Regional mapping of vegetation structure for biodiversity monitoring using airborne lidar data

Xuan Guo a,⁎, Nicholas C. Coops a, Piotr Tompalski a, Scott E. Nielsen b, Christopher W. Bater c, J. John Stadt c

a Department of Forest Resource Management, 2424 Main Mall, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada
b Department of Renewable Resources, Faculty of Agricultural, Life, and Environmental Sciences, University of Alberta, Edmonton, Alberta, T6G 2H1, Canada
c Forest Management Branch, Forestry Division, Alberta Agriculture and Forestry, 9920-108 Street NW, Edmonton, Alberta, T5K 2M4, Canada

Abstract

Vegetation structure is identified as an important biodiversity indicator providing the physical environment that generates, supports, and maintains forest biodiversity. Airborne lidar systems (light detection and ranging) have the capacity to accurately measure three-dimensional vegetation structure, and have been widely used in wildlife habitat mapping and species distribution modeling. Large-area structural inventories using lidar-derived variables that characterize generic habitat structure have rarely been done, yet would be helpful for guiding biodiversity monitoring and conservation assessments of species at regional levels. This study provides a novel approach for processing regional-scale lidar data into categorical classes representing natural groupings of habitat structure. We applied cluster analysis on six lidar-derived habitat-related variables to classify vegetation structure into eight classes for the forested areas of ten natural subregions in boreal and foothill forests in Alberta, Canada. Structure classes were compared across different natural subregions and under anthropogenic/non-anthropogenic disturbance regimes. We found that the Lower Foothills Natural Subregions had the most complex vegetation structure, and wildfire was the most prevalent disturbance agent for all classes except for the rarest class (i.e. stands with high standard deviations of height and low canopy cover) which was more heavily altered by timber harvesting. This data product provides continuous, regional mapping of vegetation structure directly measured from lidar, with a spatial resolution (30 m) relatively finer than what was provided by polygon-based forest inventories. This vegetation structure classification and its associated spatial distribution address the fundamental issue of habitat structure in biodiversity monitoring. It can serve as a base layer used together with species and land cover data for forest resources planning, species distribution and animal movement modeling, as well as prioritization of conservation efforts on critical habitat structures.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Vegetation structure is considered an important component in wildlife and forest management (Noss, 1990; McCleary and Mowat, 2002; Lindenmayer et al., 2006). For example, forest canopies mediate microclimate, provide perching, nesting, foraging, and covering habitats for many animal species, and influence food quality, diversity and accessibility (Hamer and Herrero, 1987; Johnson et al., 2002). Forest structure is also inextricably affected by disturbance regimes, especially wildfire, harvesting, and road development, which may favour certain species while discouraging others (Tews et al., 2004; Devictor et al., 2008; Boutin et al., 2009; Desrochers et al., 2012). Forest horizontal and vertical structures have therefore been identified as essential biodiversity indicators across a broad range of forest ecosystems around the world (Ozanne et al., 2003; Chirici et al., 2011; Gao et al., 2014). In general, forest species diversity is positively associated with vegetation structural diversity because different vertical strata and structural heterogeneity of forest stand provide ecological niches for species of various habitat specializations (MacArthur, 1958; Hunter, 1999; Culbert et al., 2013). This linkage between forest biodiversity and forest structure is a central assumption in ecosystem-based management approaches to forestry, where forest managers attempt to maintain the diversity of forest structural attributes at both landscape and stand scales in order to maintain forest biodiversity (Hunter, 1993).

MacArthur (1972) identified productivity, climatic stability, and habitat structure as three primary drivers of biodiversity whose effects can be reflected in three aspects: composition, structure, and function (Franklin et al., 1981; Noss, 1990). Spectral information acquired from optical remote sensing data has been widely used to assess compositional and functional components of biodiversity over broad spatial scales (Cohen and Goward, 2004; Duro et al., 2007; Coops et al., 2008; Schuster et al., 2015). Habitat classifications based on land cover types (Wessels et al., 2000; Franklin et al., 2001; McDermid et al., 2009; Riggo et al., 2013), and habitat suitability indices derived from vegetation productivity and seasonality (Nilsen et al., 2005; Coops et al., 2008).
have contributed significantly to species distribution models and animal movement studies. As opposed to correlations with spectral indices, the structural component of biodiversity has principally been assessed through forest resource inventories that require labor-intensive field surveys and/or aerial photo interpretation (Fensham et al., 2002; Hyde et al., 2006; Clawges et al., 2008; Nijland et al., 2015b). In addition, resource inventories submitted by multiple forest management stakeholders may lack consistency in interpretation standards, update schedule, and aerial coverage when collectively used for large-area habitat mapping (McDermid et al., 2009). Moreover, vegetation height estimates from photo interpretation are reported at the polygon level where the within-polygon variations in height and structure are not readily assessed (Culbert et al., 2013). More detailed, fine-scale mapping of vegetation structure is needed to allow a broader range of biodiversity values to be included in forest management planning.

Airborne lidar (light detection and ranging) is an active remote sensing technology that can accurately measure three-dimensional vegetation structure (Lim et al., 2003). Lidar-derived canopy height, canopy height variation, and canopy cover metrics have been used widely in forest ecological studies to determine or predict a number of important forest attributes, including: forest vertical layering and overall architecture (Maltamo et al., 2005); forest successional stages (Falkowski et al., 2009); vegetation strata and forest genera (Morsdorf et al., 2010; Kim et al., 2011); tree species abundance (Ewijk et al., 2014); forest volume, biomass and carbon storage (Zald et al., 2014); vegetation regeneration; and response after timber harvesting (Nijland et al., 2015b).

Although lidar technology cannot directly measure forest biodiversity, previous studies have examined the hypothesis that vegetation structure is an important indicator of species diversity as postulated by MacArthur and MacArthur (1961) and Erdelen (1984). Species
distribution and habitat models built on lidar-derived structural variables indicate strong relationships between bird species occurrence, richness, and canopy height and cover metrics (Wulder et al., 2008; Graf et al., 2009; Goetz et al., 2010; Hovick et al., 2014; Hill and Hinsley, 2015). For example, positive correlations have been found between avian species richness and foliage height diversity (Clawges et al., 2008; Bergen et al., 2009), as well as canopy height (Goetz et al., 2007). Coops et al. (2016) found that adding variables correlated with vegetation structure to existing bird richness models improved model predictive power dramatically. Lidar-derived structural metrics have also been used to examine habitat suitability for plant, bird and mammal species, including grizzly bears (Ursus arctos) and African lions (Panthera leo) (Bergen et al., 2009; Mueller et al., 2009; Lucas et al., 2010; Simonson et al., 2012; Jung et al., 2012; Gao et al., 2014; Nijland et al., 2015a; Davies et al., 2016).

Biodiversity monitoring programs focused on a limited list of indicator species may not provide an adequate measure for biodiversity as a whole due to the fact that these species may not represent the full range of trophic levels and habitat specializations (Marcot et al., 1994; Boutin et al., 2009; Marchese 2015). Furthermore, current lidar-based species habitat studies with fine-spatial resolution data tend to focus on small study sites, which may in turn restrict the generality of the analyses to broad-scale, multispecies, habitat-based monitoring (Boutin et al., 2009; Zellweger et al., 2013; Vogeler et al., 2014). In contrast, habitat monitoring projects at regional or national levels commonly use generalized variables from remote sensing data of coarser spatial resolution (McDermid et al., 2005).

Broad-scale, multispecies, and multi-dimensional biodiversity monitoring is difficult and costly (Gaston 1996, Green et al., 2005). A habitat-based approach may better lend itself to the conservation of plant and wildlife species diversity than species-specific monitoring of biodiversity. Therefore we propose to examine vegetation structure across a broad area as an indicator to infer overall habitat potential. Vegetation structure is a major driver of forest species diversity by representing the physical context that generates, supports, and maintains biodiversity (Morriz, 2001; Marchese, 2015). Noss (1990) also emphasized that a top-down assessment of biodiversity status should start with a broad-scale inventory of vegetation composition, habitat structure, and their landscape patterns. Without a comprehensive understanding of what types of habitat structure exist over the landscape, biodiversity monitoring based solely on relationships between species diversity, environmental gradients, vegetation composition, and productivity developed from localized studies could be biased and incomplete.

In this study, we utilized a large-area discrete-return lidar dataset covering the managed forests of Alberta, Canada, to develop an inventory of vegetation structure that efficiently synthesizes forest vertical variation into distinct categorical classes for use in conservation and management activities. To do so, we selected six structure-related lidar metrics that measure canopy height, cover, and height variation at a 30 m spatial resolution (i.e. grid cell size) covering the study area (the majority of Alberta's forests). Second, we applied cluster analysis to classify forest structure into eight classes and used discriminate function analysis to contrast variable importance within and between identified classes. Finally, we interpreted and compared forest attributes of species and age composition, wetland presence, and the effects of disturbance (both anthropogenic and non-anthropogenic) regimes in the derived classes.

### 2. Materials and methods

#### 2.1. Alberta natural regions and subregions

Alberta's landscape is divided into six natural regions based on the influence of climate, topography, and geology (Natural Regions Committee, 2006). In this study, we focused on the Boreal Forest and Foothills Natural Regions, which together cover ~44 million ha of the provincial land base (68%), and support Alberta's biodiversity by providing critical seasonal and permanent habitats to numerous wildlife species (Brandt, 2009).

The Boreal Forest Natural Region (Fig. 1) is the largest natural region in Alberta. This region is influenced by short summers and long, cold winters, with deciduous, mixedwood and coniferous forest intertwined with extensive wetlands (Natural Regions Committee, 2006). Eight natural subregions occur within this region, with elevations ranging from 150 m to 1100 m above sea level. Upland forests are composed of trembling aspen (Populus tremuloides), balsam poplar (Populus balsamifera), white spruce (Picea glauca), and jack pine (Pinus banksiana), with more pine-dominant stands in areas with higher elevation. Treed wetlands are characterized by black spruce (Picea mariana) and western larch (Larix occidentalis), with both bogs and fens being common across the landscape.

Prominent wildlife species of the Boreal Forest Natural Region include moose (Alces alces), caribou (Rangifer tarandus), black bear (Ursus americanus), wolf (Canis lupus), beaver (Castor canadensis), snowshoe hare (Lepus americanus), and many migratory birds such as whooping crane (Grus americana). About 110 rare plant species were reported in this natural region across a range of upland and wetland ecosystems (Natural Regions Committee, 2006). Major land uses in this region include agricultural cultivation, oil and gas extraction and significant harvesting of timber.

### Table 1

<table>
<thead>
<tr>
<th>Lidar metrics</th>
<th>Explanation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD of height</td>
<td>Standard deviation of vegetation height</td>
<td>Clawges et al., 2008</td>
</tr>
<tr>
<td>Canopy cover</td>
<td>% of total first returns above 1.37 m</td>
<td>Vogeler et al., 2014</td>
</tr>
<tr>
<td>Canopy height density - 1.37 to 5 m</td>
<td>% of first return between 1.37 and 5 m</td>
<td>Bergen et al., 2009</td>
</tr>
<tr>
<td>Canopy height density - 5 to 10 m</td>
<td>% of first return between 5 and 10 m</td>
<td>Næsset &amp; Gobakken, 2008</td>
</tr>
<tr>
<td>Canopy height density - 10 to 20 m</td>
<td>% of first return between 10 and 20 m</td>
<td>Næsset &amp; Gobakken, 2008</td>
</tr>
<tr>
<td>Canopy height density - 20 to 30 m</td>
<td>% of first return between 20 and 30 m</td>
<td>Næsset &amp; Gobakken, 2008</td>
</tr>
</tbody>
</table>

### Table 2

Pearson's correlation coefficients for six lidar variables ($p$ value = 0.05).

<table>
<thead>
<tr>
<th></th>
<th>SD of height</th>
<th>Canopy cover</th>
<th>1.37 to 5 m</th>
<th>5 to 10 m</th>
<th>10 to 20 m</th>
<th>20 to 30 m</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD of height</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canopy Cover</td>
<td>0.49</td>
<td>1.00</td>
<td>-0.34</td>
<td>0.02</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Canopy Height Density - 1.37 to 5 m</td>
<td>-0.34</td>
<td>0.02</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canopy Height Density - 5 to 10 m</td>
<td>0.05</td>
<td>0.52</td>
<td>0.14</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canopy Height Density - 10 to 20 m</td>
<td>0.47</td>
<td>0.68</td>
<td>-0.49</td>
<td>0.06</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Canopy Height Density - 20 to 30 m</td>
<td>0.53</td>
<td>0.36</td>
<td>-0.30</td>
<td>-0.22</td>
<td>0.19</td>
<td>1.00</td>
</tr>
</tbody>
</table>
The Foothills Natural Region (Fig. 1) extends east from the flank of the Rocky Mountains, north from the Bow River Valley (51.176°N, 115.570°W) to just south of Grande Prairie (55.170°N, 118.799°W), covering 10% of the province. The region is divided into two natural subregions characterized by a cool, moist climate and long growing seasons, with gently undulating to rolling terrain ranging from 700 m to 1700 m above sea level into the Rocky Mountains. Dominant vegetation cover in the Lower Foothills Natural Subregion is mixedwood forest, consisting of trembling aspen, white spruce, lodgepole pine (Pinus contorta) and balsam poplar, whereas the Upper Foothills Natural Subregion is more dominated by uniform lodgepole pine and black spruce stands.

Variable topography, surface and groundwater regimes create diverse plant and wildlife communities in the Foothills region. This region provides critical habitats for grizzly bear, woodland caribou (Rangifer tarandus caribou), many song bird species and about 80 rare plant species (Natural Regions Committee, 2006). However, the Foothills Natural Region is heavily managed for commercial timber production and rapidly expanding oil and gas exploration which have created a fragmented landscape across the region (Linke et al. 2005).

2.2. Lidar data and derived metrics

The Government of Alberta, Canada has acquired airborne discrete-return lidar data over > 33 million hectares of forested area in the province, covering 75% of Alberta foothills and boreal regions (Fig. 1). The data used in this study were acquired between 2003 and 2014, with more than half of the data obtained in 2007 and 2008. The pulse density ranges from 1 to 4 returns per m². These data were collected by multiple contractors with a scan angle of ~25° from nadir and a vertical accuracy of no larger than 30 cm root mean square error (Alberta Environment and Sustainable Resource Development, 2013). Bare-Earth products derived from these data were used to normalize the point elevations to height above ground level. A suite of forest canopy metrics were generated using a 30 m grid based on normalized first returns of the lidar point cloud data using a combination of software packages specialized for lidar data processing, including FUSION (McGaughey, 2015) and LAStools (Isenburg, 2016).

Based on existing literature, we selected six lidar-based metrics that were commonly related to forest biodiversity (Table 1). As height-related metrics were among the most widely used lidar metrics.
for habitat monitoring and assessments, we chose four height strata (1.37 m to 5 m, 5 m to 10 m, 10 m to 20 m, 20 m to 30 m) and calculated canopy height density (percentage of lidar returns) within each height stratum. A suite of metrics on canopy height density can reveal more details about canopy height distribution along the vertical dimension than direct lidar height measurements, thus we used a combination of canopy height metrics in this study. Short shrubs below 1.37 m were not considered. Previous studies have found strong correlations between bird species diversity and the proportional lidar returns near the forest floor (<5 m) (Clawges et al., 2008; Bergen et al., 2009; Mueller et al., 2009). Tall shrub and tree saplings below 5 m provide important habitat for many bird and mammal species. Therefore, canopy height density between 1.37 m to 5 m was used. The other three height strata used in this study represented the range of lower (5–10 m), middle (10–20 m) and upper (20–30 m) canopy height distributions in Alberta which might be utilized by wildlife species with different habitat needs (Coops et al., 2016).

Beyond canopy height density, standard deviation (SD) of height has also been used as an important predictor of bird richness and abundance (Bergen et al., 2009; Mueller et al., 2009; Culbert et al., 2013; Vogeler et al., 2014). Forest stands with low standard deviations of height tend to have less structural diversity than stands with high standard deviations of height, therefore providing less niche habitat for plant and wildlife species (August 1983). Canopy heterogeneity has also been used as an indicator of the presence of large snags for cavity-nesting species (Martinuzzi et al., 2009; Bater et al., 2009).

Canopy cover was measured as the total percentage of lidar returns above a minimum tree height threshold (1.37 m in this study). Open canopy stands with strong light penetration near ground level are associated with high exposure to solar radiation in shrub and understory layers. These stands generally have high volume of understory plant biomass and plant species diversity which provide important habitat for understory and shrub nesting birds species and arthropod communities (Blakely and Didham, 2010). In contrast, medium to closed-canopy stands help maintain a suitable moisture content for lichen growth which could provide critical winter habitat for caribou and other ungulate speices (Kershaw, 1977; Apps et al., 2001; Coops et al., 2010).

### Table 3

<table>
<thead>
<tr>
<th>Class Description 1</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>Total Count</th>
<th>User’s accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short, medium canopy cover stand</td>
<td>29351</td>
<td>3586</td>
<td>9576</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1975</td>
<td>41,535</td>
<td>0.74</td>
</tr>
<tr>
<td>Short, open canopy cover stand</td>
<td>2031</td>
<td>1225</td>
<td>1843</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10,186</td>
<td>0.64</td>
</tr>
<tr>
<td>Very short, dense canopy cover stand</td>
<td>1473</td>
<td>31761</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>36,614</td>
<td>0.74</td>
</tr>
<tr>
<td>Very tall, complex stand</td>
<td>413</td>
<td>0</td>
<td>0</td>
<td>20,709</td>
<td>398</td>
<td>1192</td>
<td>0</td>
<td>1381</td>
<td>24,093</td>
<td>0.86</td>
</tr>
<tr>
<td>Very tall, open canopy stand</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1975</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1975</td>
<td>1</td>
</tr>
<tr>
<td>Tall, dense canopy cover stand</td>
<td>5946</td>
<td>868</td>
<td>3970</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>76,442</td>
<td>0.847</td>
</tr>
<tr>
<td>Tall, closed canopy cover stand</td>
<td>203</td>
<td>0</td>
<td>86</td>
<td>0</td>
<td>0</td>
<td>108</td>
<td>2716</td>
<td>0</td>
<td>27,222</td>
<td>0.998</td>
</tr>
<tr>
<td>Very tall, closed canopy stand</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1881</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13,797</td>
<td>15,681</td>
<td>0.88</td>
</tr>
<tr>
<td>Total Count</td>
<td>41,535</td>
<td>99,315</td>
<td>44,798</td>
<td>27,116</td>
<td>6694</td>
<td>66,187</td>
<td>36,390</td>
<td>15,183</td>
<td>337,218</td>
<td></td>
</tr>
<tr>
<td>Average Accuracy:</td>
<td>84%</td>
<td>Producer’s accuracy</td>
<td>0.72</td>
<td>0.96</td>
<td>0.71</td>
<td>0.96</td>
<td>0.76</td>
<td>0.96</td>
<td>0.76</td>
<td></td>
</tr>
</tbody>
</table>

2.3. Disturbance and vegetation inventory plots

Three data sources were used in the study to interpret vegetation attributes and disturbance regimes for lidar-based structure classification. The Alberta historical wildfire database is a polygon-based GIS (geographic information system) layer with spatial information of large fires (>200 ha) since 1961 (Alberta Environmental Protection, 1991). We used in our analyses wildfire information from this database between 1980 and 2010. An anthropogenic disturbances database was derived from a nationwide inventory of polygonal disturbance types across the Canadian boreal ecosystem produced through the interpretation of 2008–2010 Landsat images at 1:50,000 viewing scale (Pasher et al., 2013). Cutblocks, agricultural development, oil and gas exploration, and urban settlement were the major disturbance types recorded in the database. The Alberta Biodiversity Monitoring Institute (ABMI) has established a provincial monitoring system with 1656 rectangular permanent photo plots (3 km north to south, 7 km west to east) distributed in a 20 km systematic grid covering 5% of Alberta’s land base to evaluate and report biodiversity status in the province (Castilla et al., 2013). We chose photo plots with images collected between 2008 and 2011 to generally match the year of the anthropogenic disturbance layer in this analysis. Within each photo plot, detailed vegetation inventory and land use classification with a minimum mapping unit of 2 ha were produced based on high-resolution (0.30 m spatial resolution) aerial photo interpretation.
Cluster analysis is an exploratory, unsupervised classification technique that agglomerates objects based on perceived intrinsic similarity of the data to identify natural groupings (Kaufman and Rousseeuw, 2009). It is used to reveal the general characteristics and underlying structure of data. An ideal cluster contains a set of objects which are similar to each other within the cluster, but isolated from objects in other clusters. This method has been widely used in multivariate-based ecosystem classifications (Fitterer et al., 2012; Thompson et al., 2016) and ecoregion zoning (Coops et al., 2009). In our analysis, a two-step clustering analysis (Chiu et al., 2001) was used to accommodate our large dataset and associated computing time. A k-means clustering algorithm was applied as the first step to classify the data into pre-clusters, replacing raw data. These pre-clusters were then grouped using a commonly used agglomerative hierarchical clustering method (Unweighted Pair Group Method with Arithmetic Mean) based on Euclidean distance.

Before cluster analysis, lidar data points were filtered to remove vegetation heights above 45 m (air points) and below 1.37 m (which corresponds to the height at which stem diameters are measured in the field). All data were standardized to be within the same range by multiplying a scalar. Pearson’s correlation test (at \( p \text{-value} = 0.05 \)) was performed to assess correlated relationship between the six variables. The k-mean clustering algorithm generated 100 pre-clusters, which were in turn further reduced to form the final structure classes through the hierarchical clustering process. The two-step clustering process was conducted in R software with the packages “bigmemory” and “biganalytics” (Kane and Emerson, 2010).

To test the uniqueness of the structure classes, a non-parametric Kruskal-Wallis rank sum test (Breslow 1970), followed by a Dunn’s post hoc multiple comparison test, was used to examine if the overall clustering was significant for each selected lidar variable, and if the pairwise difference between clusters was significant for each individual variable. When considering all variables together, a two-sample location test based on marginal ranks (Sen and Puri, 1971) was used as a non-parametric multivariate statistical test to determine if the clustering results were significant with consideration of all selected variables. Linear discriminate function analysis was applied on 10% of the data to verify the separability of clustering results and assess variable importance from a modeling perspective. The predicted structure classes and the six variables used in the discriminant function analysis were plotted against the first two axes of the discriminate coefficients to contrast structural characteristics within and between structure classes and assess the strength of each variable in discriminating these classes.

To interpret the dominant forest features in each structure class, such as dominant species, layering architecture, density class, age composition, and wetland coverage, we sampled and overlaid 58 ABMI photo plots (Fig. 1) containing > 15,000 polygons with our classification. To examine the influence of anthropogenic and non-anthropogenic disturbance on each of the structure classes, the historical wildfire and anthropogenic disturbance databases were overlaid with the structure classification. At the polygon level, the information in the ABMI photo plots and the two disturbance inventories were summarized to each structure class by pixel in the overlaid area.

**3. Results**

Pearson’s coefficients for the six lidar variables were significantly lower than 0.7 (\( p \text{-value} = 0.05 \)), indicating no significantly strong linear correlation between selected variables (Table 2). Fig. 2 shows the dendrogram from the two-step clustering and highlights where the natural breaks for eight structure classes were apparent in terms of maximizing the dissimilarity between an appropriate number of structure classes. Both the Kruskal-Wallis rank sum test and the two-sample location test based on marginal ranks indicated a significant difference between the eight classes for each individual variable and for all six variables (\( p \text{-value} = 0.05 \)). The Dunn’s post-hoc test between paired samples was also significant for all variables (\( p \text{-value} = 0.05 \)), except for the standard deviation of height between classes 4 and 5, and the canopy height density between 20 and 30 m for classes 3 and 7 (Fig. 3). The discriminate function analysis on the test dataset indicated an accurate classification (average accuracy: 84%) for all the classes except class 5 which was often misclassified into class 1 or 2 (Table 3). The eight structure classes were all well separated when plotted against the first two discriminate coefficient axes (Fig. 4) which explained 80% of the total variance. Total canopy cover and canopy height density between 20 and 30 m were the most influential variables in terms of discriminatory power among different structure classes.

The spatial distribution of the eight classes is shown in Fig. 5(1) and an illustration of the stand profiles of the eight classes using samples of the lidar point cloud is demonstrated in Fig. 5(2).
natural subregions exhibit different spatial patterns of structure classification. For example, black spruce-dominated wetland forest type associated with structure classes 2 and 3 was more commonly found in the Upper Boreal Highland Natural Subregion and Central Mixedwood Natural Subregion (A and B), whereas aspen-dominated upland forest type associated with structure class 8 and 6 was more often in the Dry Mixedwood Natural Subregion (C). From the stand profiles, it was clear that four of the eight classes were driven by differences in canopy height density in each height stratum (classes 3, 6, 7 and 8). And the other four classes were differentiated by canopy cover and standard deviation (SD) of tree height (classes 1, 2, 4, and 5).

Short and open canopy stands with low standard deviations (SD) of height and low canopy cover (class 2) were the most common stand structure type in all natural subregions, except the Dry Mixedwood, Lower Foothills and Upper Foothills Natural Subregions (Fig. 6). This class was characterized by black spruce-populated wetlands with wildfire as the most common disturbance agent (Table 4). The multimodal age distribution of this class might indicate cohorts of regenerated upland forest stand type dominated by aspen species after disturbance (Fig. 7; Table 4). Stands with tall, dense canopy cover and dominant heights between 10 and 20 m (class 6) were associated with the second largest class by aerial extent and were most commonly found in...
the Central and Dry Mixedwood, Lower and Upper Foothills Natural Subregions, and in the higher elevated areas in the Lower Boreal Highlands Natural Subregion (Fig. 6). Based on the ABMI photo plots, this class was characterized by dense, aspen-dominated upland forest stands with uneven age composition (Fig. 7). Two aspen dominated structure classes (class 4 and class 5) with high standard deviations of height, but contrasting canopy cover, were the most structurally complex stands in the study area. In fact, the stand type represented by structure class 5 accounted for <1% of the study area and was scattered throughout the Central and Dry Mixedwood, Lower and Upper Foothills Natural Subregions (Fig. 6), with strong associations with harvesting activities (40%) and recent wildfires (15%) (Fig. 8).

As indicated from Fig. 5(1), tall and closed-canopy, aspen-dominat-ed, upland stand type (class 8) were spatially concentrated along river floodplains (A), at a lower slope position along transitional areas between Dry Mixedwood and Lower Boreal Highland Natural Subregions (B), and between Central Mixedwood and Lower Boreal Highland Natural Subregions (C). Forest stands with dense canopy cover and very short vegetation heights ranging from 1.37–5 m (class 3) were mostly black spruce-dominated wetlands with wildfires being a common disturbance agent (Table 4). This stand type was most abundant in Central Mixedwood and Lower Boreal Highlands Natural Subregions (Fig. 6). Closed-canopy stands with dominant tree heights ranging from 5 to 10 m (class 7) were most encompassing with respect to stand structure of a broad range of species and age compositions (Fig. 7). Structure class 1 associated with stands of short tree height and medium canopy cover was mostly aspen-dominated forest type with moderate influence of human and wildfire disturbance (Table 4; Fig. 8). We speculated that the multimodal age composition of structure classes 1 and 3 indicated cohorts of recently generated forest stands following disturbances (Fig. 7).

4. Discussion

4.1. Structure classification

Canopy height density and canopy cover are the two major factors discriminating different structure classes. The contribution of standard deviation of height is less significant indicating some degree of multicollinearity between canopy cover, canopy height and standard deviation of height. In general, tall stands tend to have high standard deviations of heights. Also, single-layer stands with high canopy cover usually have lower standard deviations in heights than more complex stands. The low predictive accuracy of structure class 5 in discriminant function analysis may be due to the extremely low areal coverage of this class and large variations in heights. With canopy height distributed evenly along the vertical dimension, this structure class was more likely to be misclassified with other classes whose dominant height is within the range of height variations of the class 5 forest structure type.

4.2. Habitat value and associated ecological process

Vegetation height variation in different natural subregions is predominately influenced by elevation, climate regimes and tree species distribution (Hansen et al., 2014). Optimal slope position, soil nutrients and drainage condition (Little et al., 2002, Zhang et al., 2011) also

Table 4

Interpretation of each structure class from ABMI photo plots (species code: AW = trembling aspen; SB = black spruce. Density class (overstorey/understorey): A = 6–30% canopy closure; B = 30–50% canopy closure; C = 50–70% canopy closure. Age composition: the number in the bracket indicates the year of origin).

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
<th>Leading species</th>
<th>Age composition</th>
<th>Density class</th>
<th>Percent wetlands</th>
<th>Percent fire</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Short, medium canopy cover stand</td>
<td>Aw</td>
<td>Multimodal (1950,2000)</td>
<td>A/-</td>
<td>16%</td>
<td>13%</td>
</tr>
<tr>
<td>2</td>
<td>Short, open canopy stand</td>
<td>Sb</td>
<td>Multimodal (1950,2000)</td>
<td>A/-</td>
<td>64%</td>
<td>28%</td>
</tr>
<tr>
<td>3</td>
<td>Very short, dense canopy cover stand</td>
<td>Sb</td>
<td>Multimodal (1950,1980)</td>
<td>A/B</td>
<td>44%</td>
<td>21%</td>
</tr>
<tr>
<td>4</td>
<td>Very tall, complex stand</td>
<td>Aw</td>
<td>Unimodal (1910)</td>
<td>C/A</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>5</td>
<td>Very tall, open canopy stand</td>
<td>Aw</td>
<td>2000</td>
<td>A/-</td>
<td>5%</td>
<td>17%</td>
</tr>
<tr>
<td>6</td>
<td>Tall, dense canopy cover stand</td>
<td>Aw</td>
<td>Unimodal (1960)</td>
<td>C/-</td>
<td>4%</td>
<td>2%</td>
</tr>
<tr>
<td>7</td>
<td>Short, closed canopy stand</td>
<td>Aw &amp; Sb</td>
<td>Unimodal (1950)</td>
<td>A/-</td>
<td>18%</td>
<td>18%</td>
</tr>
<tr>
<td>8</td>
<td>Very tall, closed canopy stand</td>
<td>Aw</td>
<td>Unimodal (1910)</td>
<td>C/-</td>
<td>1%</td>
<td>1%</td>
</tr>
</tbody>
</table>
facilitated the formation of spatial clusters of tall and dense forest stands (class 8) across the study area (Fig. 5a) which may provide cover and protection for certain wildlife species such as great gray owl (Strix nebulosa) and woodland caribou (Servheen and Lyon, 1989; Duncan, 1997). Large trees, the presence of which is positively related to tree height, may form snags and cavities, providing important habitat for many cavity-nesting species (Conner and Adkisson, 1977; Nelson et al., 2005; Jung et al., 2012).

Dense stands with the most complex vegetation structure (class 4) were highly abundant in the Lower Foothills Natural Subregion where rolling terrain and various topographic features were found. These areas span a broad range of environmental gradients creating diverse microhabitats for vegetation structural development (Opedal et al., 2015). Similarly, Hansen et al. (2014) confirmed that mature forest with fully-developed canopy structures were more likely found in mountainous areas than on flat landscapes in southeastern US. Areas of rough terrain are often associated with a high level of plant species richness (Opedal et al., 2015) with terrain shading as an important factor affecting the occurrence of individual plant species (Nijland et al., 2014). As structurally complex areas tend to accommodate a higher level of biodiversity (Vierling et al., 2008), this structure class (class 4) with its topographic context may be of high interest for biodiversity monitoring and conservation.

Tall, well-developed, semi-open forest stands were previously reported to have higher levels of plant, bird and mammal species diversity compared to dense, closed-canopy stands (Hobson and Bayne, 2000; Gil-Tena et al., 2007; Smart et al., 2012). Semi-open stands with different levels of light penetration allow both light-adapted and shade-adapted species to flourish, and also provide a variety of food resources and sufficient open space for wildlife species to forage (Nielsen et al., 2004). Intuitively, these findings may indicate that the tall, multi-layered stands with structure class 4 and 5 have more overall habitat
potential than dense, closed-canopy stands with structure class 8. However, with different disturbance regimes altering vegetation structure pattern, further research is need to quantitatively assess the changes of canopy cover, vertical architecture, stand dynamics and landscape patterns after disturbance on forest biodiversity. In this study, we found human disturbance increased complex vegetation structure (class 5) which was geographically rare across the landscape.

4.3. Implication of the data product

This study integrates canopy cover, canopy height and height variation into a single classification scheme based on natural variation in vertical structure of vegetation in Alberta’s boreal and foothills forest. The 30 m resolution (grain or cell size) of the structure classification is sufficient to reflect subtle differences in vegetation structure, especially in fragmented landscapes such as cutblocks and other human disturbed areas (Fig. 9a). Satellite-based optical land cover mapping is relatively insensitive to this level of structure variation (Wulder, 1998; Lim et al., 2003; Goetz et al., 2007). As opposed to polygon-based vegetation inventory plots, our approach captures within-polygon variations in terms of tree height and canopy gaps which can be reported to reflect fine-scale habitat values (e.g. snags and cavity trees) (Fig. 9b). Furthermore, this study follows a consistent and objective classification scheme that does not require costly, labor-intensive manual digitalization and photo interpretation. In addition, the qualitative structural inventory and stratification can serve as a descriptive tool to augment quantitative resource inventories when accurate estimation of timber attributes (e.g. volume, stocking) is not required for management activities, such as wildlife conservation and forest protection where timber harvesting is not the top priority (Reque and Bravo, 2008).

The structure classification that we derived has the capacity to help conservation planners identify areas of high conservation value and interest for follow-up remote sensing and ground data collection. Our classification of vegetation structure describes habitat-related vegetation features independent of a specific wildlife species, adding flexibility to wildlife monitoring programs across different taxonomic groups. However, each structure class is likely to encompass a variety of biological and ecological conditions, making labeling classes with accurate biodiversity traits difficult. As a result, this classification should be considered as a base layer used to generate a series of value-added data products when integrated with spatial information of wildlife species distribution, land cover types, forest inventories, ecosystem classification and human disturbance data. Without this additional information, the vegetation structure classification alone may lack biological and ecological relevance. For example, a black spruce-dominated bog could have a similar stand structure to that of a burned or harvested aspen stand with vegetation regeneration, but provides different habitat values.

The lidar data used in this study is a provincial compilation collected through multiple years and seasons by various data providers using...
different sensors, however following the same data acquisition standards. Wildfire and land use changes during the period of data collection may not be fully represented in our analysis because of temporal misregistration. However, with the majority of the data collected in 2007 and 2008, this discrepancy should be alleviated. Also, time series of image products containing information about land cover changes locally or regionally can be overlaid with the classification to update the spatial distribution of vegetation structure to a more contemporary status. Photogrammetric point cloud data can also be processed to develop similar structural variables to classify vegetation structure for local areas with fine-resolution, cost-efficient sensors. This study demonstrates how regional lidar data can be utilized to classify and map vegetation structure for large-area wildlife and biodiversity monitoring initiatives.

5. Conclusions

Our approach to classify forest structure using lidar data across the majority of the forested area in Alberta, Canada highlights the maturity of the technology and is indicative of its widespread uptake and utilization. The eight unique structure classes reflect the general vegetation structure types in Alberta across different natural subregions. The inclusion of information on vegetation structure derived from lidar remote sensing can improve our ability to describe forest biodiversity patterns over that of optical sensor data. When combined with species and land cover information, the derived classification should be very useful for forest management planning, biodiversity monitoring and prioritization of conservation programs.

Acknowledgements

This work was funded by the Government of Alberta (GOA: 16GREFMB08) and a Natural Sciences and Engineering Research Council (NSERC) (RGPIN 311926-13) Discovery grant to N. Coops. The ALS data were provided by Alberta Agriculture and Forestry. We would like to thank the anonymous reviewer and the editor for their insightful comments on the paper, which helped us improve the work.

References


