Remote Sensing of Environment 124 (2012) 730–741

Contents lists available at SciVerse ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Prediction of understory vegetation cover with airborne lidar in an interior ponderosa pine forest

Brian M. Wing [a,*], Martin W. Ritchie [b], Kevin Boston [a], Warren B. Cohen [c], Alix Gitelman [d], Michael J. Olsen [e]

[a] Department of Forest Engineering, Resources, and Management, Oregon State University, 204 Peavy Hall, Corvallis, OR 97331, United States
[b] U.S.D.A. Forest Service, Pacific Southwest Research Station, 3644 Avtech Parkway, Redding CA 96002, United States
[c] U.S.D.A. Forest Service, Pacific Northwest Research Station, 3200 SW Jefferson Way, Corvallis OR 97731, United States
[d] Department of Statistics, Oregon State University, 44 Kidder Hall, Corvallis, OR 97331, United States
[e] School of Civil and Construction Engineering, Oregon State University, 220 Owen Hall, Corvallis, OR 97331, United States

ARTICLE INFO

Article history:
Received 5 February 2012
Received in revised form 26 June 2012
Accepted 27 June 2012
Available online 21 July 2012

Keywords:
Understory vegetation cover
Lidar
Intensity
Beta regression
Weighted regression

ABSTRACT

Forest understory communities are important components in forest ecosystems providing wildlife habitat and influencing nutrient cycling, fuel loadings, fire behavior and tree species composition over time. One of the most widely utilized understory component metrics is understory vegetation cover, often used as a measure of vegetation abundance. To date, understory vegetation cover estimation and prediction has proven to be inherently difficult using traditional explanatory variables such as: leaf area index, basal area, slope, and aspect. We introduce airborne lidar-derived metrics into the modeling framework for understory vegetation cover. A new airborne lidar metric, understory lidar cover density, created by filtering understory lidar points using intensity values increased traditional explanatory power from non-lidar understory vegetation cover estimation models (non-lidar R^2-values: 0.2–0.45 vs. lidar R^2-values: 0.7–0.8). Beta regression, a relatively new modeling technique for this type of data, was compared with a traditional weighted linear regression model using a leave-one-out cross-validation procedure. Both models provided similar understory vegetation cover accuracies (±22%) and biases (~0%) using 40.5 m^2 circular plots (n = 154). The method presented in this paper provides the ability to accurately obtain census understory vegetation cover information at fine spatial resolutions over a broad range of stand conditions for the interior ponderosa pine forest type. Additional model enhancement and the extension of the method into other forest types warrant further investigation.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Forest understory communities play many important roles in forest ecosystems (Suchar & Crookston, 2010). They provide habitat and forage for wildlife, are important factors in nutrient cycling and fire behavior, and help determine overstory species composition and structure over time (Falkowski et al., 2009; Legare et al., 2002; Scott & Reinhardt, 2001). Thus, understory communities are often considered good ecological indicators of forest health (Kerns & Ohmann, 2004; Tremblay & Larocque, 2001). To properly utilize understory components in the assessment of the above criteria, predictive models are needed for these characteristics (Suchar & Crookston, 2010). Unfortunately, most of the significant variables found to be useful for the prediction of the above criteria have been limited in explanatory power and spatial extent (Eskelson et al., 2011; Kerns & Ohmann, 2004; Russell et al., 2007; Suchar & Crookston, 2010; Venier & Pearce, 2007).

Understory vegetation cover, often used as an abundance measure, is an important metric used for wildlife habitat and fuel load characterization, fire behavior modeling, and understanding forest competition dynamics (Chen et al., 2008a). It is often laborious and costly to measure, which has resulted in it being sampled in a variety of ways (Eskelson et al., 2011). Traditional sampling methods include ocular estimation, line-intercept sampling, and fixed plot sampling (Bonham, 1989). All of these result in a percentage estimate for a unit area covered by understory vegetation.

Estimation and prediction of understory vegetation cover using field-derived explanatory variables has proven to be inherently difficult. To date, there have been two types of explanatory variables used in the estimation and prediction of understory vegetation cover; 1) topographically-derived (e.g. slope, aspect, digital terrain synthesis (DTS)), and 2) overstory-derived (e.g. basal area (BA), trees per hectare, leaf area index (LAI), canopy cover). The coefficient of determination associated with these models has been relatively poor (R^2-values ranging from 0.2 to 0.45) and their spatial extents are often limited to local study areas (e.g. Eskelson et al., 2011;

* Corresponding author. Tel.: +1 916 710 1054; fax: +1 541 737 4316.
E-mail addresses: brian.wing@oregonstate.edu (B.M. Wing), mritchie@fs.fed.us (M.W. Ritchie), kevin.boston@oregonstate.edu (K. Boston), warren.cohen@oregonstate.edu (W.B. Cohen), gitelman@stat.oregonstate.edu (A. Gitelman), michael.olsen@oregonstate.edu (M.J. Olsen).

0034-4257/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2012.06.024

Kerns & Ohmann, 2004; Russell et al., 2007; Suchar & Crookston, 2010; Venier & Pearce, 2007).

Traditional passive remote sensing techniques (e.g. optical imaging) have shown potential for providing information on forest characteristics, such as wildlife habitat, over broad areas at lower costs than traditional field inventories (Cohen & Goward, 2004; Kerr & Ostrovsky, 2003; Schroeder et al., 2007). In terms of estimation or prediction of understory vegetation cover, traditional remote sensing methods have been used to derive useful explanatory variables in cover modeling. Unfortunately, these methods are not sufficiently sensitive to 3D vegetation structure, which restricts their ability in the direct assessment of smaller areas or objects (Kerr & Ostrovsky, 2003; McDermid et al., 2005; Pesonen et al., 2008; Wulder & Franklin, 2003). They also have coarse spatial resolutions (>20 m), which can often constrain their usefulness.

A very promising fine spatial resolution remote-sensing technology for increasing the accuracy and efficiency of large-scale forest inventories is airborne discrete-return lidar (Maltamo et al., 2006; Næsset, 2002). Airborne lidar can be used to directly measure the three-dimensional structure of terrestrial and aquatic ecosystems across large spatial extents (Lefsky et al., 2002). Airborne lidar data produce three-dimensional characterizations of objects in the form of point clouds that are defined by precise x, y and z coordinates. They also help characterize the reflectance and surface properties of intersected objects by providing intensity values, which are a measure of return-signal strength, for each point. These attributes are useful for forest inventory and characterization, because in theory, every object in a forest with a vertical dimension can be detected if adequate lidar point densities are collected within all forest canopy layers (e.g. understory, overstory) (Pesonen et al., 2008).

In recent years, airborne lidar has been used successfully to estimate many standing tree characteristics such as biomass and volume (Heurich et al., 2004; Hyyppä et al., 2001; Maltamo et al., 2006; Næsset, 2002; Packalén & Maltamo, 2006), as well as canopy cover and height profiles (Coops et al., 2007; Goetz et al., 2007; Lim et al., 2003). Airborne lidar has also been incorporated into assessments of biodiversity (Clawges et al., 2008; Goodwin et al., 2007; Hill & Broughton, 2009; Maltamo et al., 2005; Zimble et al., 2003), fire behavior models (Andersen et al., 2005; Mutlu et al., 2008; Riaño et al., 2003), and wildlife habitat models (Goetz et al., 2007; Vierling et al., 2008). Estimation and prediction of understory components such as vegetation cover with airborne lidar has received less study. Martinuzzi et al. (2009) studied the presence and absence of understory shrub cover (cover>25%) on 20 m×20 m pixels using airborne discrete-return lidar. They found presence accuracies of 83% using two airborne lidar understory metrics along with a transformed slope aspect variable. Hill & Broughton (2009) examined the presence and absence of understory vegetation using two separate airborne discrete-return lidar datasets collected at the same location; one collected in leaf-on and one collected in leaf-off conditions. They found accuracies of 77% using a combination of both lidar datasets and 72% using only the leaf-off lidar on 20 m×20 m plots. In another recent study, Morsdorf et al. (2010) used airborne discrete-return lidar height and intensity information to identify individual vegetation strata on 5 m×5 m pixels in various forest conditions and had some success detecting the presence of the understory vegetation strata. Detection of coarse woody debris (CWD) with airborne lidar has also been studied with some promising results (Pesonen et al., 2008; Seielstad & Queen, 2003).

Intensity values are an often underexploited feature of lidar data, due to the difficulty and variability associated with acquisition settings and calibration. Intensity is the power of the returned laser light per unit area. It is primarily a measure of surface reflectance and is a function of the wavelength of the source energy, path distance, and the composition and orientation of the surface or object which the laser pulse intersects (Boyd & Hill, 2007). Currently, airborne lidar sensors use variable gain controls to compensate for variations in ground brightness and surface object reflectance to help ensure the sensor is adequately detecting returns. They affect the quality and the usefulness of intensity values. Variable gain settings can either be manually or automatically adjusted throughout an acquisition (automatic more prevalent), which can result in intensity values that lack calibration or normalization (often referred to as radiometric calibration) into the same reference scale. Gain settings are currently proprietary, thus they are unavailable to end users making radiometric calibration dependent on vendors (Boyd & Hill, 2007; Donoghue, Watt, Cox & Wilson, 2007; Kaasalainen et al., 2009). At the time of this study, the majority of lidar vendors do not calibrate the intensity information; thus they rely solely on variable gain and acquisition settings to provide useful intensity information. The quality of the intensity data is also dependent upon additional lidar acquisition parameters. Laser beam divergence, type of source energy, and path lengths all affect the quality of the intensity information and thus must be adjusted for different acquisition scenarios to ensure useful intensity information is obtained. These attributes have resulted in a broad range of quality and limited the use of intensity data. As vendor calibration and acquisition techniques become more robust and end user calibration becomes possible, the use of intensity information will likely increase.

Even with these difficulties, intensity information has been used successfully in many forestry applications to differentiate between tree species, estimate biomass, and predict basal area (Donoghue et al., 2007; Holmgren & Persson, 2004; Hudak et al., 2006; Kim et al., 2009; Lim et al., 2003; Morsdorf et al., 2010). Lim et al. (2003) used an intensity threshold to remove lower NIR intensity returns when estimating the live biomass of a northern hardwood forest in Ontario, Canada. In that study, the mean height of the higher intensity returns was the best predictor of basal area, biomass and volume. More recently, Kim et al. (2009) used intensity value threshold filtering to successfully estimate live and dead standing tree biomass. All of these studies point toward the great potential of intensity information to help characterize many forest attributes. In this study, we explore the ability of intensity information to filter lidar points associated with various understory components.

This study seeks to expand on previous work and exploit the additional information available in airborne lidar to predict understory vegetation cover. The primary objectives of this study are to: 1) analyze the potential of airborne lidar-derived metrics to estimate and predict understory vegetation cover; 2) explore the use of intensity values to filter understory component lidar points, 3) compare two modeling approaches for the prediction of understory vegetation cover using the airborne lidar-derived metrics; and 4) develop a practical method that utilizes airborne lidar-derived metrics to predict understory vegetation cover. New understory airborne lidar metrics are introduced and explored.

2. Materials and methods

2.1. Study area

The study was conducted at Blacks Mountain Experimental Forest (BMEF) in northeastern California (Fig. 1). The experimental forest (40°40′N, 10 121°10′W), managed by the USDA Forest Service Pacific Southwest Research Station, is located approximately 35 km northeast of Mount Lassen Volcanic National Park and ranges between 1700 and 2100 m elevations. Stands are dominated by ponderosa pine (*Pinus ponderosa* Dougl. ex P. and C. Laws) with some white fir (*Abies concolor* (Gord. and Glend.) Lindl.) and incense-cedar (*Calocedrus decurrens* (Torr.) Florin) at higher elevations. At lower elevations, Jeffrey pine (*Pinus jeffreyi* (Grev. and Balf.); Oliver, 2000) can also be found in some stands. Classified as an interior ponderosa pine forest type (Forest Cover Type 237) (Eyre, 1980), the 4358 ha forest has a wide range of stand conditions as a result of past research and

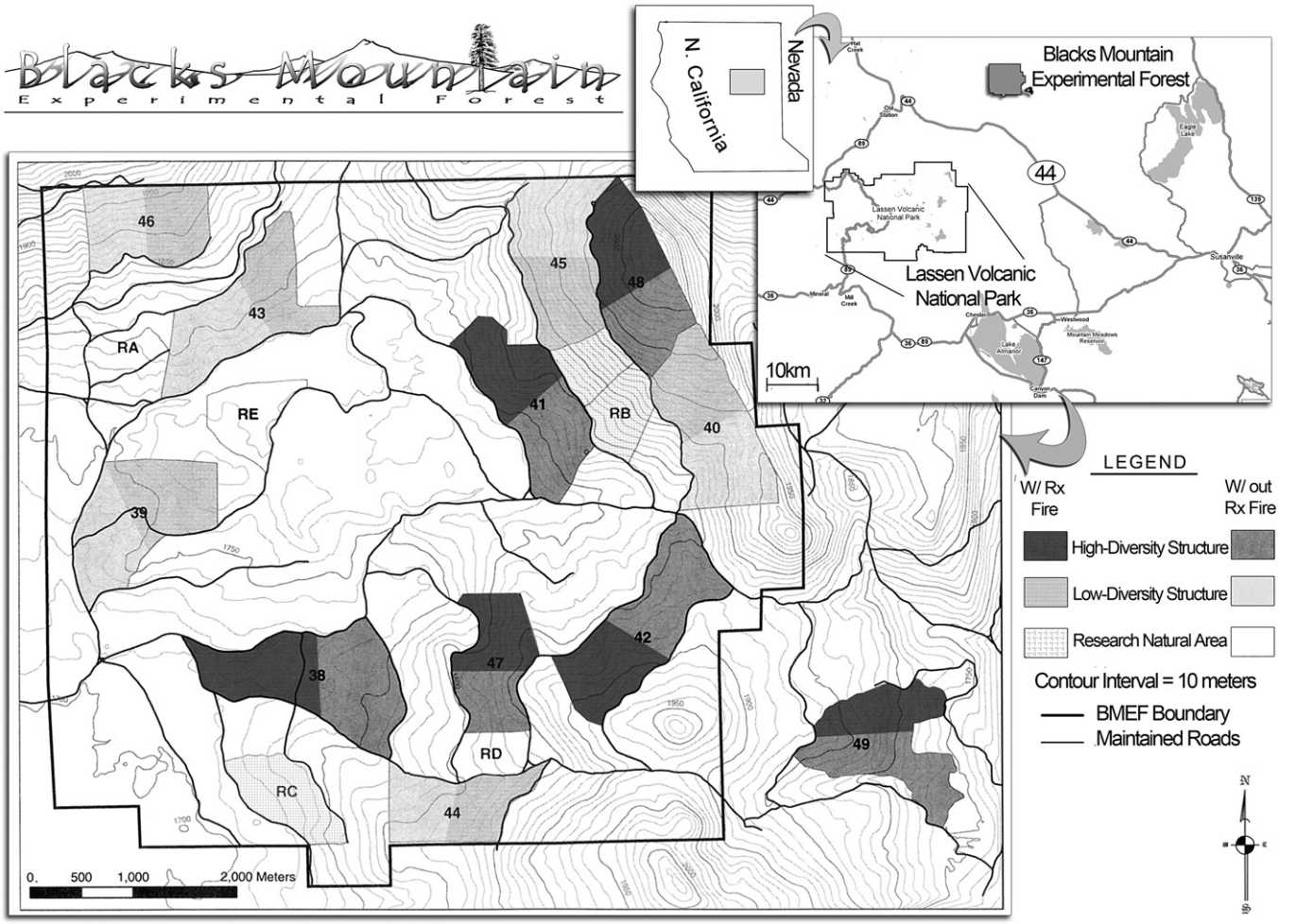

Fig. 1. Geographic location of the Blacks Mountain Experimental Forest and layout of the Blacks Mountain Long-Term Ecological Research Project in northeastern California.

management activities, as well as disturbance events (Ritchie et al., 2007).

As part of a large-scale, long-term interdisciplinary experimental design at BMEF initiated in 1991, two contrasting stand structures were created: low structural diversity (LoD) and high structural diversity (HiD) (Oliver, 2000). LoD stands were thinned to maintain a single canopy layer of intermediate trees, with the goal of simplifying forest tree structure. At the time of treatment implementation, stands were thinned to a uniformly spaced density of approximately 40 trees ha^{-1}, maintaining trees with heights ranging from 12 to 30 m and crown ratios generally greater than 50%. At the time of our study, LoD stand densities ranged from 25 to 430 trees ha^{-1} based on plot-level data (DBH > 9 cm). In contrast, the HiD units retained all canopy layers, which resulted in stands that feature multiple age classes and varying crown structures (Oliver, 2000). All large old trees were maintained with one smaller tree retained within the larger tree's crown circumference. Tree densities ranged from 60 to 95 trees ha^{-1} at the initial implementation and ranged from 90 to 1400 trees ha^{-1} at the time of our study based on plot-level data (DBH > 9 cm). Plots with higher tree densities are associated with a few spatially scattered dense thickets (0.4–0.8 ha) containing smaller trees that were left as part of the HiD prescription.

Six research units were each randomly assigned from both the LoD and HiD treatments ranging in size from 77 to 144 ha. Each unit was then split in half with one randomly assigned half receiving prescribed fire treatments (Fig. 1). Due to the large unit size, treatment implementation took several years. The three individual treatment blocks, each with four units, were created in 1996, 1997, and 1998, respectively.

Also included at BMEF, are four research natural areas (RNA) each approximately 40 ha in size (RA, RB, RC, RD). The RNAs were set aside to serve as unmanaged, qualitative controls representative of the interior ponderosa pine type. They have never received mechanical treatment, but fire exclusion has greatly increased their understory tree densities. Two of the four RNAs (RB and RC) received one application of prescribed fire in the late 1990s. RNA stand densities ranged from 420 to 1220 trees ha^{-1} for trees ≥9 cm DBH at the time of our study.

As part of the experimental design all 16 research units at BMEF have permanently monumented grid markers located within them on a 100 × 100 m lattice pattern. The permanent grid markers serve as the center points for all plot level research being conducted on the forest. Each grid was located by conventional survey methods and placed within 15 cm of their predetermined UTM coordinates using the High Precision Geodetic Network along with survey grade GPS (Oliver, 2000). These provide a solid foundation for researchers to conduct airborne lidar research, because plot location errors are minimized.

2.2. Field data

Field data were collected on five of the LoD units, six of the HiD units and 2 randomly selected RNAs in July 2009 (RC and RD). Standing live tree (DBH ≥ 9 cm) attributes for all three structure types at

the time of our study are summarized in Fig. 2. Using the BMEF permanent grid system, plot locations were assigned systematically with a random start within each unit on every other grid point in all intercardinal directions (282 m spacing). At each selected grid point location two nested circular plots were established: 1) a 40.5 m^2 circular plot to measure understory vegetation, and 2) a 805 m^2 circular plot to measure standing trees and coarse woody debris (CWD). A total of 154 plots were measured (LoD = 65, HiD = 79, RNA = 10). Every shrub with a height greater than 0.3 m was measured for crown dimensions and was stem mapped (Fig. 3). These measurements included the azimuth and distance from the plot center to the center of the shrub, two perpendicular crown width measurements, and two height measurements (maximum height and average height). Maximum height was defined as the top height of the shrub, and average height was determined by ocular estimation measured with a tape measure. Shrub species found in our study included (listed in order of abundance): greenleaf manzanita (*Arctostaphylos patula* Greene), antelope bitterbrush (*Purshia tridentate* (Pursh)), snowbrush (*Ceanothus veluntinus* Dougl. ex Hook.), wax current (*Ribes cereum*), Pacific serviceberry (*Amelanchier alnifolia* Nutt.), rabbitbrush (species) (*Chrysothamnus sp.*), common snowberry (*Symphoricarpos albus* (L.) S.F. Blake), and Sierra gooseberry (*Ribes rozelii* Regel). Greenleaf manzanita and snowbrush exhibit denser foliage with larger leaf areas, more branching complexity, and tend to grow taller and wider than the other shrub species.

A geographically registered shrub cover layer was then constructed in ArcGIS using shrub locations coupled with crown dimensions (Fig. 4). For each shrub the arithmetic mean of the two perpendicular crown widths was used as the shrub diameter. Next, a circle was assumed for the general two-dimensional shrub shape and each shrub's circular area was incorporated into the layer. Lastly, the circular shrub areas were merged to create one shrub cover layer for each plot. This technique accounts for overlapping shrub crowns, edge effects, and should result in a more accurate estimation of shrub cover when compared to many traditional sampling methods.

Seedlings over 0.3 m in height and all saplings were tallied for each plot. Saplings were tallied into two diameters at breast height (DBH) classes (2.54 and 5.08 cm). For seedling and sapling cover estimates, predetermined cover values were used (seedlings = 0.15 m^2; saplings (2.54 cm class) = 0.5 m^2; saplings (5.08 cm class) = 1 m^2). These cover values were based on average values from a subsample of seedling and sapling crown dimensions. Saplings greater than 6.35 cm in DBH were considered to have crowns above the understory layer based on field observations. Total understory vegetation cover was determined by summing all three of the cover areas together and dividing by the total plot area. This method does not account for overlapping seedling and sapling crowns which could slightly affect the accuracy of the plot measured understory vegetation cover values when seedlings and saplings were present.

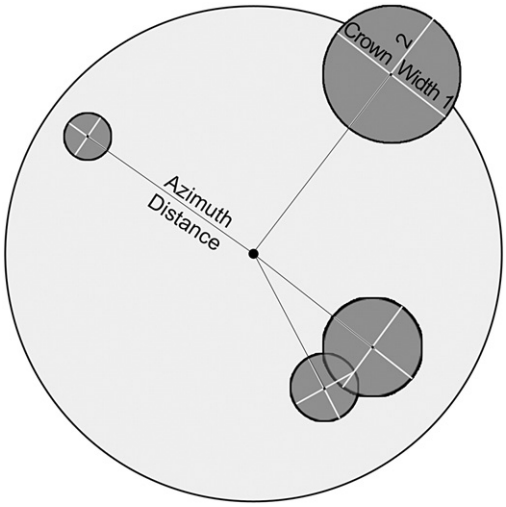

Fig. 3. Field sampling design for understory shrub cover.

In addition to the understory vegetation measured, all coarse woody debris with at least one end height above 0.3 m and one end diameter greater than 0.3 m were measured at every understory vegetation plot location only using a larger plot size (809 m^2 circular). Azimuth and distance was measured to the middle of each end from the plot center and each end's width and height were also measured for cover and volume estimation. The spatial characteristics of the data enable direct determination of the geographic spatial arrangement associated with each piece of CWD. These attributes provided the ability to determine the quantity, cover and volume of CWD located within the 40.5 m^2 circular understory vegetation plots. In addition, prominent stumps (height > 0.5 m) were located using the azimuth and distance from the plot center.

Standing live and dead trees ≥ 9 cm DBH were also measured on the 809 m^2 circular plots. All trees were stem mapped from the plot center and measured for total height, DBH, crown width, and height to live and dead crown. These data were used for verification of plot locations.

2.3. Lidar data

Discrete multiple return airborne lidar data were provided by Watershed Sciences Inc. in LAS file format (version 1.1). The lidar data were acquired over the entire BMEF study area in late July 2009 using a Leica ALS50 Phase II laser system mounted on a fixed wing aircraft. The aircraft was flown at 900 m above ground level following topography. Data were acquired using an opposing flight line side-lap of 50% and a sensor scan angle 14° from nadir to provide good penetration of laser shots through the canopy layers. On-ground laser beam

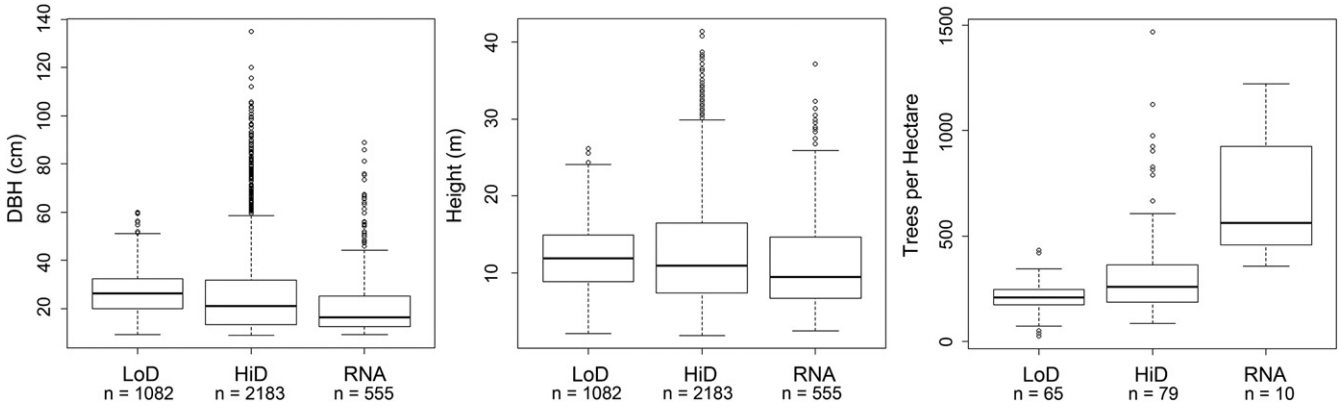

Fig. 2. Standing live tree (DBH > 9 cm) attributes from all plots per treatment type (LoD, HiD, RNA) at BMEF.

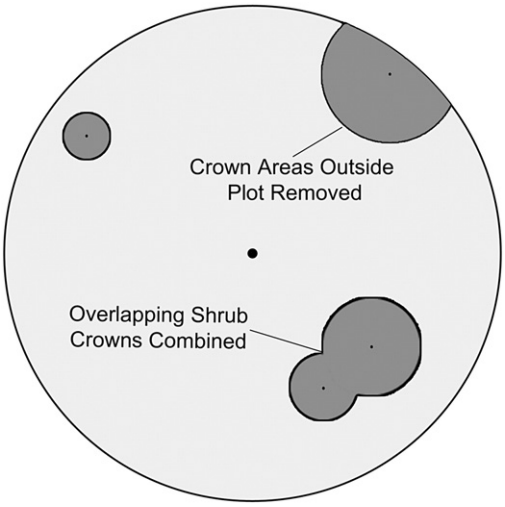

Fig. 4. ArcGIS understory vegetation cover layer created from field-measured shrub data.

diameter was approximately 25 cm (narrow beam divergence setting), which resulted in a low percentage of multiple returns (higher order than first returns: 9.2%) and a high percentage of first and single returns (first: 9.4%; single: 81.4%). The high ratio of first and single returns helped provide better quality intensity information, because calibration problems associated with laser pulse energy are reduced for these returns (for review: Morsdorf et al., 2010). An average of 6.9 points m^{-2} was obtained for the entire study area, with a standard deviation of 5.6 points m^{-2}. Ground survey data were collected to enable the geo-spatial correction of the aircraft positional coordinate data collected throughout the flight, and to allow for quality assurance checks on final LiDAR data products. Simultaneous with the airborne data collection mission, multiple static (1 Hz recording frequency) ground surveys were conducted over monuments with known coordinates to enable geo-spatial data correction. Indexed by time, these GPS data were used to correct the continuous onboard measurements of aircraft position. To enable the assessment of LiDAR data accuracy, ground truth points were collected using GPS based real-time kinematic (RTK) surveying.

The vendor post-processed lidar data utilized proprietary software (TerraScan) coupled with manual methods to identify ground points for the development of the digital terrain model (DTM). Vertical DTM accuracy for BMEF was approximately 15 cm at a 95% confidence level. The vendor used an automatic variable gain setting during acquisition and did not calibrate the intensity values post-acquisition. In past acquisitions, where the vendor used similar acquisition methods, the intensity information was successfully used to differentiate between live and dead biomass (Kim et al., 2009).

2.4. Data analysis

An important step in any airborne lidar data analysis for forestry applications is verification of geo-registered plot locations. Inaccurate plot locations can be one of the largest sources of model error found in many types of airborne lidar analysis. Even though the permanent grid system at BMEF helps to minimize the need for this step, every plot location was manually inspected using the standing tree stem maps for each 809 m^2 circular plot. Every 809 m^2 circular plot point cloud was compared to the field-measured standing tree stem map to assess the validity of the plot location. All plot locations were found to be highly accurate (0.2 m) based on the manual inspection.

Once plot locations were verified, the lidar point cloud heights were normalized using the DTM and points corresponding to the 40.5 m^2 circular plots were extracted from the normalized lidar dataset. These plot point clouds were used to derive all potential explanatory lidar metrics used in the understory vegetation cover modeling analysis.

2.4.1. Understory lidar metrics

Martinuzzi et al. (2009) found the use of two understory airborne lidar metrics along with a common slope-aspect transformation variable that could accurately estimate the presence of shrub cover (accuracy: 83%). The two understory lidar metrics utilized in their study were the percentage of ground points and percentage of points between 1 and 2.5 m compared to all plot points. We introduce a new understory lidar metric that combines the inherent information found in these two metrics.

The new metric, understory lidar cover density (ULCD), can be derived using a series of standardized steps that can be automated (Fig. 5). First, the height range for the understory layer is determined from the average and maximum shrub height data collected from field measurements. The minimum height for the understory layer was determined by rounding the field-measured minimum average shrub height value down to the nearest 0.1 m. The maximum height for the understory layer of each plot was determined by rounding the field-measured upper 99th percentile maximum shrub height value to the nearest 0.1 m. By using the 99th percentile value of the maximum height range the maximum height threshold for the understory layer was reduced by 0.4 m and was determined to better represent the overall shrub crown height distributions for the site. The maximum height for the understory layer also served as the cut-off level between understory and overstory points.

For this study, average shrub heights ranged from 0.25 to 1.45 m and maximum heights ranged from 0.5 to 1.85 m (Fig. 6). This resulted in an understory layer that ranged from 0.2 to 1.5 m. There were a total of 8 shrubs (shrub sample size: m = 821) measured that had portions of their crowns above the maximum understory layer height threshold. All lidar points located within the understory layer were then extracted from the plot point cloud for further analysis. The average plot-level percentage of the first and single returns in this layer was similar to that of the entire acquisition (first: 6.7%; single: 83.4%). Theoretically, points in this layer can intersect one of eight understory components for this forest type: shrubs, the base of standing tree boles, seedlings, saplings, CWD, taller herbaceous vegetation, low hanging tree branches, and stumps (Fig. 7). Other intersected components were considered too rare to be significant in this study area.

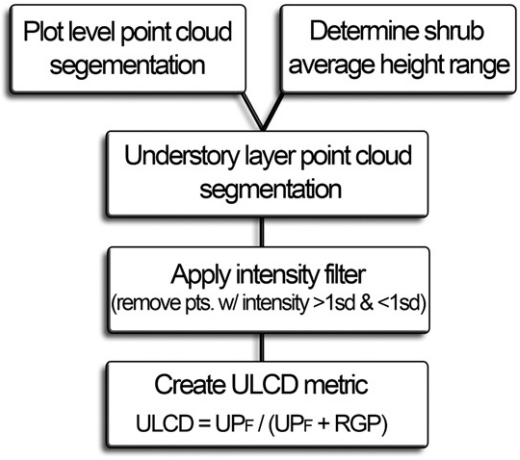

Fig. 5. Procedure for the creation of the understory lidar cover density (ULCD) metric. sd = standard deviation, UP_F = number of understory points remaining after filtering, RGP = relative ground points.

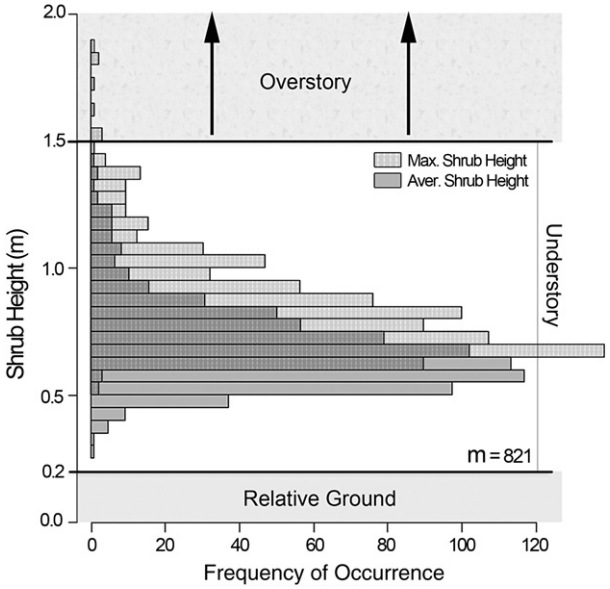

Fig. 6. Frequency histogram for average and maximum shrub heights and the depiction of how understory layer height range is determined. m = number of shrubs sampled.

We explored the use of intensity values to filter and remove points associated with unwanted understory components (e.g. non-vegetation and herbaceous points) from the understory layer point cloud. It was hypothesized that the intensity values would differ for the various understory components, thus providing a technique to identify the points associated with understory vegetation. For each plot, all points below the understory threshold (<1.5 m) were used to calculate understory intensity mean and standard deviation values. It was determined from manual inspection of understory lidar point clouds, that points associated with live vegetation typically contained intensity values within one standard deviation of the mean intensity value, and points associated with other understory components were often outside this range. Based on this observation, all points with intensity values beyond one standard deviation from the plot's mean intensity value were removed from the understory point cloud. The understory lidar cover density metric is then obtained using the formula:

$$\mathrm{ULCD} = \frac{\mathrm{UP_f}}{(\mathrm{UP_f} + \mathrm{RGP})} \quad (1)$$

where UP_f is the number of remaining understory points after applying the intensity filter, and RGP is the number of relative ground points (points under 0.2 m).

Two additional understory lidar metrics derived from the understory point cloud (heights<1.5 m) were understory point density (UPD) and effective plot coverage (EPC). Understory point density was defined as the number of lidar points per square meter under 1.5 m, and ranged from 1.5 to 24.2 points m^{-2} with a mean of 5.4 points m^{-2} and a standard deviation of 3.1 points m^{-2}. Typically, point densities are used to assess the adequacy of plot point cloud coverage. High point densities for a plot are often associated with adequate point coverage over the entire plot. In an understory context, it is possible to obtain high point densities while areas within the plot have no representative points because of scanning attributes (e.g. scan angle and path distance) and overstory obstructions. In an attempt to overcome this trait, the EPC metric was derived to measure the size of the plot area where the point coverage was adequate. This metric contains inherent information from overstory characteristics (e.g. canopy cover and structure, species composition, etc.) and acquisition methodologies (e.g. point densities, scan angles, pulse rate and pattern). Two assumptions must be made to derive the metric. First, how much area an individual point should represent, and second, what the significant minimum shrub cover area is (i.e. the minimum cover area associated with a shrub that would meet sampling requirements). A balance between these two assumptions must be determined. After initial exploration, it was assumed that an area of 0.09 m^2 was a good representative area in the determination of EPC, because it coincided with the smallest shrub crown area sampled. To derive the metric, plots were gridded into 0.3×0.3 m cells and each grid cell was evaluated to

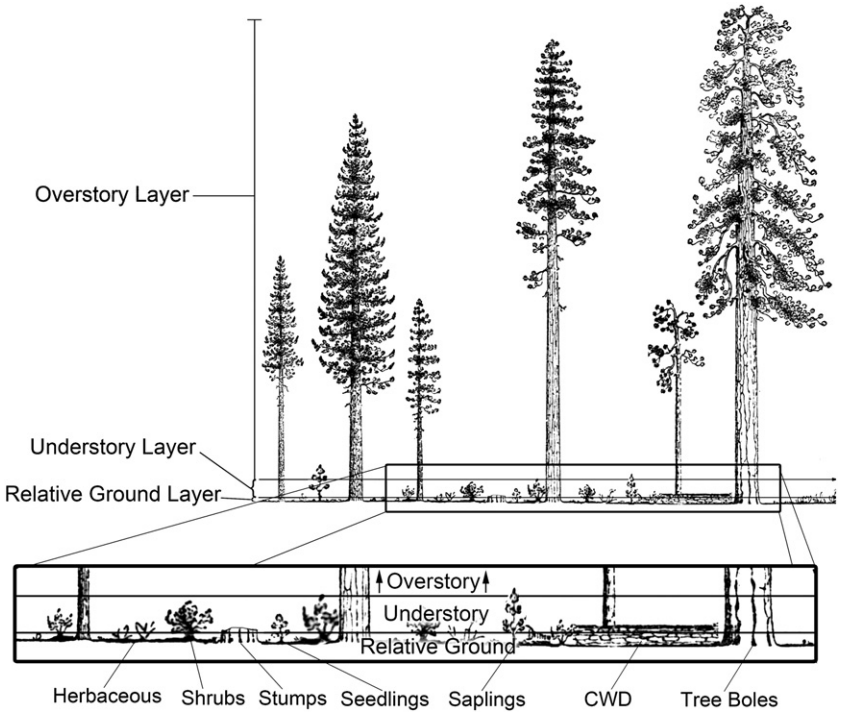

Fig. 7. Depiction of possible understory components that airborne lidar pulses can intersect. Artwork derived from Dunning (1928).

determine if it contained an understory point. Grid cells containing at least one point were summed up to determine the effective area covered by understory lidar points (Fig. 8). EPC was then determined by dividing the effective area covered by the total plot area. The effective plot coverage ranged from 0.12 to 0.88 with a mean of 0.38.

2.4.2. Overstory lidar metrics

Many previous understory vegetation cover studies found variables associated with the overstory (e.g. standing basal area, tree density, species composition) to be significant in the estimation of the understory vegetation cover (Eskelson et al., 2011; Kerns & Ohmann, 2004; Suchar & Crookston, 2010; Venier & Pearce, 2007). From previous lidar studies, the following overstory metrics were derived from the first, last and combined return overstory point clouds (heights > 1.5 m): 1) the quantiles corresponding to the 01, 10,..., 90 percentiles of the canopy heights; 2) the maximum height values; 3) the mean height values; 4) the standard deviation and coefficient of variation of height values; 5) the proportion of points above the 01, 10,..., 90 canopy height percentiles; 6) the proportion of points located within six predetermined canopy height intervals (s1 = 1.5–5 m, s2 = 5–10 m, s3 = 10–20 m, s4 = 20–30 m, s5 = 30–40 m, s6 > 40 m), and 7) overstory canopy cover determined by the proportion of first returns over the 1.5 m understory height threshold (Falkowski et al., 2009; Hudak et al., 2008; Næsset, 2002).

2.4.3. Topographic and stand attribute variables

Seven independent topographic and stand attribute variables were used in the model selection procedure. Topographic variables are often used for the estimation and prediction of understory cover (Eskelson et al., 2011; Martinuzzi et al., 2009). Five independent topographic variables were derived from the airborne-lidar-generated DTM for each plot: elevation, slope, aspect, and two commonly used slope aspect transformations [slope · cosine(aspect); slope · sin(aspect)] (Stage & Salas, 2007). Stand attribute variables included the research unit number and strata (LoD, HiD, RNA).

2.5. Estimation and prediction modeling

Although cover is frequently sampled in vegetation surveys, the theoretical and statistical basis underlying cover measures are not well understood (Chen et al., 2006). Understory vegetation cover data, including data used in this study, are characterized by two key distributional features that do not conform to the assumptions of standard statistical procedures (Damgaard, 2009). They are bounded between 0 and 1, and have heteroscedastic error variances. In ordinary least squares (OLS), parameter estimates are unbiased but are inefficient when heteroscadastic error variances are present; in addition the usual parameter estimate variance–covariance estimators are biased. There are a number of alternative adjustment methods to deal with the unequal error variance problem in the OLS linear regression setting. The two most common adjustment methods are applying independent and dependent variable transformations and the use of weighted regression (WR) (Kmenta, 1986).

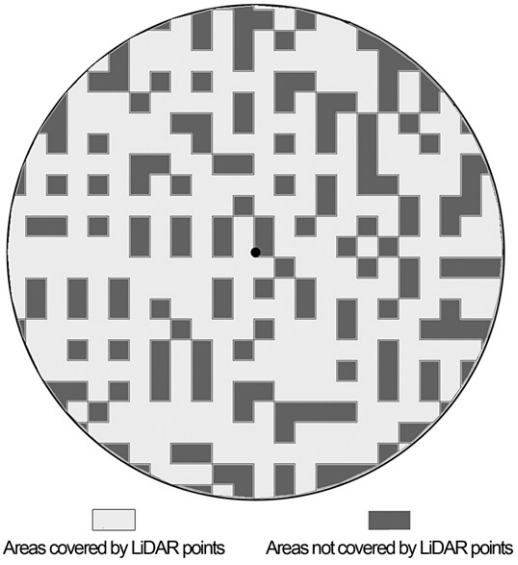

Fig. 8. Depiction of the ArcGIS layer created to derive the lidar effective plot coverage (EPC) metric using 0.3 × 0.3 m grid cells.

A theoretically correct way to model cover data is by using the properties of the beta distribution, a flexible and useful tool for modeling continuous random variables that assume values in the standard unit interval (0, 1), such as rates, percentages and proportions (Kieschnick & McCullough, 2003). Thus, it can be appropriate for modeling vegetation cover data because it adequately describes the frequency distribution of cover for various individual plant species or plant communities (Bonham, 1989; Chen et al., 2006; Damgaard, 2009; Pielou, 1977). While most of the work with the beta distribution has been completed for grasslands and crop fields (Chen et al., 2006, 2008a, 2008b), it has recently been applied in forestry applications. For example, Eskelson et al. (2011) used beta regression (BR) in the estimation of riparian understory vegetation cover and found that it performed better than the OLS model. Korhonen et al. (2007) also successfully estimated forest canopy cover with beta regression.

Based on the characteristics of the study's understory vegetation cover data (i.e. heteroscedastic error variance), weighted and beta regression models were specified for the estimation and prediction of understory vegetation cover using the airborne lidar-derived metrics. All three treatment strata (LoD, HiD, RNA) were grouped together to test the models robustness to varying forest structure and canopy densities.

2.5.1. Model specification

Weighted least squares regression can be used when the unequal error variance assumption of the linear regression model is violated. The theory behind this method is based on the assumption that the weights are known exactly (Kmenta, 1986). This is rarely the case, so estimated weights must be used instead. For this study, the equally sized group iterative procedure described in Kmenta (1986) to stabilize and determine final model parameter estimates was followed (5 iterations). Five groups of sizes approximately 31 were used in the procedure. Fitting this model is equivalent to minimizing:

$$Q = \sum_{i=1}^{n} \omega_i \left[y_i - f\left(x_i : \hat{\beta}\right) \right]^2 \tag{2}$$

where ω_i are weights $= 1/var(\varepsilon_i)^2$ for $(i = 1, \ldots, 154)$ from the 5 weighted groups, y_i is a vector of dependent variables, and $f\left(x_i : \hat{\beta}\right)$ is from the OLS linear model: $Y = X\beta + \varepsilon$.

Using a parameterization of the beta distribution, Ferrari and Cribari-Neto (2004) introduced a beta regression model similar to the approach for generalized linear models (McCullagh & Nelder, 1989), except that the distribution of the response is not a member of the exponential family. In the extended generalized linear model approach, $y_1, \ldots, y_n$ are independent random variables with each y_i is a parameterization of the beta probability density function with mean μ_i and variance ϕ. The beta regression model is specified:

$$g(\mu_i) = x_i^T \beta = \eta_i \tag{3}$$

where x_i is a vector of explanatory variables, β is a vector of unknown regression parameters, η_i is a linear predictor, $g(\cdot)$ is a

strictly increasing and twice differentiable link function that maps $(0, 1)$ into the real line, and T indicates the transpose of the vector. A variety of link functions $g(\cdot)$ are available, but the logit link $g(\mu) = \log(\mu/(1-\mu))$ is particularly useful, because μ_i is obtainable in closed form (Espinheira et al., 2008).

2.5.2. Model selection

Weighted and beta multiple regression models were fit for estimation and prediction of understory vegetation cover. The models were fit to all strata grouped together into one dataset to test the robustness of the models to varying forest structures and canopy densities. Traditional logarithmic transformations were applied to independent overstory lidar metrics which have been shown to be useful in previous studies (Næsset, 2002).

Selection of significant independent variables was completed via a two-stage procedure. First, a forward and backward stepwise model selection procedure was performed using OLS linear regression to reduce the field of explanatory variables to twenty based on the Bayesian information criterion (BIC) model performance. Twenty was used as the cut-off level to help ensure significant variables ($P<0.05$) would not be eliminated in this step. In the second step, BR models were fit using different sets and combinations of the twenty explanatory variables to identify the most significant variables based on BIC model performance. Because understory shrub cover data included zero values, the following commonly used transformation was applied to the understory vegetation cover dependent variable (Smithson & Verkuilen, 2006):

$$y_i^* = (y_i(n-1) + 0.5)/n \tag{4}$$

where y_i is field-measured estimate of understory vegetation cover and n is the number of sample plots ($n = 154$). Independent predictor variables with associated P-values greater than 0.05 were removed after this step. The final models were selected based on the lowest BIC value while also taking variable interactions into account. Variable interactions were assessed using a standard principal component analysis procedure (Weisberg, 1985). Two models were selected for further analysis with both WR and BR; one containing only the most significant variable based on the lowest partial F-statistic value and one containing all significant variables.

Both models were fit using the *R project* (v. 2.14.0) using the '*stats*' package (v. 2.13.2) for weighted regression (R Development Core Team, 2009), and the '*betareg*' package (v. 2.4.0) for beta regression (Cribari-Neto & Zeileis, 2010). Model comparisons were also conducted using *R*.

2.5.3. Model comparison

No independent data were available to assess the accuracy of the regression equations used for prediction. Therefore, leave-one-out cross-validation was used to assess the prediction accuracy of the models. For each step in the validation procedure, one sample plot was removed from the dataset at a time and the selected models were fitted to the remaining plots ($n-1$). Understory vegetation cover was then predicted for the removed plot. This procedure was repeated until predicted values were obtained for all plots. Two reliability figures were used to determine the accuracy of predictions. The absolute bias (AB), and root mean squared prediction error (RMSPE) were reported:

$$\mathrm{AB} = \sum_{i=1}^{n} \frac{(\mathrm{predicted} - \mathrm{observed})}{n} \tag{5}$$

$$\mathrm{RMSPE} = \sqrt{\sum_{i=1}^{n} \frac{(\mathrm{predicted} - \mathrm{observed})^2}{n}} \tag{6}$$

3. Results

The final selected model contained three variables: 1) ULCD 2) the standard deviation of overstory lidar first return point heights (OH_{sdf}); and 3) the density of overstory lidar first return points in the predetermined fifth height strata (OD_{s5f}) (Table 1). The signs of the coefficients correspond to the responses between understory vegetation cover and the independent variables. The regression parameters β have different interpretations in the BR model compared to the WR model, while model predictions have identical interpretation.

ULCD explained the greatest amount of variability for understory vegetation cover followed by the standard deviation of overstory first return point heights (OH_{sdf}) and then the density of overstory first return points in the predetermined fifth height strata (OD_{s5f}). The WR model containing only ULCD had a BIC value of (-466.2), while inclusion of the two significant overstory estimators decreased the value to (-488.6). For the BR model the BIC value went from (-530.0) to (-543.3) with the inclusion of the two overstory estimator variables. BIC values for the two model families (WR and BR) can only be used to compare within model performance. According to Raferty (1995) and Kass and Raferty (1995), a difference in BIC values (ΔBIC) of ≤ 2 between models is "not worth more than a bare mention" and a $\Delta\mathrm{BIC} > 10$ implies very strong evidence that the models are different.

Prediction accuracy was very similar for both the WR and the BR models. Overall, RMSPE was 0.003 larger for BR2 compared to WR2, which equates to an average understory vegetation cover prediction difference of approximately 0.3% (Table 2). Absolute bias was virtually zero for both models with the BR models displaying a slightly lower AB (BR2 -0.0001 vs. WR2 0.0005). RMSPE increased slightly for the models containing only the ULCD variable. Both models performed well in the prediction of understory vegetation cover with root mean square prediction errors ranging from 0.0640 to 0.0735, which translates to average understory vegetation cover prediction errors of approximately $\pm 7\%$. The overall accuracy for prediction of understory vegetation cover was $\pm 22\%$ for all model forms. AB was not significantly different from zero for any of the model forms. No trends were found between understory vegetation cover prediction errors and canopy cover for any of the models (Fig. 9). A small trend, which should be viewed with caution, was found between understory vegetation cover prediction errors and understory point densities. The errors seemed to decrease with increasing understory point densities, although as point densities increased the sample size diminished (Fig. 10).

Residuals for WR models were normally distributed and centered on zero with no obvious dependencies or patterns that might reveal improper model specification besides the unequal error variance issue in the linear model, which was dealt with by using the WR procedure. BR residuals displayed similar traits, except the residual errors displayed equal variance across all values. The BR residual distribution also displayed a slightly more pronounced negative tail. Larger residual errors from both the WR and BR models were most often associated with plots that contained CWD.

4. Discussion and conclusions

Understory vegetation cover has been difficult to estimate and predict, especially over large spatial extents. The method presented in this paper greatly increases the ability to estimate and predict

Table 1
Final statistical model summaries.

Model	Dependent variable	Intercept	Independent variables			BIC
			ULCD	OH_{sdf}	OD_{s5f}	
WR1	Understory vegetation cover	0.0048 (0.004)	1.0537 (0.066)			−466
WR2	Understory vegetation cover	0.0251 (0.006)	1.0260 (0.063)	−0.0064 (0.001)	0.1559 (0.031)	488
BR1	Understory vegetation cover	−3.0847 (0.109)	7.2843 (0.562)			−530
BR2	Understory vegetation cover	−2.5740 (0.141)	6.4823 (0.537)	−0.1719 (0.033)	2.6719 (0.868)	−543

*Independent variable standard errors are given in parenthesis. WR = Weighted Regression; BR = beta regression; BIC = Bayesian information criterion. Regression parameters and BIC values have different interpretations for BR and WR.

understory vegetation cover in interior ponderosa pine forests. Both the WR and BR models produced satisfactory errors for the prediction of understory vegetation cover. Only a simple independent variable transformation was necessary for the beta regression modeling framework, which should not result in any prediction bias. Theoretically the BR model seems to be the most appropriate choice; however the WR model performed equally well. This is most likely due to the most significant variable (ULCD) being a proportion bounded between 0 and 1, which essentially measures the same metric (e.g. the proportion of an area covered by shrub crowns). In theory, there should be a one-to-one type of relationship between these two variables. To demonstrate this point a simple linear regression model is presented in Fig. 11 between ULCD and field-measured understory vegetation cover.

The method was robust in terms of applicability to different forest structures in this forest type based on the model performance combining all three BMEF treatment strata (LoD, HiD, RNA). Understory vegetation cover prediction errors did not show any obvious relationships with canopy cover in this forest type (Fig. 9). This fact seems somewhat counterintuitive, since areas with higher overstory canopy densities typically occlude laser pulses from reaching the understory. Previous airborne lidar studies have identified this occlusion problem as a significant limiting factor in characterizing understory components (Hill & Broughton, 2009; Morsdorf et al., 2010). The problem was less evident in this forest type and likely resulted from a combination of unique characteristics associated with this study. First, the most significant variable, ULCD, is relative to the number of points that reach the understory. A proportion bounded by 0 and 1 itself, the ratio between the number of understory cover points to the total number of points below the understory maximum height threshold remains relatively stable under different overstory conditions. Even though laser pulses are less likely to intersect understory vegetation in denser canopy conditions, they are also less likely to reach the ground. Overstory cover ranged from 0 to 90% with a mean of 32% (standard deviation: 24%), based on the plot-level lidar data. The second key characteristic is that the likelihood of encountering understory vegetation decreases with increasing overstory cover for this forest type. Therefore, it is less important that a lower number of laser pulses are reaching the understory in these situations, because there is less probability of the area containing understory vegetation. The third key characteristic is associated with the relationship between the overstory and understory layers in this forest type. There tends to be a distinct height difference between overstory and understory layers in this forest type, which makes it easier to identify and analyze the understory vegetation cover layer separately.

Although it is yet untested, we hypothesize that this method will not perform as well in forest types that contain an abundance of understory vegetation under dense overstory cover conditions, or where the understory and overstory layers intermix. Obtaining higher understory point densities would help to alleviate these problems, but this is difficult in areas with dense overstory conditions and would also increase acquisition costs. The use of small-footprint full-waveform airborne lidar (SFFW) might also provide more explanatory value in identifying and ultimately predicting understory characteristics, such as cover. SFFW provides more than just coordinate and intensity information, it also provides echo width information. Echo width information has proven to be useful for classifying ground and vegetation returns and warrants investigation (Ducic et al., 2006; Wagner et al., 2008). Costs associated with SFFW acquisition are relatively high, making its use prohibitive for most broad-scale forest inventory applications. These costs will likely decrease over the next decade. In forests with significant proportions of deciduous trees, lidar acquisition completed during overstory leaf-off conditions and understory leaf-on conditions, if and when available, should increase understory lidar point densities and provide better results. In areas with taller herbaceous vegetation, acquisition should be completed while the herbaceous cover has yet to reach the minimum height requirement for the understory layer.

While both models performed well, additional modeling strategies should be investigated for this type of data to help determine the best approach. Two potential models not investigated in this study are the zero-inflated beta regression model (Ospina & Ferrari, 2012), and the

Table 2
Leave-one-out cross-validation results for the prediction of understory vegetation cover using individual models.

Model	RMSPE	AB
WR1	0.0678	0.0004 (P = 0.957)
WR2	0.0640	0.0005 (P = 0.946)
BR1	0.0735	0.0000 (P = 0.997)
BR2	0.0671	−0.0001 (P = 0.988)

*P-value for *t* test; testing whether the bias is significantly different from 0. RMSPE = root mean square prediction error; AB = absolute bias.

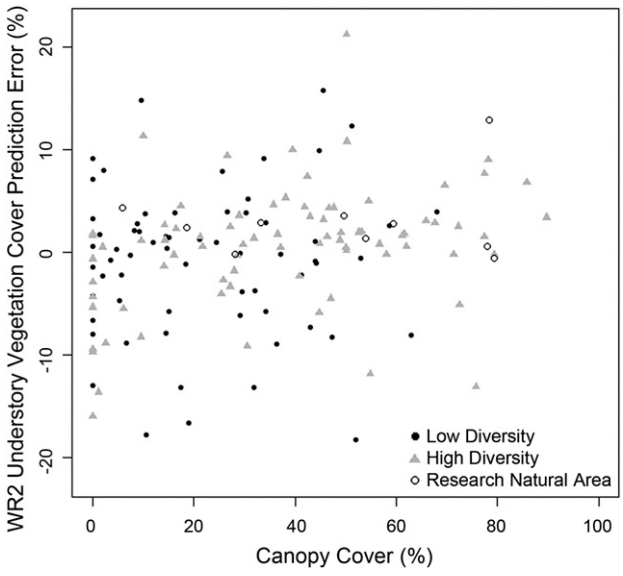

Fig. 9. Percent canopy cover (determined by the proportion of first returns over 1.5 m in height) versus WR2 understory vegetation cover model prediction errors for the 40.5 m^2 circular plots ($n = 154$).

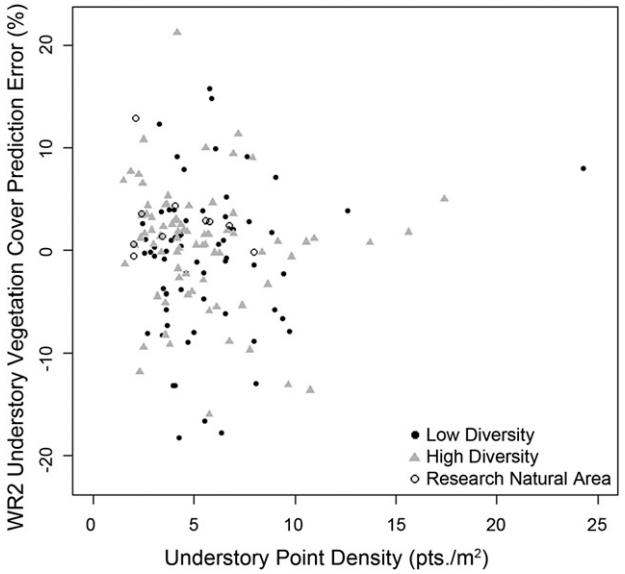

Fig. 10. Understory point density versus WR2 understory vegetation cover model prediction errors for the 40.5 m^2 circular plots ($n = 154$).

Copula model (Nelsen, 2006). The zero-inflated beta regression model is an extension of BR that incorporates the probability of observing a zero value in the model. Copulas are multivariate distribution functions whose one-dimensional margins are uniform on the interval (0, 1) (Nelsen, 2006). Eskelson et al. (2011) had promising results applying a multivariate Gaussian copula model to understory vegetation cover data which also accounted for spatial dependence. Both these models seem well suited for understory vegetation cover data since zero values are likely, but we leave this for future work.

The two overstory explanatory variables selected in the final model (OH_{sdf}, OD_{s5f}) indicate that there are significant interactions between overstory and understory vegetation. This trend coincides with previous understory vegetation studies (Eskelson et al., 2011; Martinuzzi et al., 2009). It is also interesting to note that Hopkinson et al. (2006) found the standard deviation of the vertical point structure to be the most powerful predictor of canopy height for various forest structures, and concluded that it should be used as a universal lidar canopy height metric. This study found the same metric (OH_{sdf}) to be the most significant overstory estimator variable, thus supporting their hypothesis. The metric seemed to replace a combination of canopy height distribution quantile metrics. When the metric was removed from the model and the model was refit with the remaining estimator variables, two canopy height distribution quantile metrics took its place (one representing the lower portion and one representing the higher portion of the crown profile). This demonstrates the usefulness of OH_{sdf} in areas containing variable forest structure and its ability to replace multiple lidar derived height metrics.

Plots containing CWD produced the largest residual errors. To demonstrate the importance of CWD, plot-level CWD cover was added to the plot-level understory vegetation cover dependent variable and the simple linear model (Fig. 11) was refit (Fig. 12). The explanatory power (R^2) increased from 0.74 to 0.81. This suggests that the filtering method was not successful in filtering out all points associated with CWD. To further solidify this point, the residual errors from the understory vegetation cover models were found to be the most significant estimators in a CWD presence and absence logistic regression estimation model (40.5 m^2 plot-level CWD volume > 1.5 m^3). Theoretically, the understory vegetation cover model residual errors should predominately be associated with the CWD lidar points, since other understory component points were successfully filtered based on visual inspection of the point cloud data. Residual errors coupled with two other significant independent variables, slope · sin(aspect) and the proportion of discarded intensity filtered points, produced presence accuracies of approximately 70% for estimating cumulative CWD volumes greater than 1.5 m^3 on the 40.5 m^2 circular plots. While the intensity filter successfully removed a portion of the CWD points from the ULCD variable, the understory vegetation cover model residual errors displayed more explanatory power than the proportion of discarded intensity filtered points in the CWD model. This suggests that there was a higher proportion of unfiltered CWD points using the filtering method in this study. If improved lidar point filtering techniques can be created, it might become possible to predict CWD in addition to understory vegetation cover using a similar method to the one outlined in this paper. A linear object recognition filter (Vosselman et al., 2004) coupled with intensity filtering might be successfully utilized.

Even though the intensity filtering method used in the study did not successfully remove all non-vegetation understory component points, it was successful at removing a large portion of them. The filter still might be improved in a number of ways. The first resides in the

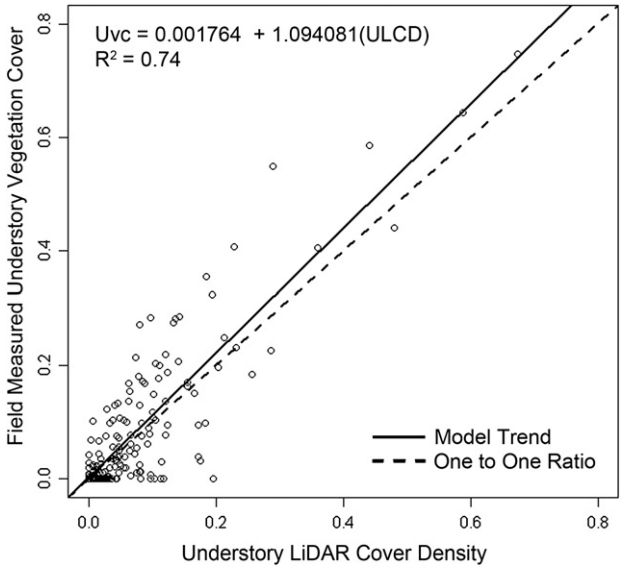

Fig. 11. Field-measured understory vegetation cover versus understory lidar cover density and the reference one to one ratio line.

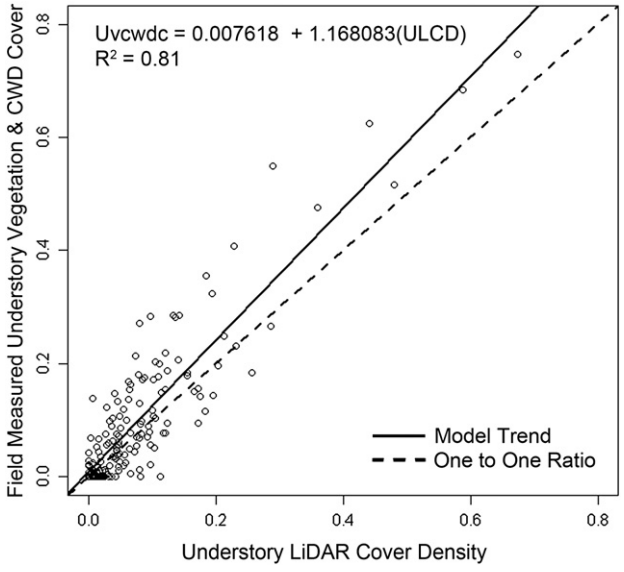

Fig. 12. Field-measured understory vegetation and coarse woody debris cover combined versus understory lidar cover density and the reference one to one ratio line.

fact that the filtering method used all lidar points associated with each plot's understory and relative ground layers (points<1.5 m height). The filter might perform better if only understory vegetation layer points are used. This would require a larger plot size or higher understory point densities to ensure enough understory component points are available for the creation of the plot-level intensity filtering statistics (e.g. mean and standard deviation). The use of only first and single understory returns could also provide a way of improving the filtering method, since intensity information associated with these returns has been shown to be of better quality, especially when no post-acquisition radiometric calibration is applied. This theory was tested using the data from this study and the results were neither improved nor diminished, which was likely due to the high percentage of first and single understory returns (90.1%). It was also found that the understory layer intensity data displayed both bi-modal and uni-modal distributions before filtering. A filtering method that treated them separately might provide another way of improving the intensity filter. It is also important to note that the lidar data in this study was acquired during understory leaf-on conditions, which likely made differentiating intensity values associated with understory vegetation and other understory components easier.

The use of intensity information to characterize forest attributes is dependent upon the quality of the intensity information. The intensity data in this study were acquired while using the variable gain setting and were not calibrated post-acquisition, if they had been, filtration results would likely improve. Even without calibration, intensity displayed great potential in distinguishing lidar points associated with the various understory components. Currently, airborne lidar vendors are just beginning to develop and apply post-acquisition intensity calibration techniques. As vendors continue to develop acquisition and calibration techniques and provide end users with the necessary information to calibrate the data, intensity values will likely become a much more valuable feature of airborne lidar data.

The field sampling design and data analysis steps used to obtain field-measured understory vegetation cover are simple to implement, provide accurate estimates, and fit well into most traditional forest inventory sampling designs. The shrub stem map and crown dimensional measurements provided a good method to estimate shrub cover accurately by accounting for overlapping crowns and incorporated well with the traditional airborne lidar plot-based standing tree inventory sampling design. The method can be improved in the future by stem mapping and dimensionally measuring seedlings and saplings. Then seedlings and saplings can be incorporated into the field based spatial understory shrub cover layer for more accurate field estimation of understory vegetation cover. Matching the actual shrub, seedling, and sapling crown shapes could also provide more accurate estimations. Although a circle seems like an appropriate assumption for shrub, seedling, and sapling cover shapes, the two perpendicular crown width measurements could be used to better match the actual crown shape of individual shrubs spatially. This might result in more accurate understory vegetation cover estimates. Plot size, sampling efficiency and costs associated with this sampling design should also be examined in further detail. Understanding the effects of increasing plot size on model variability would help to determine the most efficient sampling design.

Application of the prediction model to entire forested stands can be completed following the traditional airborne lidar two-stage plot-based gridding procedure outlined in Næsset (2002). In this procedure, stands of interest are first divided into grid cells that match the prediction model's plot size. Then significant independent variable values are obtained for each grid cell and the prediction model is applied using weights for each grid cell to minimize edge bias associated with the smaller boundary-edge grid cells. The end result can be used to; 1) identify areas that meet understory vegetation cover habitat criteria and create habitat maps over entire forest stands, 2) determine understory fuel loadings over entire stands, which can then be used to refine fire behavior models, 3) accurately estimate and predict understory vegetation biomass and carbon stocks, 4) help assess forest health and biodiversity, and 5) assess competition dynamics between understory vegetation and standing trees.

The characteristics of this study's sampling design, airborne lidar acquisition and intensity value calibration provided a unique opportunity to examine the capability of airborne lidar to predict understory vegetation cover. The method presented in this paper was practical and efficient, and showed promise for predicting understory vegetation cover at fine spatial resolutions over large spatial extents in the interior ponderosa pine forest type. Incorporating airborne lidar with other remote sensing techniques such as aerial photography, or utilizing small-footprint full-waveform airborne lidar could also enhance the ability to characterize and predict understory components such as vegetation cover. The new ULCD metric displayed a strong relationship with understory vegetation cover and was robust to various forest structures and densities in this forest type. The filtering of lidar points using intensity information helped to remove a portion of understory component points not associated with understory vegetation cover (e.g. CWD, stumps, live tree boles). The extension of the method to additional forest types warrants further investigation.

Acknowledgements

This study was funded through a cooperative agreement between Oregon State University and the U.S. Forest Service Pacific Southwest Research Station. The authors gratefully acknowledge the field work and additional help of Thomas Fisher, David McClung and Travis Springer. The authors would also like to acknowledge the three anonymous reviewers for providing extremely useful comments on earlier drafts of the manuscript.

References

Andersen, H. -E., McGaughey, R. J., & Reutebuch, S. E. (2005). Estimating forest canopy fuel parameters using lidar data. *Remote Sensing of Environment*, *94*, 441–449.
Bonham, C. D. (1989). *Measurements for terrestrial vegetation.* New York: John Wiley & Sons 338 pp.
Boyd, D. S., & Hill, R. A. (2007). Validation of airborne lidar intensity values from a forested landscape using Hymap data: Preliminary analysis. *IAPRS, XXXVI, Part 3/W52*.
Chen, J., Shiyomi, M., Yamamura, Y., & Hori, Y. (2006). Distribution model and spatial variation of cover in grassland vegetation. *Grassland Science*, *52*, 167–173.
Chen, J., Shiyomi, M., Hori, Y., & Yamamura, Y. (2008a). Frequency distribution models for spatial patterns of vegetation abundance. *Ecological Modelling*, *211*, 403–410.
Chen, J., Shiyomi, M., Bonham, C. D., Yasuda, T., Hori, Y., & Yamamura, Y. (2008b). Plant cover estimation based on the beta distribution in grassland vegetation. *Ecological Research*, *23*, 813–819.
Clawges, R., Vierling, K., Vierling, L., & Rowell, E. (2008). The use of airborne lidar to assess avian species diversity, density, and occurrence in a pine/aspen forest. *Remote Sensing of Environment*, *112*, 2064–2073.
Cohen, W. B., & Goward, S. N. (2004). Landsat's role in ecological applications of remote sensing. *Bioscience*, *54*(6), 535–545.
Coops, N. C., Hilker, T., Wulder, W. A., St-Onge, B., Newnham, G., Siggins, A., et al. (2007). Estimating canopy structure of Douglas-fir forest stands from discrete-return lidar. *Trees-Structure and Function*, *21*(3), 295–310.
Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, *34*(2), 1–24.
Damgaard, C. (2009). On the distribution of plant abundance data. *Ecological Informatics*, *4*, 76–82.
Donoghue, D. N. M., Watt, P. J., Cox, N. J., & Wilson, J. (2007). Remote sensing of species mixtures in conifer plantations using lidar height and intensity data. *Remote Sensing of Environment*, *110*(4), 509–522.
Ducic, V., Hollaus, M., Ullrich, A., Wagner, W., & Melze, T. (2006). 3D vegetation mapping and classification using full-waveform laser scanning. *International workshop 3D remote sensing in forestry, Vienna, Feb. 14-15, 2006* (pp. 211–217).
Dunning, D. (1928). A tree classification for the selection forests of the Sierra Nevada. *Journal of Agricultural Research*, *36*(9), 755–771.
Eskelson, B. N. I., Madsen, L., Hagar, J. C., & Temesgen, H. (2011). Estimating riparian understory vegetation cover with beta regression and copula models. *Forest Science*, *57*(3), 212–221.
Espinheira, P. L., Ferrari, S. L., & Cribari-Neto, F. (2008). On beta regression residuals. *Journal of Applied Statistics*, *35*(4), 407–419.
Eyre, F. H. (1980). *Forest cover types of the United States and Canada.* Washington, DC: Society of American Foresters 148 pp.
Falkowski, M. J., Evans, J. S., Martinuzzi, S., Gessler, P. E., & Hudak, A. T. (2009). Characterizing forest succession with lidar data: an evaluation for the inland Northwest USA. *Remote Sensing of Environment*, *113*, 946–956.

Ferrari, S. L., & Cribari-Neto, F. (2004). Beta regression modeling rates and proportions. *Journal of Applied Statistics*, *31*(7), 799–815.
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.
Goodwin, N. R., Coops, N. C., Bater, C., & Gergel, S. E. (2007). Assessment of sub-canopy structure in a complex coniferous forest. *Proceedings of the ISPR Workshop "Laser Scanning 2007 and SilviLaser 2007", Espoo, September 12–14, 2007, Finland*, *Vol. XXXVI*. (pp. 169–172) ISSN:1682-1777, P3/W52.
Heurich, M., Persson, Å., Holmgren, J., & Kennel, E. (2004). Detecting and measuring individual trees with laser scanning in mixed mountain forest of central Europe using an algorithm developed for Swedish boreal forest conditions. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *36*, 307–312.
Hill, R. A., & Broughton, R. K. (2009). Mapping the understorey of deciduous woodland from leaf-on and leaf-off airborne Lidar data: A case study in lowland Britain. *ISPRS Journal of Photogrammetry and Remote Sensing*, *64*, 223–233.
Holmgren, J., & Persson, Å. (2004). Indentifying species of individual trees using airborne laser scanning. *Remote Sensing of Environment*, *90*, 415–423.
Hopkinson, C., Chasmer, L., Lim, K., Treitz, P., & Creed, I. (2006). Towards a universal lidar canopy height indicator. *Canadian Journal of Remote Sensing*, *32*, 139–152.
Hudak, A. T., Crookston, N. L., Evans, J. S., Falkowski, M. J., Smith, A. M. S., Gessler, P. E., et al. (2006). Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral satellite data. *Canadian Journal of Remote Sensing*, *32*, 126–138.
Hudak, A. T., Crookston, N. L., Evans, J. S., Hall, D. E., & Falkowski, M. J. (2008). Nearest neighbor imputation modeling of species-level, plot-scale structural attributes from lidar data. *Remote Sensing of Environment*, *112*, 2232–2245.
Hyyppä, J., Kelle, O., Lehikoinen, M., & Inkinen, M. (2001). A segmentation based method to retrieve stem volume estimates from 3-dimensional tree height models produced by laser scanner. *IEEE Transactions on Geoscience and Remote Sensing*, *39*, 969–975.
Kaasalainen, S., Hyyppä, H., Kukko, A., Litkey, P., Ahokas, E., Hyyppä, J., et al. (2009). Radiometric calibration of LiDAR intensity with commercially available reference targets. *IEEE Transactions on Geoscience and Remote Sensing*, *47*(2), 588–598.
Kass, R. E., & Raferty, R. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.
Kerns, B. K., & Ohmann, J. L. (2004). Evaluation and prediction of shrub cover in coastal Oregon forests (USA). *Ecological Indicators*, *4*, 83–98.
Kerr, J. T., & Ostrovsky, M. (2003). From space to species: Ecological applications for remote sensing. *Trends in Ecology and Evolution*, *18*, 299–305.
Kieschnick, R., & McCullough, B. D. (2003). Regression analysis of variates observed on (0, 1): Percentages, proportions and fractions. *Statistical Modelling*, *3*, 193–213.
Kim, Y., Yang, Z., Cohen, W. B., Pflugmacher, D., Lauver, C. L., & 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.
Kmenta, J. (1986). *Elements of econometrics* (2nd ed.). New York: Macmillan 733 pp.
Korhonen, L., Korhonen, K. T., Stenberg, P., Matlamo, M., & Rautianen, M. (2007). Local models for forest canopy cover with beta regression. *Silva Fennica*, *41*(4), 671–685.
Lefsky, M. A., Cohen, W. B., Harding, D. J., & Parker, G. G. (2002). Lidar remote sensing for forest ecosystem studies. *BioScience*, *52*, 19–30.
Legare, S., Bergeron, Y., & Pare, D. (2002). Influence of forest composition on understory cover in boreal mixed-wood forests of Western Quebec. *Silva Fennica*, *36*, 353–365.
Lim, K., Treitz, P., Baldwin, K., Morrison, I., & Green, J. (2003). Lidar remote sensing of biophysical properties of tolerant northern hardwood forests. *Canadian Journal of Remote Sensing*, *29*, 648–678.
Maltamo, M., Packalén, P., Yu, X., Eerikäinen, K., Hyyppä, J., & Pitkänen, J. (2005). Identifying and quantifying structural characteristics of heterogeneous boreal forests using laser scanner data. *Forest Ecology and Management*, *216*, 41–50.
Maltamo, M., Malinen, J., Packalén, P., Suvanto, A., & Kangas, J. (2006). Nonparametric estimation of stem volume using airborne laser scanning, airborne photography, and stand-register data. *Canadian Journal of Forest Research*, *36*(2), 426–436.
Martinuzzi, S., Vierling, L. A., Gould, W. A., Falkowski, M. J., Evans, J. S., Hudak, A. T., et al. (2009). Mapping snags and understory shrubs for a lidar-based assessment of wildlife habitat suitability. *Remote Sensing of Environment*, *113*, 2522–2546.
McCullagh, P., & Nelder, J. A. (1989). Generalized linear models. (2nd ed.). *Monographs on statistics and probability*, *37*, London: Chapman & Hall 511 pp.
McDermid, G. J., Franklin, S. E., & LeDrew, E. F. (2005). Remote sensing for large-area habitat mapping. *Progress in Physical Geography*, *29*, 449–474.
Morsdorf, F., Mårell, A., Koetz, B., Cassagne, N., Pimont, F., Rigolot, E., et al. (2010). Discrimination of vegetation strata in a multi-layered Mediterranean forest ecosystem using height and intensity information derived from airborne laser scanning. *Remote Sensing of Environment*, *114*, 1404–1415.
Mutlu, M., Popescu, S., Stripling, C., & Spencer, T. (2008). Mapping surface fuel models using lidar and multispectral data fusion for fire behavior. *Remote Sensing of Environment*, *112*, 274–285.
Næsset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sensing of Environment*, *80*, 88–99.
Nelsen, R. B. (2006). *An introduction to copulas.* New York: Springer 239 pp.
Oliver, W. W. (2000). Ecological research at Blacks Mountain Experimental Forest in northeastern California. *Pacific Southwest Research Station, Albany, Calif. USDA Forest Service General Technical Report. PSW-GTR-179*.
Ospina, R., & Ferrari, S. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics and Data Analysis*, *56*(6), 1609–1623.
Packalén, P., & Maltamo, M. (2006). Predicting the plot volume by tree species using airborne laser scanning and airborne photographs. *Forest Science*, *52*(6), 611–622.
Pesonen, A., Maltamo, M., Eerikäinen, & Packalén, P. (2008). Airborne laser scanning based prediction of coarse woody debris volumes in a conservation area. *Forest Ecology and Management*, *255*, 3288–3296.
Pielou, E. C. (1977). *Mathematical ecology.* New York: John Wiley & Sons 384 pp.
R Development core team (2009). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing.
Raferty, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163.
Riaño, D., Meier, E., Allgöwer, B., Chuvieco, E., & Ustin, S. L. (2003). Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sensing of Environment*, *86*, 177–186.
Ritchie, M. W., Skinner, C. N., & Hamilton, T. A. (2007). Probability of wildfire-induced tree mortality in an interior pine forest: effects of thinning and prescribed fire. *Forest Ecology and Management*, *247*, 200–208.
Russell, R. E., Saab, V. A., & Dudley, J. (2007). Habitat suitability models for cavity-nesting birds in a postfire landscape. *The Journal of Wildlife Management*, *71*, 2600–2611.
Schroeder, T. A., Cohen, W. B., & Yang, Z. (2007). Patterns of forest regrowth following clearcutting in Western Oregon as determined from a Landsat time-series. *Forest Ecology and Management*, *243*, 259–273.
Scott, J. H., & Reinhardt, E. D. (2001). *Assessing crown fire potential by linking models of surface and crown fire behavior.* Res. Pap. RMRS-RP-29. Fort Collins, CO: U. S. Department of Agriculture, Forest Service, Rocky Mountain Research Station 59 pp.
Seielstad, C. A., & Queen, L. P. (2003). Using airborne laser altimetry to determine fuel models for estimating fire behavior. *Journal of Forestry*, *101*(4), 10–15.
Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, *11*(1), 54–71.
Stage, A. R., & Salas, C. (2007). Interactions of elevation, aspect, and slope in models of forest species composition and productivity. *Forest Science*, *53*(4), 486–492.
Suchar, V. A., & Crookston, N. L. (2010). Understory cover and biomass indices predictions for forest ecosystems of the Northwestern United States. *Ecological Indicators*, *10*, 602–609.
Tremblay, N. O., & Larocque, G. R. (2001). Seasonal dynamics of understory vegetation in four eastern Canadian forest types. *International Journal of Plant Sciences*, *162*, 271–286.
Venier, L. A., & Pearce, J. L. (2007). Boreal forest land birds in relation to forest composition, structure, and landscape: Implications for forest management. *Canadian Journal of Forest Research*, *37*, 1214–1226.
Vierling, K. T., Vierling, L. A., Martinuzzi, S., Gould, W., & Clawges, R. (2008). Lidar: Shedding new light on habitat modeling. *Frontiers in Ecology and the Environment*, *6*, 90–98.
Vosselman, G., Gorte, B. G. H., Sithole, G., & Rabbani, T. (2004). Recognizing structure in laser scanning point clouds. *The Information of the Science*, *46*(8), 1–6.
Wagner, W., Hollaus, M., Briese, C., & Ducic, V. (2008). 3D vegetation mapping using small-footprint full-waveform airborne laser scanners. *International Journal of Remote Sensing*, *29*(5), 1433–1452.
Weisberg, S. (1985). *Applied linear regression* (2nd ed.). New York: Wiley 324 pp.
Wulder, M., & Franklin, S. (2003). *Remote sensing of forest environments: Concepts and case studies.* Dordrecht/Boston/London: Kluwer Academic Publishers 519 pp.
Zimble, D. A., Evans, D. L., Carlson, G. C., Parker, R. C., Grado, S. C., & Gerard, P. D. (2003). Characterizing vertical forest structure using small-footprint airborne LIDAR. *Remote Sensing of Environment*, *87*, 171–182.