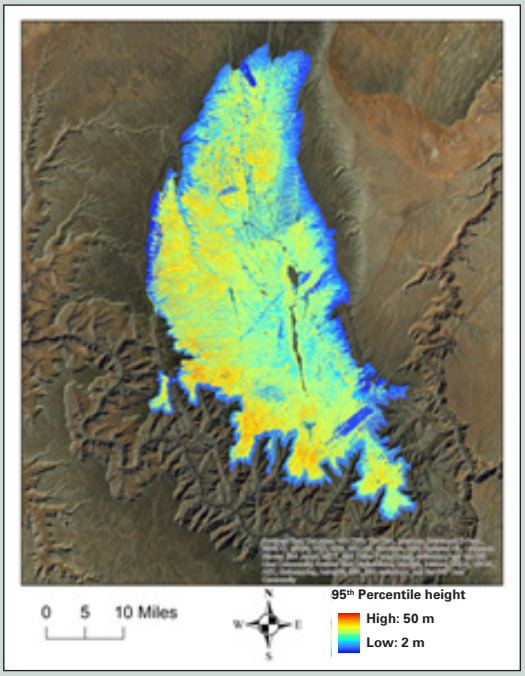
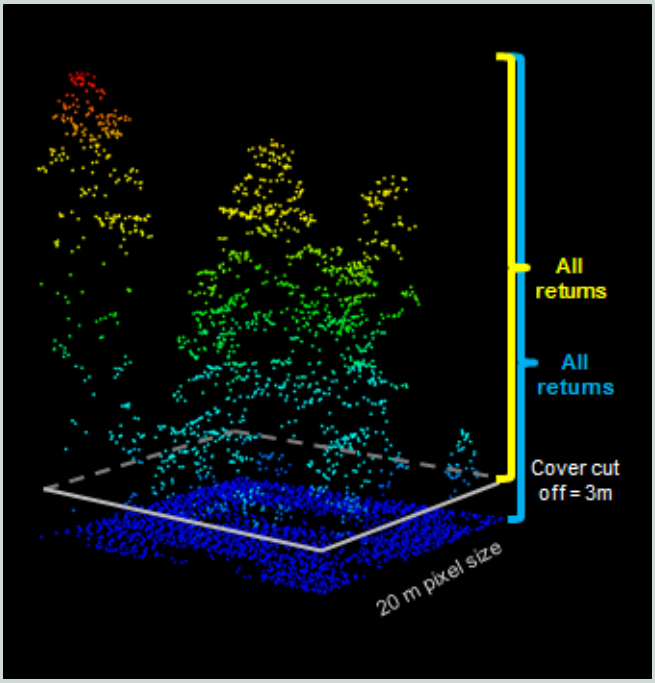


Figure 2—The point cloud graphic on the left illustrates the logic for computing the 95th percentile height. For height statistics, returns from the ground and any vegetation below 2 m are excluded from the calculation. The GIS layer on the right depicts the 95th percentile height calculated at 20 m resolution for the study site.

Vegetation density (all returns above 3 m / total number of all returns)*100



Vegetation density landscape metric

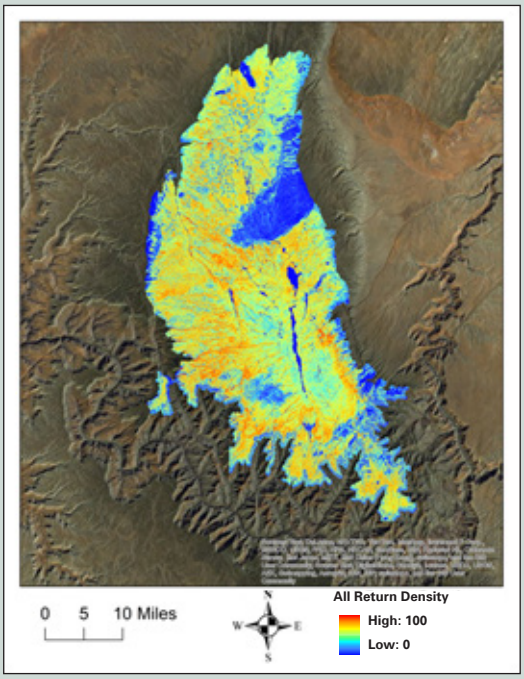


Figure 3—The point cloud graphic on the left illustrates the logic for computing the all-return vegetation density above 3 m. The GIS layer on the right depicts the all-return vegetation density metric calculated at 20 m resolution for the study site.

Field Work Preparation

Location maps for all plots to be measured were created to assist the field crew. For each plot, an aerial photo was used as backdrop with the plot marked by a circle in the predetermined location. A subset was also clipped from the lidar data at each location, which provided the field crew an additional 3-D visualization of the desired plot location. The field crew used these map products and the predetermined plot coordinates to navigate to approximate plot locations. The GPS coordinates of the plot center as identified on site became the official coordinates used for further analysis.

Plot Protocol

At each of the 116 randomly selected locations a fixed-radius plot with a radius of 37.2 ft was delineated. Each plot consisted of an outer (37.2 ft radius) and inner plot (16.7 ft radius). At each plot the slope, aspect, and landform position was recorded; plus four photographs in each cardinal direction were taken facing towards plot center from the outer plot perimeter. Within the full plot all trees (live and dead) greater than or equal to 8" in diameter were measured for species, diameter at breast height (DBH), total height, height to live crown base, crown diameter in two directions, and distance and compass direction from plot centers. Within the inner plot, all trees less than 8" DBH were measured and tallied by species into trees over or under 6 ft tall and dominant shrubs were tallied by species and height (over or under 6 ft), and measured for average height for each category. In addition, within the inner plot, percentages of ground cover, total vegetation cover (trees, shrubs, herbaceous, and graminoid), and cover of dominant trees, dominant shrubs, and herbaceous, and graminoid vegetation were occularly estimated. Coarse woody debris (over 3", large end diameter) were also inventoried in the inner plot. Each piece of woody debris was measured for small and large end diameters and total length.

Table 1—Sampling strata, height and density thresholds, and number of plots collected

Stratum	P95 Height	Density	Number of Plots Collected
1Q1	2 to 17 m	<29%	6
2Q1	17 to 22 m	<29%	6
3Q1	22 to 27 m	<29%	6
4Q1	>27 m	<29%	6
1Q2	2 to 17 m	29% to 40%	6
2Q2	17 to 22 m	29% to 40%	6
3Q2	22 to 27 m	29% to 40%	6
4Q2	>27 m	29% to 40%	6
1Q3	2 to 17 m	40% to 49%	6
2Q3	17 to 22 m	40% to 49%	6
3Q3	22 to 27 m	40% to 49%	6
4Q3	>27 m	40% to 49%	6
1Q4	2 to 17 m	49% to 100%	7
2Q4	17 to 22 m	49% to 100%	6
3Q4	22 to 27 m	49% to 100%	6
4Q4	>27 m	49% to 100%	6

GPS Data Collection

When using lidar and field data to model forest inventory parameters it is imperative that sub-meter plot location accuracy be obtained. This relatively high level of positional accuracy is needed to minimize error and maximize correlation between field and lidar data in the modeling environment.

At plot center a sub-meter GPS location was collected using a Trimble GeoXH6000. Each point was collected using GPS + GLONASS, accuracy based logging settings, for a minimum of 10 minutes. Logging was paused if accuracy was worse than 1 m, and resumed when accuracy was 1 m or better. Differential correction was applied using Pathfinder Office referenced from a single base station in Fredonia, Arizona.

All plot data were entered into Trimble Juno Personal Data Recorders (PDR) using FSVeg software so that forest inventory parameters could be generated for each plot using the Forest Vegetation Simulator (FVS).

Lidar Data

High pulse-density lidar data were acquired for approximately 185,000 ha within the Kaibab Plateau study area between August 25 and September 15, 2012. The dataset met or slightly exceeded the minimum recommended specifications for forest inventory modeling[1] with a nominal pulse density of 10 pulses/m^2, greater than 50 percent side lap, and a scan angle within 14° of nadir. The lidar data was projected in UTM zone 12 North.

Methods

Generating Lidar 3-D Canopy Structure Layers at the Landscape Scale

FUSION software (McGaughey 2014) was used to generate a collection of lidar-derived 3-D canopy structure derivatives that describe the height and density of the canopy (table 2). All the canopy structure derivatives were generated from the raw lidar point cloud within the boundaries of each 3-D grid cell, with the grid spanning the entire lidar collection ("gridmetrics"). All gridmetrics were computed using a 20-m cell size, which corresponds to approximately the same area as a 0.1 acre field plot. All height gridmetrics were calculated for the canopy using returns above a height cut-off of 2 m to exclude the ground and low-lying vegetation from influencing the "canopy height descriptor" statistics. Density gridmetrics were calculated using a static cover threshold (canopy cut-off) of 3 m and the dynamic thresholds of mean and mode values within each grid cell. Density strata (height slices) were selected based on the natural history of the goshawk and its foraging behaviors (Reynolds and others 1992). Figures 2 and 3 provide a look into the logic and the resulting landscape-scale outputs for the 95th percentile height and the percentage of all returns above 3 m (vegetation density), respectively. Landscape products such as these were generated for all the metrics listed in table 2. In addition to the strata layers listed in table 2, "relative density strata" layers were also generated for all the same height breaks. Relative density layers differ from the standard density strata layers in that the proportion is calculated by taking the number of returns in the stratum of interest and dividing it by the sum of points in the stratum and the remaining points below the stratum of interest. This is done in an attempt to normalize the effects of the canopy above the strata of interest. For additional explanation of the logic used for calculating the lidar metrics please refer to "First Order Lidar Metrics: A supporting document for lidar deliverables" (http://www.fs.fed.us/eng/rsac/lidar_training/pdf/LidarMetricsDescriptionOfDeliverables_Generic_12_15_14.pdf)

Generating Field Inventory Attributes at the Plot Scale

For the field data to be useful in lidar modeling applications the "area-based approach" methodology needs to be applied. This approach sums individual tree metrics to the plot level and creates relationships between plot sums and lidar descriptor metrics for those same plot areas. To get the plot-based estimates from the field data, individual tree data collected at each field plot were entered into the Region 3 variant of the Forest Vegetation Simulator (FVS) and estimates of forest inventory attributes were created for each of our 116 field plots (table 3). Note that forest inventory attributes are in English units per request of cooperators.

It should be mentioned that a majority of the plots were collected in the Ponderosa Pine forest type and the inventory models will perform the best in that forest type. The number of plots collected in each of the different forest types is presented in table 4. In an effort to provide more insight into the composition of the forest and the sampled plots the dominance types of the plots is summarized in appendix A.

Generating Lidar Predictor Statistics at the Plot Scale

In the area-based modeling approach, each plot needs to be "clipped from the lidar" and normalized to height above ground. Summary statistics describing the density and height distributions are then calculated for each plot, using exactly the same parameters and thresholds as for calculating the gridmetrics described in the earlier section (statistics listed in table 2). Using the x, y coordinates and the plots radius of 37.2 ft, the plots were clipped from the lidar data and normalized to height above ground using FUSION's *Clipdata* command. The height and density metrics were then calculated using FUSION's *Cloudmetrics* command. The resulting product is a table with each plot associated with a complete list of canopy metrics for height and density that will be used as potential predictors in the inventory modeling efforts.

[1] Discrete lidar data continues to prove useful in many natural resource applications. However, not all lidar datasets are equal. The most important single characteristic, perhaps, that determines the appropriate use of a lidar dataset is the mean number of pulses/m^2. For example, relatively low pulse-density data (0.5 to 1 pulse/m^2) is typically only useful for bare earth or terrain models. Medium pulse-density data (1 to 3 pulses/m^2) has the additional potential of providing canopy height models. Forest structure information for resolving individual trees, however, requires relatively high pulse-density data (typically 3 pulses/m^2 or higher).

Table 2—Lidar metrics generated at landscape and plot scales using FUSION software

Category	Output variable
Height distribution	Total number of returns
	Count of returns by return number
	Minimum
	Maximum
	Mean
	Median (output as 50th percentile)
	Mode
	Standard deviation
	Variance
	Coefficient of variation
	Interquartile distance
	Skewness
	Kurtosis
	AAD (Average absolute deviation)
	L-moments (L1, L2, L3, L4)
	L-moment skewness
	L-moment kurtosis
Height percentiles	1st, 5th, 10th , 20th, 25th, 30th, 40th, 50th, 60th, 70th, 75th, 80th, 90th, 95th, 99th percentiles
Density indices	Percentage of first returns above a specified height (canopy cover estimate)
	Percentage of first returns above the mean height
	Percentage of first returns above the mode height
	Percentage of all returns above a specified height (canopy density estimate)
	Percentage of all returns above the mean height
	Percentage of all returns above the mode height
	Number of all returns above a specified height / total first returns * 100
	Number of all returns above the mean height / total first returns * 100
	Number of all returns above the mode height / total first returns * 100
Strata density layers	Proportion of returns from 0–1 m compared to total returns
	Proportion of returns from 1–2 m compared to total returns
	Proportion of returns from 2–4 m compared to total returns
	Proportion of returns from 4–7 m compared to total returns
	Proportion of returns from 7–10 m compared to total returns
	Proportion of returns from 10–13 m compared to total returns
	Proportion of returns from 13–16 m compared to total returns
	Proportion of returns from 16–19 m compared to total returns
	Proportion of returns from 19–22 m compared to total returns
	Proportion of returns from 22–25 m compared to total returns
	Proportion of returns from 25–28 m compared to total returns
	Proportion of returns from 28–31 m compared to total returns
	Proportion of returns from 31–34 m compared to total returns
	Proportion of returns from 34–37 m compared to total returns
	Proportion of returns from 37–40 m compared to total returns
	Proportion of returns from 40–43 m compared to total returns

Table 3—Inventory attributes calculated for individual trees and summarized to the plot level

Attribute	Description/Units of Measurement
Biomass (STBIOMASS)	Tree biomass Units: Tons per acre
Quadratic mean diameter (QMD)	DBH of the tree of average basal area, based on live trees only Units: Inches
Stand basal area (SBA)	Units: Square feet per acre
Total cubic foot volume (TCuFT)	Units: Cubic feet per acre
Merchantable board foot volume (MBdFT)	Units: Cubic feet per acre
Merchantable cubic foot volume (MCuFT)	Units: Cubic feet per acre
Stand density index–Reineke (SDI_R)	An expression of relative stand density based on the predictable relationship between average tree size and trees per unit area in dense stands Units: Percentage
Stand density index–Zeide (SDI_Z)	An expression of relative stand density based on the predictable relationship between average tree size and trees per unit area in dense stands Units: Percentage
Largest diameter tree (LDIA)	Units: Inches
Height of largest diameter tree (HgtLDia)	Units: Feet
Nominal DBH (NDia)	Computed by averaging the 9 sample trees centered on the 70th percentile tree Units: Inches
Nominal height (NHgt)	Computed by averaging the 9 sample trees centered on the 70th percentile tree Units: Feet
Height of tallest tree (HgtTTree)	Units: Feet
Crown base height (CrBsHgt)	Units: Feet
Crown bulk density (CrBlkDn)	Units: Pounds per cubic foot
Trees per acre (TPA)	Units: Number per acre
Canopy cover (CanCov)	Corrected for overlap, random tree distribution Units: Percentage
Surface fuels of herbs and shrubs (HerbShrb)	Live surface fuels of herbs and shrubs Units: Tons per acre
Canopy load biomass (CLBiomss)	Includes crown and foliage, not stem Units: Tons per acre
Lorey's mean height (BA_WT_HGT)	Mean stand height weighted by basal area Units: Feet
Vegetation structural stage (VSS)	Units: Categorical

Forest Inventory Model Development

The field inventory attributes from FVS and the lidar-derived plot-level height and density metrics were combined into one modeling table in preparation for exploring their relationships. Many of the lidar statistics are very similar; therefore, to reduce collinearity during regression modeling, variables with a correlation of greater than 0.9 were identified and all but one were removed.

Table 4—Number of field plots collected in each forest type.

Forest type	Number of field plots collected
Ponderosa pine	73
Dry mixed-conifer	18
Wet mixed-conifer	18
Aspen	5
Pinyon, juniper mixed with ponderosa	2

Next, potential predictors for each field inventory parameter were identified from the available lidar predictors. *Regsubsets,* which is contained within the *leaps* R statistical package (http://cran.r-project.org), was used to narrow down our potential predictors. *Regsubsets* is a regression subset selection process and was implemented using a sequential replacement method to identify the top four predictor combinations with up to three variables. Linear regression models were then created with the selected predictor combinations and evaluated based on R^2 values, collinearity, standard error and model significance. We favored the simplest model while still trying to maximize model fit. Logic was applied during model selection to try to include both height and density metrics as predictors to keep the models robust in a variety of canopy structure scenarios.

Results and Discussion

Final Inventory Models

The final inventory models are highlighted in table 5 and are ordered based on their R^2 values, highest to lowest. Overall the models for volume were the strongest, with R^2 values of 0.74 and above. Mean height and density above mean height were the most strongly correlated lidar predictor variables for volume, which makes sense since taller trees with larger upper canopies generally have higher timber volume. The tree biomass model was also strong, with an R^2 value of 0.68, with the same two lidar predictors of mean height and density above mean height. Volume and biomass work well in the area-based modeling approach as they are both calculated with relatively objective field measurements and are driven by height and density directly. The remaining models that had R^2 values above 0.6 were the tree height inventory parameters: height of tallest tree, height of largest diameter tree and nominal height. Again these are directly measured in the field and are driven by the lidar parameter of maximum height. It is worth noting that lidar accurately measures dominant tree height to within generally 1 m accuracy and using the area-based modeling approach might be unnecessary for at least the height of the tallest tree. Rather, it might be more prudent to use the lidar-derived 3-D canopy structure derivatives of maximum height or the 95th height percentile when making inferences about the tree canopy height across large landscapes.

The models for basal area and Lorey's mean height had R^2 values of 0.55 and 0.58, respectively. The best lidar predictors for basal area were percentage of returns above the mean, relative density of returns between 2 and 4 m, and relative density of returns between 19 and 22 m. Of interest is that the relative density of returns between 2 and 4 m was negatively correlated with basal area, which would indicate that plots with larger trees had less dense understory. The last models with R^2 values above 0.5 were the stand density indices as calculated using the Zeide and Reineke methods. Similar to basal area, they were both positively correlated with density above mean height and negatively correlated with the relative density of returns between 2 and 4 m. The rest of the models had R^2 of less than 0.5 and will not be discussed specifically model by model. In general, the poorly performing models ($R^2 < 0.5$) are either not directly related to height and density in a linear way or they are measured in the field using subjective techniques or estimated with models that make large assumptions. For example, we would expect canopy biomass ($R^2 = 0.41$) to have strong agreement with the lidar data using the area-based modeling approach, but canopy biomass is challenging to measure in the field and is probably inconsistently measured. On the other hand, lidar samples the canopy very consistently, so disagreement between the two datasets is not surprising. This would also be the case for canopy cover ($R^2 = 0.35$) and crown bulk density ($R^2 = 0.2$). As mentioned above for the height metrics, canopy cover is probably better estimated across the landscape using the 3-D canopy density metrics derived from lidar first returns.

Table 5—Summary of the best linear regression forest inventory models created, using lidar predictor variables to model field-derived forest inventory attributes, sorted by descending linear fit R^2 values ***(continued on next page)***

Predicted parameter	P1	P2	P3	Linear Fit R^2	Linear equation
Merchantable cubic foot volume (MCuFt)	mean height	density above mean height	NA	0.75	610.071658 + 1.365251 * "mean_height" - 73.800300 * "density above mean height" + 12.100059 * ("mean_height" * "density above mean height")
Total cubic foot volume (TCuFt)	mean height	density above mean height	NA	0.74	629.475303 + 2.631671 * "mean_height" - 71.618261 * "density above mean height" + 12.580401 * ("mean_height" * "density above mean height")
Merchantable board foot volume (MBdFt)	mean height	density above mean height	rel_forest_ strata_28_31	0.74	z (- 14765.4563) + 1537.0163 * "mean_ height" + 345.7255 * "density above mean height" + 29055.6962 * "rel_ forest_strata_28_31"
Height of largest diameter tree (HgtLDia)	max height	NA	NA	0.68	(- 17.72583) + 3.56277 * "max_height"
Sum of tree biomass (STBiomss)	mean height	density above mean height	NA	0.68	9.4581254 + .3977259 * "mean_height" - 0.7182464 * "density above mean height" + 0.1727576 * ("mean_height" * "density above mean height")
Height of tallest tree (HgtTTree)	max height	NA	NA	0.68	(- 19.105951) + 3.706713 * "max_height"
Nominal height is computed by averaging the 9 sample trees centered on the 70th percentile tree (NHgt)	max height	height skewness	NA	0.66	(- 1.056533) + 2.36759 * "max_height" - 11.954539 * "height_skewness"
Lorey's mean height (BA_WT_HGT)	max height	mean height	NA	0.58	(- 18.371106) + 1.588907 * "max_height" + 2.889848 * "mean_height"
Stand basal area (SBA)	density above mean height	rel_forest_ strata_2_4	rel_forest_ strata_19_22	0.55	24.707406 + 3.885735 * "density above mean height" - 116.166478 * "rel_forest_ strata_2_4" + 108.999229 * "rel_forest_ strata_19_22"
Stand density index - Reineke (SDI_R)	density above mean height	rel_forest_ strata_2_4	NA	0.53	19.658145 + 7.573155 * "density above mean height" - 153.450886 * "rel_forest_ strata_2_4"
Stand density index - Zeide (SDI_Z)	density above mean height	rel_forest_ strata_2_4	NA	0.53	19.025117 + 7.373651 * "density above mean height" - 150.590236 * "rel_forest_ strata_2_4"
Canopy load biomass, includes crown and foliage, not stem (CLBiomss)	mean height	rel_forest_ strata_19_22	NA	0.41	(- 0.3880573) + 0.8969171 * "mean_ height" + 24.6236621 * "rel_forest_ strata_19_22"
Quadratic mean diameter (QMD)	mean height	height covariance	NA	0.40	(- 2.657393) + 1.060491 * "mean_height" + 12.494121 * "height_CV"

Table 5-C*ontinued*— Summary of the best linear regression forest inventory models created, using lidar predictor variables to model field-derived forest inventory attributes, sorted by descending linear fit R^2 values

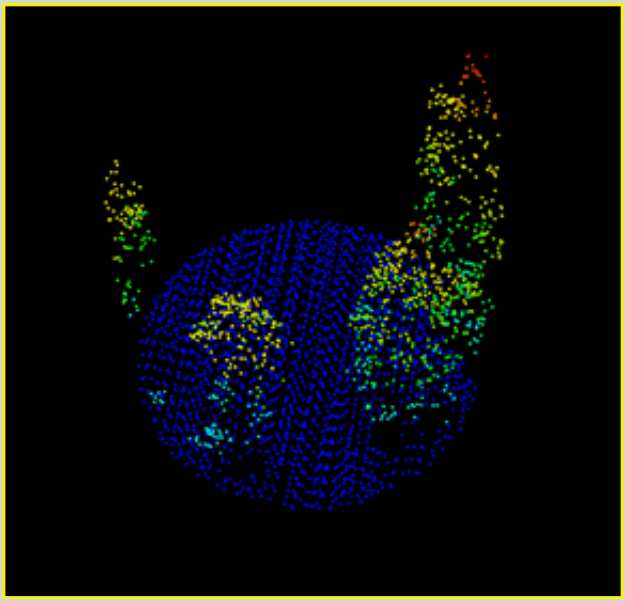
Predicted parameter	P1	P2	P3	Linear Fit R^2	Linear equation
Largest diameter tree (LDia)	max height	NA	NA	0.39	1.217215 + .871298 * "max_height"
Nominal DBH is computed by averaging the 9 sample trees centered on the 70th percentile tree (NDia)	mean height	forest_ strata_2_4	NA	0.39	1.395438 + 1.087181 * "mean_height" + 25.774645 * "forest_strata_02_04"
Canopy cover, corrected for overlap, random tree distribution (CanCov)	density above mean height	rel_forest_ strata_16_19	NA	0.35	9.6643313 + 0.5653002 * "density above mean height" + 34.2516683 * "rel_ forest_strata_16_19"
Trees per acre (TPA)	rel_forest_ strata_10_13	NA	NA	0.34	26.92284 + 174.40535 * "rel_forest_ strata_10_13"
Crown bulk density (CrBlkDn)	rel_forest_ strata_10_13	NA	NA	0.2	0.02330149 + .10922510 * "rel_forest_ strata_10_13"
Crown base height (CrBsHgt)	mean height	NA	NA	0.19	(- 0.2501845) + 1.4750417 * "mean_ height"
Live surface fuels of herbs and shrubs (HerbShrb)	mean height	NA	NA	0.19	1.27440995 - 0.04569552 * "mean_ height"

The area-based approach used to model inventory parameters has some limitations and those were highlighted in the observations identified as outliers during the modeling efforts. Based on the variability of natural landscapes, we decided to retain all observations while developing our inventory models. To illustrate one such outlier, plot 4102 had one of the highest field-calculated values for total cubic foot volume, 9,623 cubic feet per acre. Table 5 shows that the lidar predictors for total cubic volume are mean height and density above mean height. If we explore where plot 4102 ranks in these two lidar plot parameters, we see that it is approximately average in the distribution of both. This disagreement between the datasets is a result of edge effects. In most cases, a 0.1 acre plot is big enough to mitigate the edge effects but three very large trees near the edge of this plot contributed to a substantial discrepancy between the FVS calculations and the lidar-based values as a significant portion of the canopies was excluded from the clipped lidar data (figure 4). This resulted in the model under-predicting as the edge effects reduced the density and height statistics. A potential fix for this plot would be to increase its size to 0.2 acre; however, the potential benefit would probably not justify the increased field costs.

In another example, plot 1303 had a total cubic volume of zero as calculated by FVS because there were no live trees with a DBH of 8" or greater. The plot did contain a lot of young aspen trees, which create a very dense canopy and elevate the lidar predictor metric "density above mean height" to above average, as illustrated in figure 5. Given this contradiction, the inventory model over-predicted this plot and weakened the agreement between the lidar and the field observations. These plots highlight limitations of the area-based modeling approach, but on the whole there is strong agreement between the lidar and plot values for approximately half of the FVS-calculated inventory attributes.

In addition to the linear models discussed above and highlighted in table 5, we also attempted to create a predictive model for the Vegetation Structural Stage (VSS) attribute recorded at each plot. VSS is a categorical variable and doesn't lend itself to a linear regression approach like the continuous variables. VSS is made up of six categories: 1-Grass and Forb, 2-Sapling, 3-Young Forest, 4-Mid-Aged Forest, 5-Mature Forest, and 6-Old Forest. Random Forests (Breiman 2001) was used to initially investigate the predictive power of lidar for modeling VSS. A Random Forests model was created to predict VSS using the 6 standard classes with an out-of-bag-error rate of 48.28 percent. The error matrix shown in table 6 provides some insight on where errors and confusion of classes were prevalent.

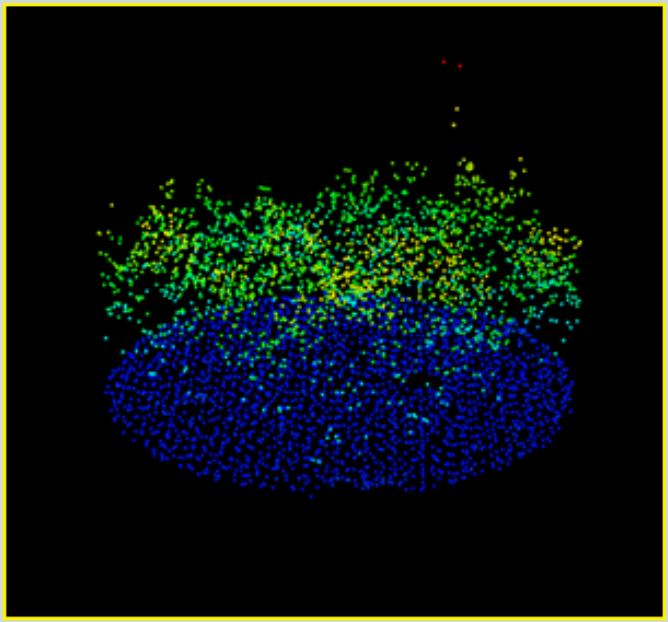
Point Cloud representation of plot 4102



Photograph of field plot 4102

Figure 4—The point cloud and photograph above illustrate the edge effects that degraded the relationship between field data calculations and lidar point cloud metrics for plot 4102.

Point Cloud representation of plot 1303



Photograph of field plot 1303

Figure 5—The point cloud and photograph above illustrate the characteristics that degraded the relationship between field data calculations and lidar point cloud metrics for plot 1303. The field-based calculations for total cubic volume excluded trees smaller than 8″ DBH, which excludes all of the trees on this plot.

As table 6 shows, most classes had a very high error rate, with class 3 being predicted wrong 100 percent of the time and class 5 being predicted wrong approximately 80 percent of the time. When we ran *Random Forests* using only plots in classes 4 and 6, the out of bag error rate was 20 percent, which indicates that lidar is able to differentiate between mid-aged forest and old forest classes. But when we included mature forest (class 5) in the classification, the error rate jumped greatly, indicating that lidar is not able to differentiate between mature forest and mid-aged or old forest. The Random Forests variable importance plots indicated that height variables had the most importance within the models (figure 6). It was beyond the scope of this project to explore VSS modeling any further, but we thought it was valuable to share initial results.

Table 6—Error matrix for the VSS Random Forests classification

	Predicted class 1	Predicted class 3	Predicted class 4	Predicted class 5	Predicted class 6	Class error
Actual class 1	7	1	1	2	5	0.56
Actual class 3	0	0	5	1	3	1.00
Actual class 4	3	0	19	3	4	0.34
Actual class 5	3	0	6	4	8	0.81
Actual class 6	1	0	6	4	30	0.27

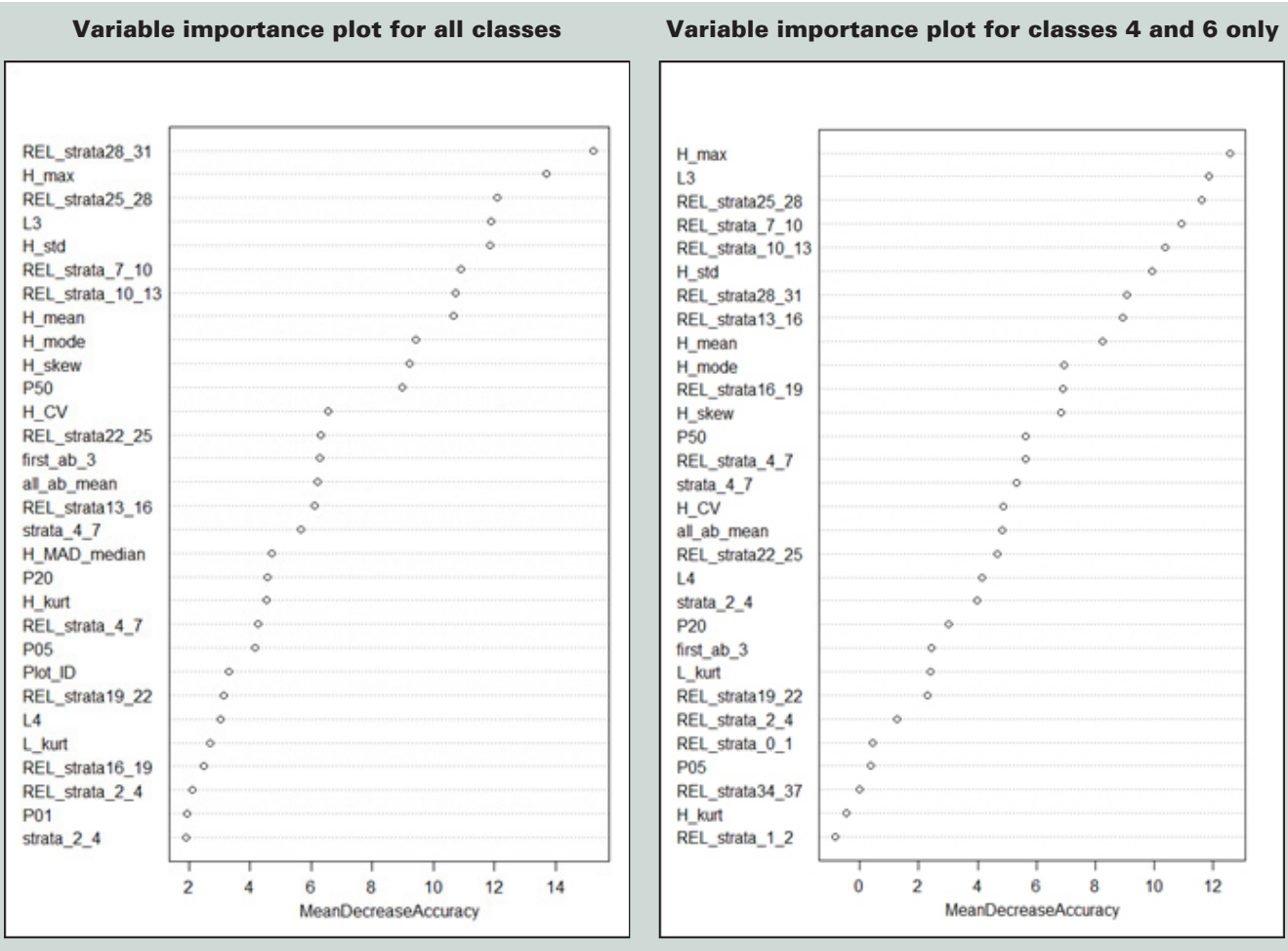


Figure 6—The variable importance plots from the Random Forests classification runs.

Applying Inventory Models at the Landscape Scale

One of the advantages of lidar is that the detailed canopy structure information used to create the forest inventory linear regression models at the plot level is also available at the landscape level. This allows us to apply the models listed in table 5 across our entire study area. The gridmetrics used to apply the models were produced at a 20 m cell size as this has approximately the same area as a 0.1 acre plot (this preserves the spatial scale of the models created with the area-based approach). Each equation in table 5 was applied to the landscape using the corresponding gridmetrics of the selected predictors, as specified by the plot-level models, to create forest inventory GIS layers. The forest inventory GIS layer for total cubic volume is depicted in figure 7. A complete collection of the forest inventory GIS layers is displayed in appendix B.

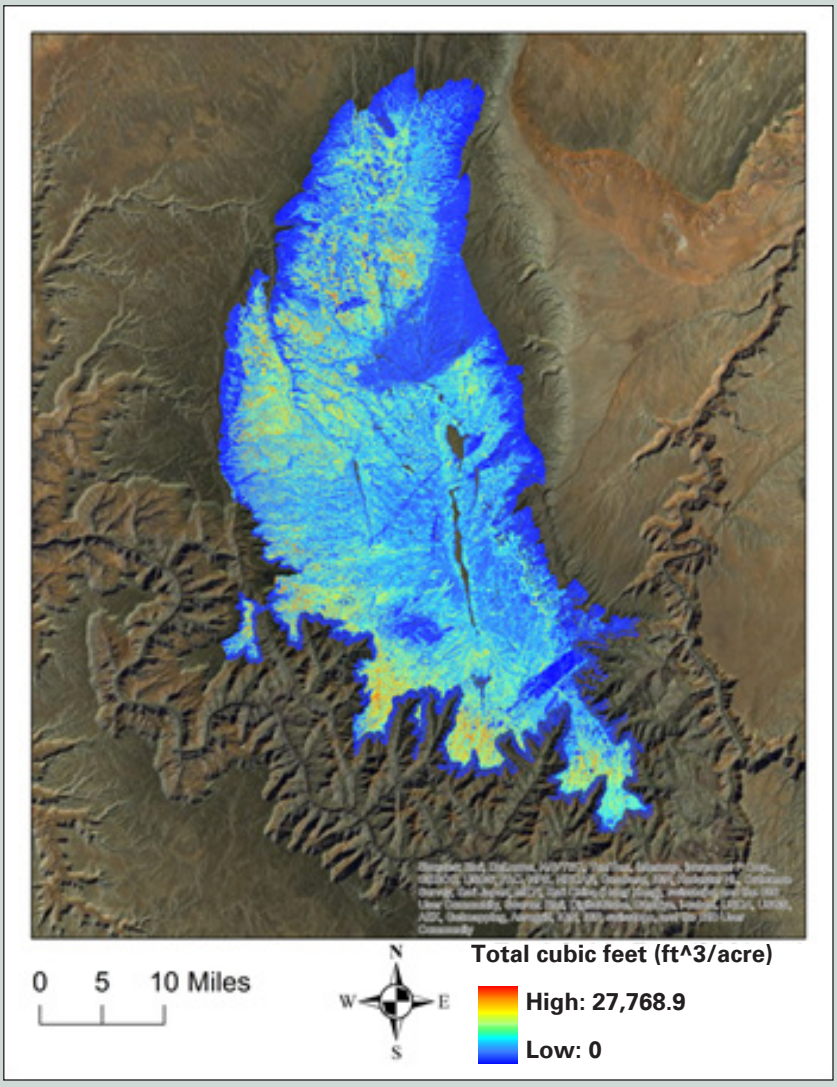


Figure 7—The GIS grid layer (20 m cell size) representing the total cubic volume forest inventory layer.

The extreme topography along the edge of the Kaibab Plateau (north rim of the Grand Canyon), produced some cliff edge artifacts that influenced the height statistics generated from the point clouds. The effect is illustrated with the 95th percentile height gridmetric in figure 8. These artifacts, if not dealt with, would propagate when the models were applied across the landscape so a cliff edge mask was applied to all final products. The mask was created using a GIS workflow that incorporated slope, curvature and lidar-derived canopy height. For a more detailed explanation and technical instructions for implementing the cliff edge masking workflow, refer to the "Masking Lidar Cliff Edge Artifacts" document (http://www.fs.fed.us/eng/rsac/lidar_training/pdf/Masking_Lidar_CliffEdgeArtifacts_06122014.pdf).

The field sampling strategy was designed to minimize model extrapolation but inevitably some extrapolation gets introduced. Model extrapolation creates values outside of the range of the training data for the inventory attribute being predicted. For example, if the maximum total cubic volume calculated from the field plot data is 10,404 ft^3/ac, but there are values exceeding that when the model is applied to the landscape, then extrapolations outside the data range used to create the model exist. The problem with extrapolation is a lack of training data to assess if the model is performing well. A summary of the extrapolated values produced when the inventory models were applied to the landscape is presented in appendix C. It is worth noting that an extremely low percentage of the pixels across the landscape were outside the training data range, with only two models extrapolating by 1 to 1.5 percent beyond the range of training data, and the others were extrapolating by less than 0.1 percent. Thus, we conclude that extrapolation is not a significant concern. To deal with the extrapolation issues the negative extrapolation values in the landscape models were changed to zero and the high-end extrapolation values were left unchanged.

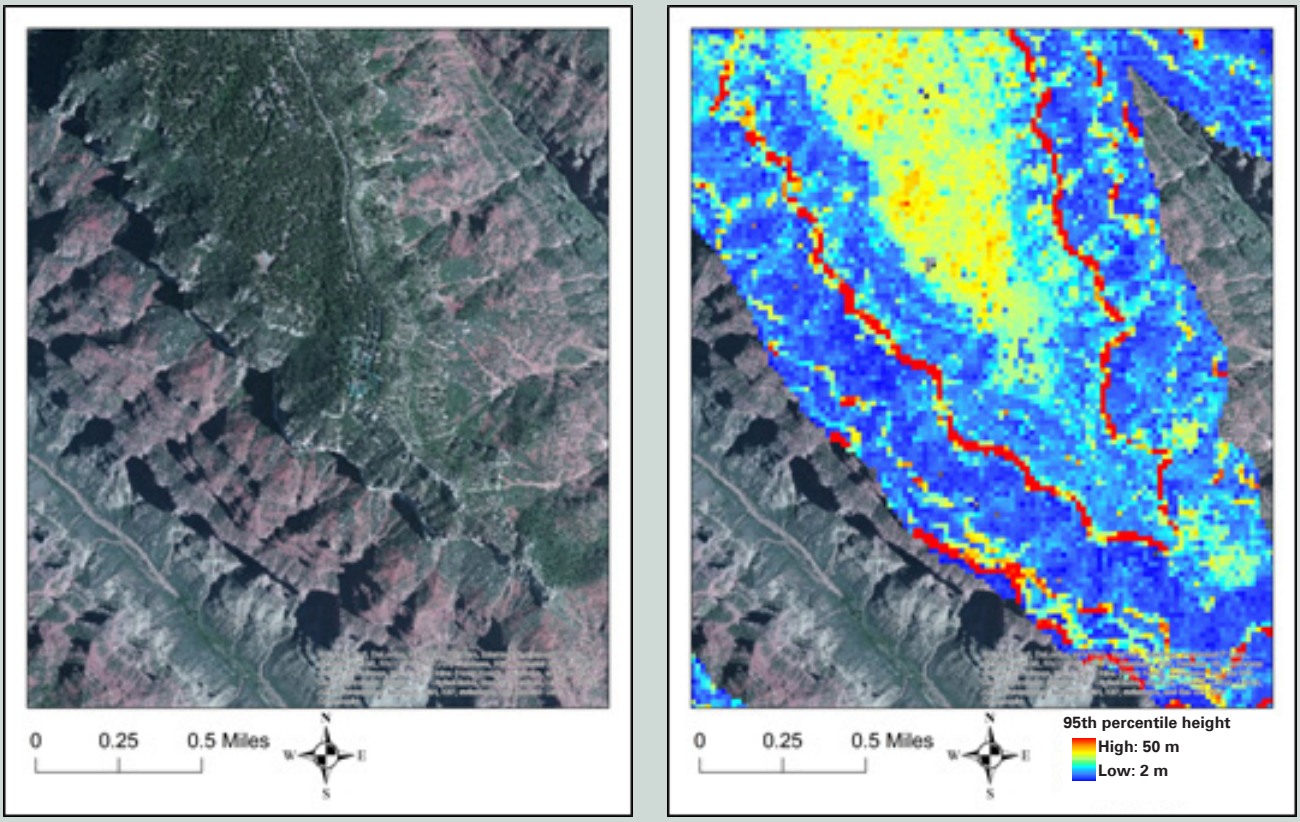


Figure 8—The NAIP imagery on the left highlights an area in the south of the Kaibab Plateau along the north rim of the Grand Canyon with a lot of cliff edges. The figure on the right is the 95th percentile height overlaid on the NAIP imagery and illustrates the linear cliff edge artifacts that are present in the height metrics.

Next Steps

The generation of lidar-derived 3-D canopy structure derivatives and lidar forest inventory models is the first phase in this effort and will provide the Rocky Mountain Research Station (RMRS) and Region 3 Regional Office (R3) with baseline data that can be used to explore and understand the links between 3-D canopy structure and goshawk demographic performance on the Kaibab Plateau. RMRS and R3 will use the baseline data products to assess their agreement with the demographic data. If they are in good agreement, predictive models can be created to provide the Forest Service with a dependable and repeatable information source to sustainably manage habitat for the goshawk on the Kaibab Plateau and similar forest types across the western U.S.

In addition to providing 3-D canopy structure information for assessing goshawk habitat, the lidar data will also provide new levels of efficiency and accuracy for a host of forest management objectives. In particular it will be helpful in the effort to restore the composition and structure in the southwestern forests as outlined by Reynolds and others (2013) in the RMRS General Technical Report, "Restoring composition and structure in Southwestern frequent-fire forests: a science-based framework for improving ecosystem resiliency."

Generally the lidar bare earth surface (high resolution DEM) and canopy derivatives will provide all resource managers on the Kaibab Plateau with a dataset that can improve efficiencies in daily management operations and decisions. RSAC is committed to providing technical consultation to resource managers on the Kaibab Plateau to ensure the lidar data are used in an appropriate and effective manner.

References

Breiman, L. 2001. Random Forests. Machine Learning. 45: 5–32.

Broughton R.K; Hinsley, S.A.; Bellamy, P.E.; Hill, R.A.; Rothery, P. 2006. Marsh tit *Poecile palustris* territories in a British broad-leaved wood. Ibis. 148: 744–752.

Goetz, S.; Steinberg, D.; Dubayah, R.; Blair, B. 2007. Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. Remote Sensing of Environment. 108: 254–263.

Hagar, J.C.; Eskelson, B.N.I.; Haggerty, P.K.; Nelson, S.K.; Vesely, D.G. 2014. Modeling marbled murrelet (*Brachyramphus marmoratus*) habitat using lidar-derived canopy data. Wildlife Society Bulletin. 38(2): 237–249.

Hawbaker, T.J.; Keuler, N.S.; Lesak, A.A.; Gobakken, T.; Contrucci, K.; Radeloff, V.C. 2009. Improved estimates of forest vegetation structure and biomass with a LiDAR-optimized sampling design. Journal of Geophysical Research. 114: G00E04. DOI: 10.1029/2008JG000870.

Hill, R.A.; Hinsley, S.A.; Gaveau, D.L.A.; Bellamy, P.E. 2004. Predicting habitat quality for great tits (*Parus major*) with airborne laser scanning data. International Journal of Remote Sensing. 25: 4851–4855.

Hinsley, S.A.; Hill, R.A.; Gaveau, D.L.A.; Bellamy, P.E. 2002. Quantifying woodland structure and habitat quality for birds using airborne laser scanning. Functional Ecology. 16: 851–857.

Hyde, P; Dubayah, R.; Peterson, B.; Blair, J.B.; Hofton, M.; Hunsaker, C.; Knox, R.; Walker, W. 2005. Mapping forest structure for wildlife habitat analysis using waveform lidar: validation of montane ecosystems. Remote Sensing of Environment. 96: 427–37.

Hyde, P.; Dubayah, R.; Walker, W.; Blair, J.B.; Hofton, M.; Hunsaker, C. 2006. Mapping forest structure for wildlife habitat analysis using multi-sensor (LiDAR, SAR/InSAR, ETM plus, Quickbird) synergy. Remote Sensing of Environment. 102: 63–73.

Lefsky, M.A.; Cohen, W.B.; Parker, G.G.; Harding, D.J. 2002. Lidar remote sensing for ecosystem studies. BioScience. 52: 19–30.

McGaughey, R. 2014. FUSION/LDV: software for lidar data analysis and visualization. Version 3.42. Seattle, WA: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. http://forsys.cfr.washington.edu/fusion/fusionlatest.html. (31 March 2015).

Nelson, R.; Keller, C.; Ratnaswamy, M. 2005. Locating and estimating the extent of Delmarva fox squirrel habitat using an airborne LiDAR profiler. Remote Sensing of Environment. 96: 292–301.

Reutebuch, S.E.; Andersen, H.; McGaughey, R.J. 2005. Light detection and ranging (LIDAR): an emerging tool for multiple resource inventory. Journal of Forestry. 103: 286–292.

Reynolds, R.T.; Graham, R.T.; Reiser, M.H.; Bassett, R.L.; Kennedy, P.L.; Boyce, D.A., Jr.; Goodwin, G.; Smith, R.; Fisher, E.L. 1992. Management recommendations for the northern goshawk in the southwestern United States. Gen. Tech. Rep. RM-217. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station. 90 p.

Reynolds, R.T.; Sánchez Meador, A.J.; Youtz, J.A.; Nicolet, T.; Matonis, M.S.; Jackson, P.L.; Delorenzo, D.G.; Andrew, D.G. 2013. Restoring composition and structure in southwestern frequent-fire forests: a science-based framework for improving ecosystem resiliency. Gen. Tech. Rep. RMRS-GTR-310. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 76 p.

Vierling, K.T.; Vierling, L.A.; Gould, W.A.; Martinuzzi, S.; Clawges, R.M. 2008. Lidar: shedding new light on habitat characterization and modeling. Frontiers in Ecology and the Environment. 6: 90–98.

Vogeler, J.C.; Hudak, A.T.; Vierling, L.A.; Vierling, K.T. 2013. Lidar-derived canopy architecture predicts brown creeper occupancy of two western coniferous forests. The Condor. 115(3):614-622.

For additional information, contact:

Haans Fisk, RSEAT Program Leader
Remote Sensing Evaluation,
Applications & Training
Remote Sensing Applications Center
2222 West 2300 South
Salt Lake City, UT 84119

phone: (801) 975-3800
email: mailroom_wo_rsac@fs.fed.us

This publication can be downloaded from the RSAC Web site: http://fsweb.rsac.fs.fed.us

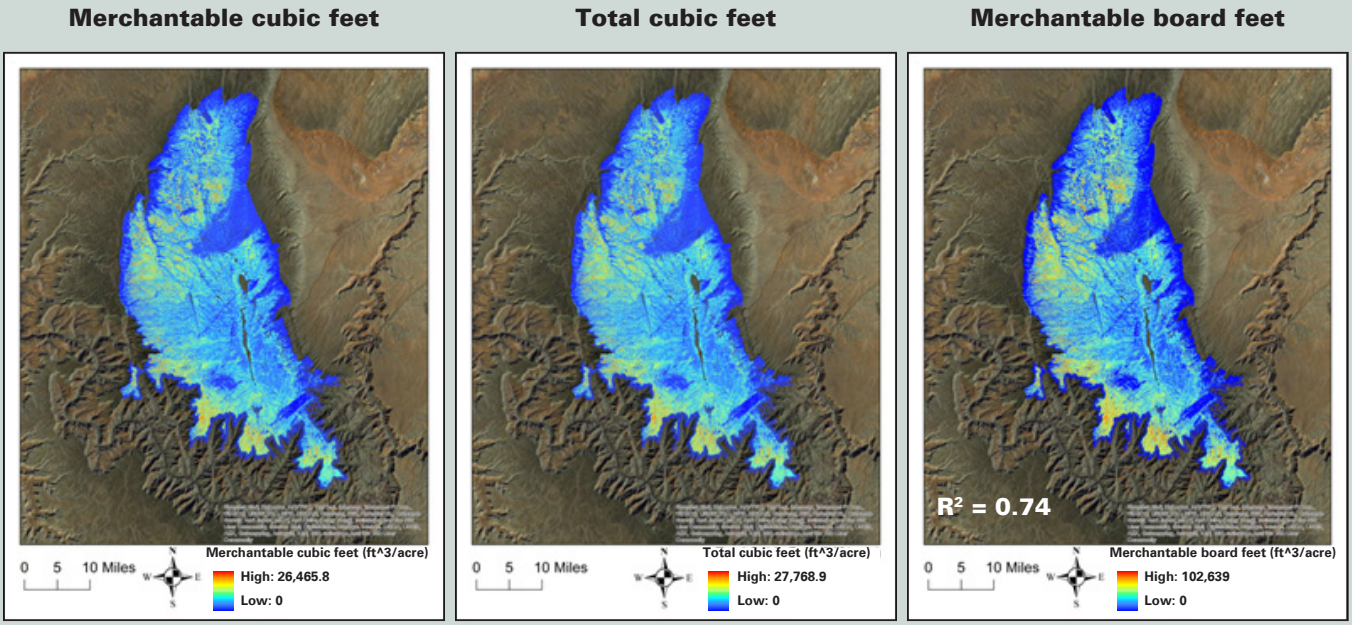
Appendix A: Field Plot Dominance Type

Table A-1—Number of field plots collected in each dominance type

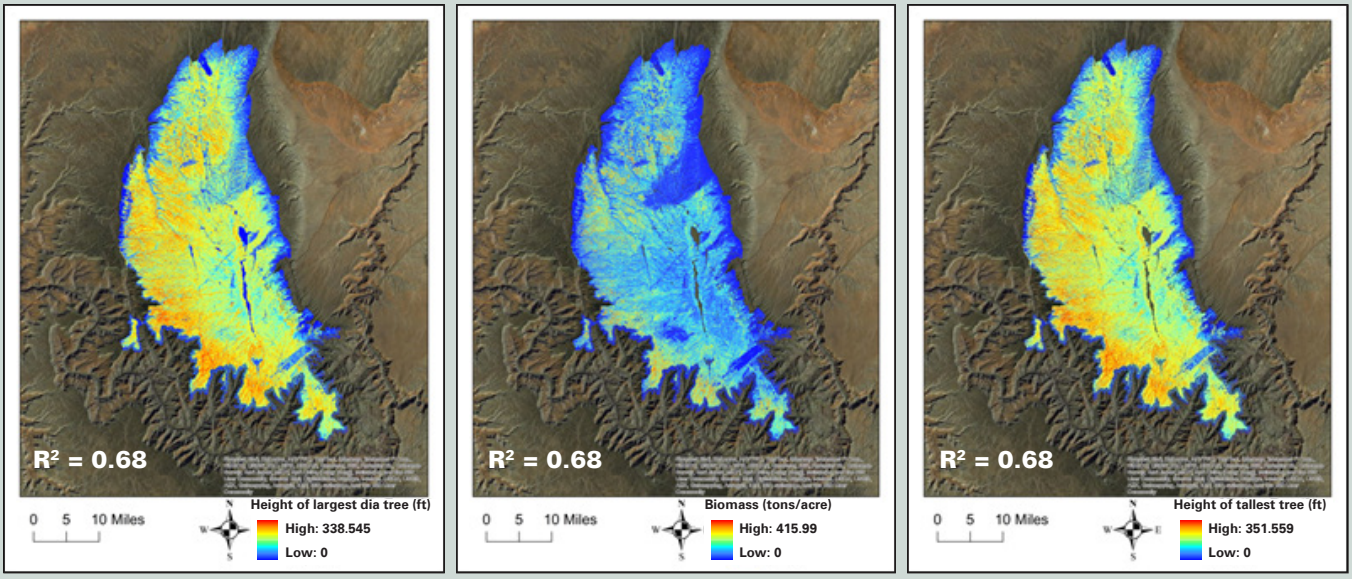
Dominance type	Number of field plots collected
Ponderosa pine	69
Quaking aspen	9
Engelmann spruce	7
Douglas-fir	4
Engelmann spruce—Quaking aspen	4
Ponderosa pine—Douglas-fir	3
Shade intolerant evergreen tree mix	3
Engelmann spruce—Ponderosa pine	2
Evergreen and deciduous tree mix	2
Gambel oak	2
Ponderosa pine—Quaking aspen	2
Quaking aspen—Douglas-fir	1
Deciduous shrub mix	1
Fir—Ponderosa pine	1
Pinyon—Ponderosa pine	1
Ponderosa pine—Engelmann spruce	1
Ponderosa pine—Shade intolerant evergreen tree mix	1
Two needle pinyon	1
White fir—Quaking Aspen	1
White fir—Singleleaf pinyon	1

Appendix B: Forest Inventory GIS Layers

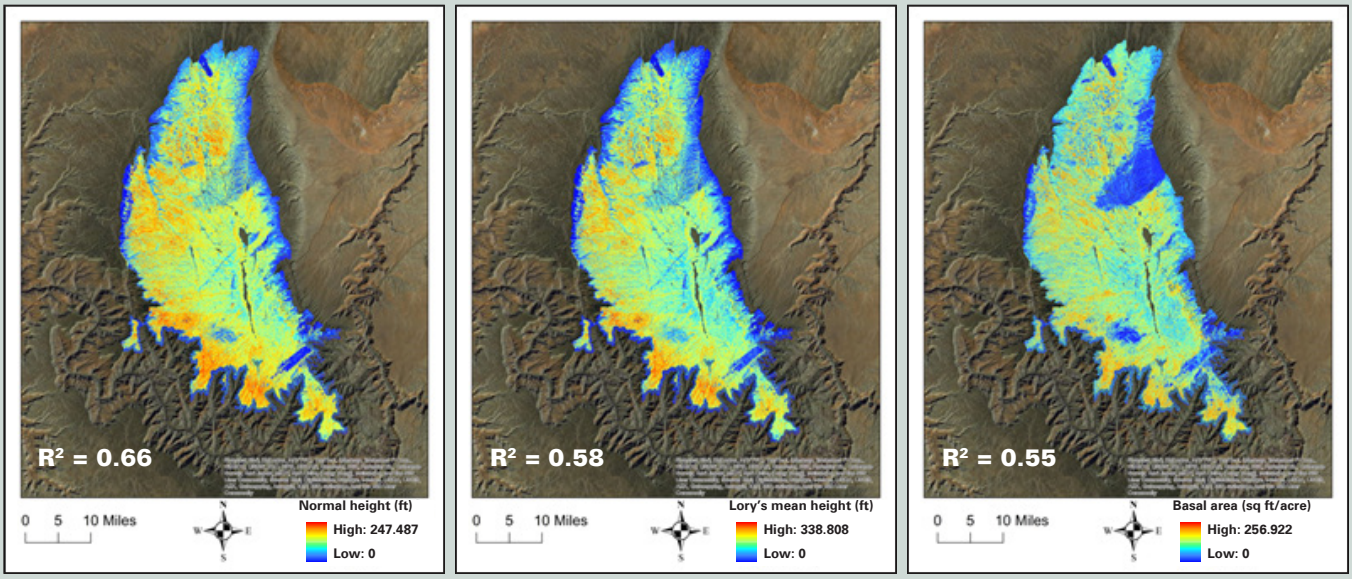
The GIS grid layers (20 m cell size) representing all inventory models presented in table A-1 applied at the landscape level.



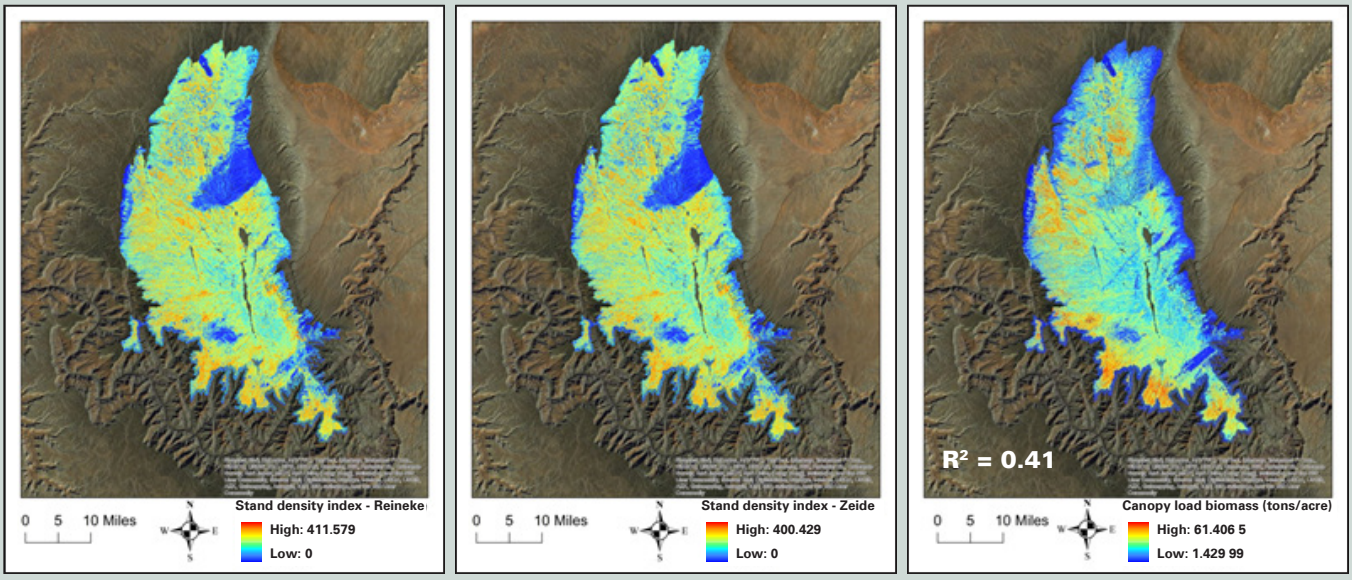
Appendix B-1—The GIS grid layers (20 m cell size) represent the forest inventory parameter models applied at the landscape level.



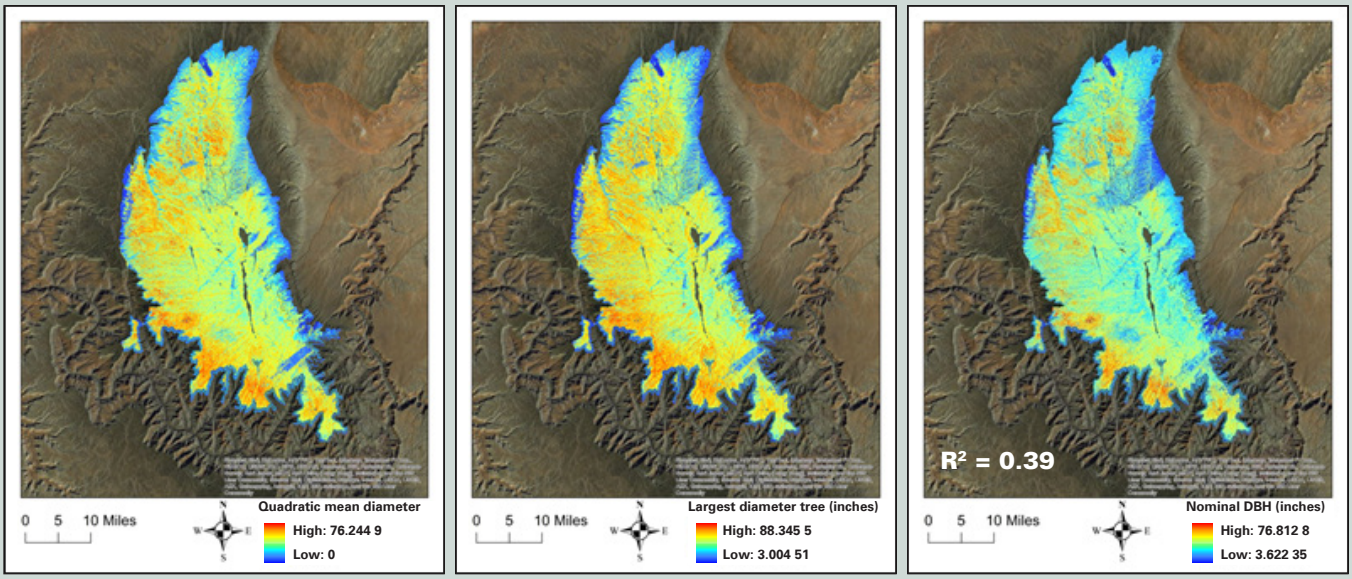
Appendix B-2—The GIS grid layers (20 m cell size) represent the forest inventory parameter models applied at the landscape level.



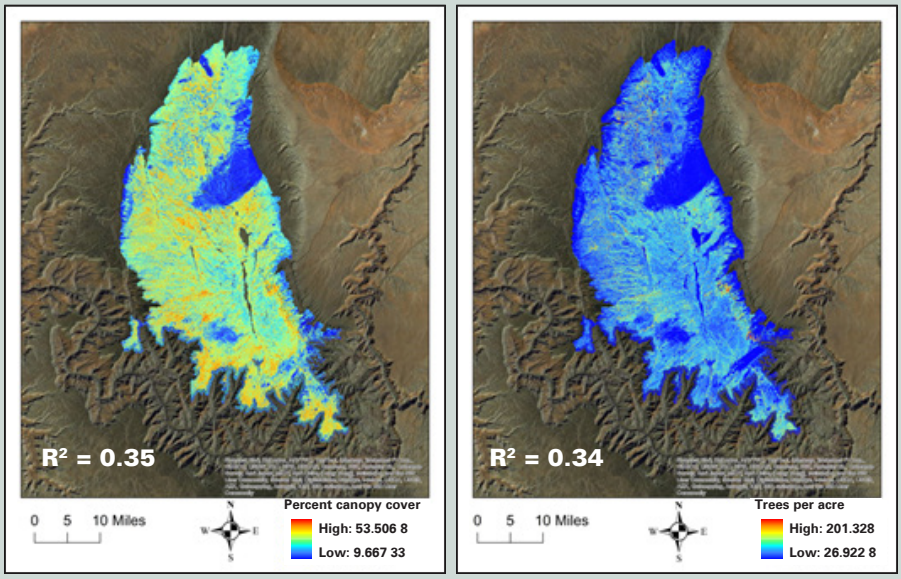
Appendix B-3—The GIS grid layers (20 m cell size) represent the forest inventory parameter models applied at the landscape level.



Appendix B-4—The GIS grid layers (20 m cell size) represent the forest inventory parameter models applied at the landscape level.

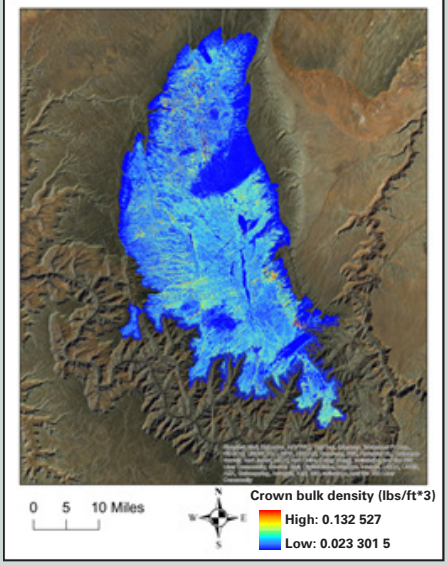


Appendix B-5—The GIS grid layers (20 m cell size) represent the forest inventory parameter models applied at the landscape level.

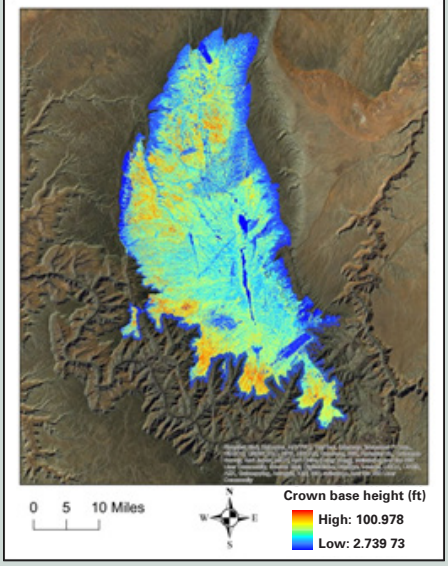


Appendix B-6—The GIS grid layers (20 m cell size) represent the forest inventory parameter models applied at the landscape level.

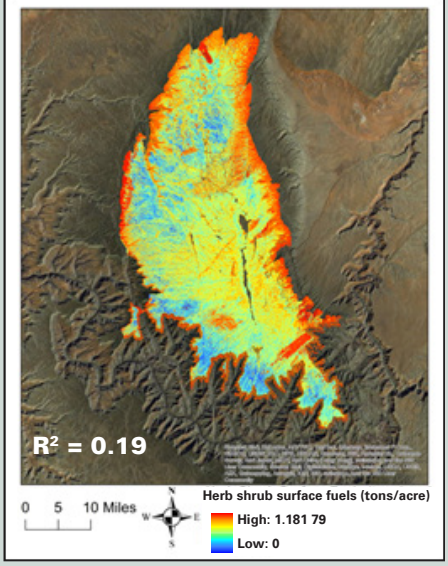
Crown bulk density



Crown base height



Herb shrub surface fuels



Appendix B-7—The GIS grid layers (20 m cell size) represent the forest inventory parameter models applied at the landscape level.

Appendix C: Extrapolation of Landscape Inventory GIS Layers

Table C-1 summarizes the effect of extrapolation when lidar inventory models were applied at the landscape level. Extrapolation values were identified after the cliff edge artifacts were masked out.

Table C-1—Summary of model extrapolation values in the landscape inventory GIS layers

Inventory parameter	Minimum plot value	Minimum predicted value	Pixels extrapolated below minimum plot value	Maximum plot value	Maximum predicted value	Pixels extrapolated above maximum plot value
MCuFt	0	-1,705.56	0.71%	9,836	29,821.40	0.11%
TCuFt	0	-1,449.05	0.34%	10,404	31,305.50	0.11%
MBdFt	0	-11,650.10	21.58%	64,100	125,779.00	0.001%
HgtLDia	0	-10.50	2.38%	121	338.55	1.04%
STBiomss	0	-1.40	0.0001%	175.81	464.08	0.05%
HgtTTree	0	-11.59	2.64%	122	351.57	1.51%
NHgt	0	-337.49	3.20%	117	252.37	0.06%
SBA	0	-65.21	0.01%	365.62	324.61	0%
SDI_R	0	-82.64	0.01%	386.08	486.67	0.001%
SDI_Z	0	-81.76	0.01%	378.85	473.34	0.001%
CLBiomss	0	1.42	0%	61.77	81.01	0.0001%
QMD	0	-0.46	0.03%	47.27	94.54	0.002
LDia	0	2.98	0%	54.6	88.35	0.05%
NDia	0	3.61	0%	46.67	100.06	0.002%
CanCov	0	9.67	0%	67.64	55.83	0%
TPA	0	26.92	0%	260	201.33	0%
CrBlkDn	0	0.02	0%	0.25	0.13	0%
CrBsHgt	-1	2.73	0%	82	133.61	0.0004%
HerbShrb	0.35	-2.87	4.94%	2.3	1.18	0%